



**HAL**  
open science

# Deux méthodes d'inférence statistique appliquées aux données de sismique réflexion profonde : détection de signaux et localisation d'onde

Mirko van Der Baan

## ► To cite this version:

Mirko van Der Baan. Deux méthodes d'inférence statistique appliquées aux données de sismique réflexion profonde : détection de signaux et localisation d'onde. Géophysique [physics.geo-ph]. Université Joseph-Fourier - Grenoble I, 1999. Français. NNT : . tel-00745500

**HAL Id: tel-00745500**

**<https://theses.hal.science/tel-00745500>**

Submitted on 25 Oct 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**OBSERVATOIRE DE GRENOBLE**  
et  
**LABORATOIRE DE GÉOPHYSIQUE INTERNE ET TECTONOPHYSIQUE**

**THÈSE**

présentée par

**Mirko VAN DER BAAN**

pour obtenir le titre de DOCTEUR de

**L'UNIVERSITÉ JOSEPH FOURIER - GRENOBLE I**

(Arrêtés ministériels du 5 juillet 1984 et du 30 mars 1992)

Spécialité: Géophysique - Géochimie - Géomécanique

---

---

**Deux méthodes d'inférence statistique  
appliquées aux données de sismique réflexion profonde:  
Détection de signaux et localisation d'onde**

---

---

Date de soutenance: 22 novembre 1999

Composition du jury:

M. M. Campillo	Président	Professeur, Université de Grenoble
M. D. Gibert	Rapporteur	Professeur, Université de Rennes
M. R. Snieder	Rapporteur	Professeur, Université d'Utrecht
M. F. Glangeaud	Examinateur	Directeur de recherche, CNRS, Grenoble
M. B. van Tiggelen	Examinateur	Chargé de recherche, CNRS, Grenoble
Mme. A. Paul	Directeur de thèse	Chargé de recherche, CNRS, Grenoble



Thèse de Doctorat  
de l'Université Joseph Fourier Grenoble I

---

*Titre de l'ouvrage:*

**Deux méthodes d'inférence statistique appliquées  
aux données de sismique réflexion profonde:  
Détection de signaux et localisation d'onde**

*Auteur:*           **Mirko VAN DER BAAN**

*Établissement:*

**Observatoire de Grenoble  
Laboratoire de Géophysique Interne et  
Tectonophysique**

*Résumé*

L'objectif de toute acquisition de données géophysiques est d'obtenir une image plus précise de l'intérieur de la Terre. Afin de préciser cette image, on utilise principalement des méthodes d'inférence déterministe ayant comme but de fournir une description exacte de chaque détail particulier.

Contraire aux méthodes déterministes, les approches statistiques n'essaient de saisir que les caractéristiques principales de l'ensemble des détails. Ceci réduit la complexité mathématique, nous permettant ainsi de mieux étudier certains phénomènes physiques. De plus, elles peuvent nous aider à sonder la Terre à une échelle plus petite que la résolution des méthodes déterministes. Finalement, elles peuvent nous fournir une appréciation de vraisemblance, nous aidant ainsi à prendre des décisions ou à reconnaître des signaux.

Dans ce mémoire, nous avons utilisé ces 3 avantages sur 2 problèmes différents. D'abord, nous les avons employés afin d'augmenter le rapport signal sur bruit en détectant des réflexions dans des données de sismique réflexion profonde d'une faible qualité. Ceci nous a permis de faciliter l'interprétation du profil considéré. Deux méthodes différentes ont été considérées, celle dite Classification par Vraisemblance Maximale et celle des Réseaux de Neurones.

Ensuite, nous les avons utilisés afin d'étudier le processus de la diffusion forte dans un milieu acoustique, stratifié et aléatoire. Afin d'étudier ce problème, nous avons appliqué la théorie de la localisation d'onde prédisant l'atténuation apparente et la dispersion d'une impulsion traversant un tel milieu. Cette théorie nous a aussi permis d'estimer la magnitude et la dimension typiques des fluctuations des vitesses et le facteur de qualité de l'atténuation intrinsèque sous le Bassin de Mauléon dans les Pyrénées Occidentales. Ceci nous a montré que la croûte supérieure y est plus hétérogène et plus atténuante que la croûte inférieure et que les dimensions typiques y sont plus petites.

*Mots clés*

Sismologie - sismique réflexion profonde - détection de signaux - reconnaissance - réseaux neuronaux - localisation d'onde - atténuation - milieux aléatoires.

OBSERVATOIRE DE GRENOBLE  
et  
LABORATOIRE DE GÉOPHYSIQUE INTERNE ET TECTONOPHYSIQUE

**THÈSE**

présentée par

**Mirko VAN DER BAAN**

pour obtenir le titre de DOCTEUR de

**L'UNIVERSITÉ JOSEPH FOURIER - GRENOBLE I**

(Arrêtés ministériels du 5 juillet 1984 et du 30 mars 1992)

Spécialité: Géophysique - Géochimie - Géomécanique

---

**Deux méthodes d'inférence statistique  
appliquées aux données de sismique réflexion profonde:  
Détection de signaux et localisation d'onde**

---

Date de soutenance: 22 novembre 1999

Composition du jury:

M. M. Campillo	Président	Professeur, Université de Grenoble
M. D. Gibert	Rapporteur	Professeur, Université de Rennes
M. R. Snieder	Rapporteur	Professeur, Université d'Utrecht
M. F. Glangeaud	Examineur	Directeur de recherche, CNRS, Grenoble
M. B. van Tiggelen	Examineur	Chargé de recherche, CNRS, Grenoble
Mme. A. Paul	Directeur de thèse	Chargé de recherche, CNRS, Grenoble

## Remer

En général,  
dernière tâche à  
complir, mais avec  
réflexion de la chose  
En effet, quelque  
fait plus de mal que

Le plus beau  
peut-être et même  
D'abord, le  
satisfaction de la  
s'agisse pas par  
deux à travailler  
tant qu'il a fait  
son bien-être de la  
Malheureusement,  
souvent de par

Enfin, je  
l'absence de  
pour toutes ces  
de l'avoir plus  
surtout de la  
accepte d'être

*"Wie nooit twijfelt, leert niets"*

*"Celui qui ne doute jamais, n'apprend rien"*

Dirk Volkertszoon Coornhert (1522-1590)

le contact de  
de tous les  
En effet, je  
souvent de la  
sur les réactions de  
en l'air et en  
particulier. Plus  
l'absence de  
de plus en plus  
le n'ont pas  
Malheureusement,  
souvent de la  
souvent de la  
souvent de la  
souvent de la

*"Beproeft en behoudt het goede"*

*"Mettez à l'épreuve et gardez ce qui est bon"*

Dicton hollandais



## Remerciements

En général, on est très content d'écrire les remerciements, puisqu'il s'agit vraiment de la dernière tâche à accomplir. En tous cas, je ne suis très content car non seulement ma thèse est rédigée, mais aussi et surtout parce que je l'ai déjà présentée. Cependant, je dirais que les préparations de la soutenance ont été (physiquement et mentalement) plus difficiles que la rédaction. En effet, quelques incidences informatiques (ce qui est normale) et sur la voie publique (ce qui fait plus de mal) ne m'ont pas facilité les choses.

Je tiens beaucoup à remercier plusieurs personnes pour ce qu'elles ont fait pour moi, avant, pendant et même après la thèse.

D'abord, je voudrais remercier Georges Poupinet pour avoir réglé toutes les affaires administratives afin que je puisse venir effectuer ma thèse au LGIT. Je regrette cependant, que nous n'ayons pas pu établir une intense collaboration. En revanche, j'espère vivement que nous réussirons à travailler ensemble dans un avenir proche. Puis, je voudrais aussi le remercier pour tout ce qu'il a fait pour moi dans la vie quotidienne. Par exemple, il y a trois ans, il m'a prêté son vélo afin de faciliter mes déplacements dans Grenoble. Le vélo m'a été vraiment très utile. Malheureusement, il a été un peu déformé lors de l'accident survenu lorsque je préparais ma soutenance. Je promets de le lui rendre dès qu'il sera réparé....

Ensuite, je tiens à remercier les membres du jury de ma thèse d'avoir accepté de juger le contenu scientifique de mon travail. En outre, je remercie Michel Campillo et François Glangeaud pour toutes nos discussions. Moreover, I would like to thank Roel Snieder and Bart van Tiggelen for having given a Dutch accent to a principally French environment. By the way, my Dutch abstract should be flawless by now. Enfin, je souhaite aussi remercier Dominique Gibert d'avoir accepté d'être rapporteur. Je suis désolé qu'il n'ait pas pu assister à ma soutenance. Finalement, je remercie Anne Paul d'avoir su tenir une tâche difficile: être mon directeur de thèse, et pour les corrections de mon français.

Je tiens également à remercier toutes les personnes qui m'ont aidé à rédiger mes articles. En outre, je suis reconnaissant à Christian Jutten pour avoir partagé ses connaissances sur le domaine des réseaux de neurones. Ceci a, en effet, permis d'aboutir à notre article de synthèse sur les réseaux de neurones. In addition, I'm very grateful to Sergei Shapiro for all discussions we have had concerning wave localization theory in general and his theoretical development in particular. Finally, I would like to thank Jeroen Goudswaard, Michiel Postema and my brother Roland for having corrected my fabulous Dutch abstract.

De plus, je souhaite remercier Jean-Noël Chardron, François Janod et Catherine Péquegnat: ils m'ont apporté une aide précieuse lors de la préparation de ma soutenance.

Finalement, il me reste à remercier tout le monde qui a fait de mon séjour en France un souvenir inoubliable. Parmi eux se trouvent bien sûr les thésards du labo, ceux de Belleville et mes copains d'escalade. Je leur dois mes connaissances du jargon français. De plus, grâce à ces grimpeurs farouches les mots 'démon' et 'abo' ne seront plus jamais les mêmes mots qu'avant (même si très longtemps je pensais qu'il s'agissait des mots 'à beau!').



Last but not least, I would like to express my gratitude to my parents and my family. Without their support this thesis wouldn't have been realized.

Maintenant, il ne me reste plus qu'à remercier Laetitia pour ses soins et inquiétudes lors de la rédaction de ma thèse et la préparation de ma soutenance et pour récrire mes remerciements. De plus, je suis très heureux qu'elle m'accompagne en Angleterre.

## Résumé en français

L'objectif de toute acquisition de données géophysiques est d'obtenir une image plus précise de l'intérieur de la Terre. Afin de préciser cette image, on utilise principalement des méthodes d'inférence déterministe ayant comme but de fournir une description exacte de chaque détail particulier. Cependant, ce type de méthodes est forcément limité par une résolution finie et, en général, une approche déterministe fait croître la complexité des mathématiques impliquées, puisqu'elle nécessite de décrire chaque détail et son influence exacte sur le processus considéré.

D'un autre côté, les approches statistiques n'essaient de saisir que les caractéristiques principales de l'ensemble des détails. C'est cette qualité qui réduit la complexité mathématique, nous permettant ainsi de mieux étudier certains phénomènes physiques. Ceci a pour conséquence additionnelle que les méthodes d'inférence statistique peuvent nous aider à sonder la Terre à une échelle plus petite que résolution des méthodes déterministes. Un troisième avantage des méthodes d'inférence statistique est qu'elles peuvent nous fournir une appréciation de vraisemblance, ce qui peut nous aider à prendre des décisions ou à reconnaître des signaux.

Dans ce mémoire, nous avons utilisé ces 3 avantages sur 2 problèmes différents. D'abord, nous les avons employés afin d'augmenter le rapport signal sur bruit en détectant et extrayant des réflexions dans des données de sismique réflexion profonde d'une faible qualité. Ceci nous a permis de faciliter l'interprétation du profil considéré en n'appliquant que de méthodes de reconnaissance automatique. Deux méthodes différentes ont été considérées, celle dite 'Classification par Vraisemblance Maximale' et celle des 'Réseaux de Neurones'. Bien que la première soit plus facile à mettre en œuvre, des signaux ayant des distributions statistiques sous-jacentes multi-modales ou asymétriques sont plus difficiles à gérer. Par contre, ce type de distribution ne pose aucun problème pour les Réseaux de Neurones, puisqu'ils peuvent approximer toute fonction continue et bornée de façon automatique, mais ces outils mathématiques sont plus difficiles à gérer.

Ensuite, nous avons employé les avantages d'une approche statistique afin d'étudier le processus de la diffusion forte dans un milieu acoustique, stratifié et aléatoire pour déterminer les caractéristiques statistiques de la croûte terrestre en utilisant à nouveau des données de sismique réflexion profonde. Afin d'étudier le processus de la diffusion forte dans de tels milieux, nous avons appliqué la théorie de la localisation d'onde qui nous prédit l'atténuation apparente et la dispersion d'une impulsion traversant un milieu chaotique. De plus, elle fait appel à des quantités auto-moyennantes, afin que nous puissions comparer les quantités prédites et observées dans des simulations numériques. Elle nous a permis d'estimer la magnitude  $\sigma_c$  et la dimension  $a$  typiques des fluctuations des vitesses et le facteur de qualité de l'atténuation intrinsèque  $Q_{in}$  sous le Bassin de Mauléon dans les Pyrénées Occidentales. La croûte supérieure y est caractérisée par  $a \sim 15$  m,  $\sigma_c \leq 7.5$  % et  $Q_{in} \geq 350$  et la croûte inférieure par  $a \sim 40$  à  $50$  m,  $\sigma_c \leq 5.5$  % et  $Q_{in} \geq 1600$ . Ceci nous montre que, dans cette région, la croûte supérieure est plus hétérogène et plus atténuante que la croûte inférieure et que les dimensions typiques y sont plus petites.



## Abstract in english

The objective of the acquisition of geophysical data is to obtain a finer image of the Earth's interior. To refine this image, mainly deterministic inference methods are used. Such methods try to describe precisely each single detail. Hence, this type of method has to be limited by a finite resolution. Moreover, generally, they tend to increase the complexity of the involved mathematics as each single detail and its influence on the considered phenomenon have to be described.

Statistical approaches, on the other hand, only try to grasp the principal characteristics of the details in their totality. Due to this particular feature, mathematical complexity may be significantly reduced, thereby allowing for a better study of certain physical processes. As an additional consequence, statistical inference methods may help us to study the Earth on a scale beyond the resolution of their deterministic equivalents. A third advantage of statistical inference methods is that they can produce an assessment of likelihood, which can be helpful to take decisions or to recognize signals.

In this thesis, these 3 advantages have been used in 2 different ways. First, they have been used to enhance the signal-to-noise ratio in noisy deep seismic reflection data by detecting and extracting present reflections. Hence, interpretation of the considered profil has been facilitated by applying automatic recognition techniques only. Two different recognition methods have been employed, namely 'Maximum Likelihood Classification' and 'Neural Networks'. The first method is simple to use, but signals characterized by multimodal or skewed distributions are less easily handled. Such types of distributions do not pose a problem for Neural Networks as they can approximate automatically any finite, continuous function. However, they are more difficult to use.

Next, the advantages of statistical inference methods have been employed to study the process of strong scattering in random, stratified, acoustic media and to determine the statistical characteristics of the Earth's crust using again deep seismic reflection data. To study the phenomenon of strong scattering, the so-called wave localization theory has been used which describes the apparent attenuation and dispersion of a pulse traversing such media. Moreover, this theory makes use of self-averaging quantities, thereby allowing for a fair comparison of theoretical predictions and observed quantities in numerical simulations. In addition, with its help, the typical magnitude  $\sigma_c$  and scale-length  $a$  of the velocity fluctuations and the quality factor of intrinsic attenuation  $Q_{in}$  could be established underneath the Mauléon Bassin in the Western Pyrenees. In that region, the upper crust is characterized by  $a \sim 15$  m,  $\sigma_c \leq 7.5$  % and  $Q_{in} \geq 350$  and the lower crust by  $a \sim 40$  to  $50$  m,  $\sigma_c \leq 5.5$  % and  $Q_{in} \geq 1600$ . This shows us that the upper crust is more heterogeneous and more attenuative than the lower crust and that the typical scale-lengths increase with depth in this region.

## Samenvatting in het Nederlands

Het doel van het verzamelen van geofysische data is het verkrijgen van een beter beeld van het innere der aarde. Om dit beeld te verduidelijken worden voornamelijk deterministische deductiemethoden gebruikt. Het doel van dit soort methoden is om een exacte beschrijving van elk afzonderlijk detail te verkrijgen. Dit leidt ongelukkigerwijs tot een eindige resolutie. Bovendien kan een deterministische aanpak leiden tot een niet ongeringe wiskundige complexiteit, aangezien elk afzonderlijk detail en zijn invloed op het onderzochte verschijnsel beschreven moeten worden.

Een statistische aanpak daarentegen probeert alleen de voornaamste kenmerken van al deze afzonderlijke details te bepalen. Dankzij deze eigenschap kan de wiskundige complexiteit verminderd worden. Dit geeft de mogelijkheid om bepaalde natuurkundige verschijnselen beter te bestuderen. Verder kunnen statistische methodieken ons helpen om de aarde te bestuderen op een schaal die kleiner is dan het oplossend vermogen van deterministische methodieken. Een derde voordeel is het feit dat statistische methodieken een waarschijnlijkheidswaardering opleveren kunnen. Hierdoor wordt het nemen van beslissingen en het herkennen van signalen vereenvoudigd.

In dit proefschrift zijn deze drie voordelen op twee verschillende manieren toegepast. Allereerst zijn ze gebruikt om de signaal/ruis verhouding te verbeteren door middel van het ontdekken en onttrekken van reflecties uit data van diep-seismische experimenten, die gekenmerkt worden door een hoog ruisniveau. Dit leidt tot een eenvoudigere interpretatie van deze data, terwijl slechts automatische herkenningmethoden gebruikt worden. Twee verschillende technieken zijn nader bekeken, te weten 'Classificatie via Maximale Waarschijnlijkheid' en 'Neurale Netwerken'. De eerste techniek is makkelijk te implementeren en toe te passen, maar zijn signalen die gekenmerkt worden door multimodale of asymmetrische verdelingen lastiger te analyseren. Neurale Netwerken hebben geen enkel probleem met dit soort verdelingen, maar zijn in praktijk veel lastiger te hanteren.

Vervolgens zijn de voordelen van een statistische aanpak gebruikt om het fenomeen van de sterke verstrooiing in chaotische, akoestische, gelaagde media te bestuderen en de statistische eigenschappen van de aardkorst te bepalen; opnieuw met behulp van diep-seismische data. Om het fenomeen van de sterke verstrooiing in dergelijke media te kunnen bestuderen is gebruik gemaakt van de zogenaamde golf localisatie theorie. Deze theorie beschrijft de schijnbare uitdoving en dispersie van een puls die een chaotisch medium doorloopt. Bovendien gebruikt deze theorie zelf-middelende variabelen die een nauwkeurige vergelijking van theoretische voorspellingen en observaties in numerieke simulaties toestaat. Dankzij deze theorie kon de typische magnitude  $\sigma_c$  en schaallengte  $a$  van de fluctuaties van de akoestische snelheden en de kwaliteitsfactor van de intrinsieke uitdoving  $Q_{in}$  bepaald worden onder het Mauléon bekken in de Westelijke Pyreneeën. In dit gebied wordt de bovenkorst gekenmerkt door  $a \sim 15$  m,  $\sigma_c \leq 7.5$  % en  $Q_{in} \geq 350$  en de onderkorst door  $a \sim 40$  à  $50$  m,  $\sigma_c \leq 5.5$  % en  $Q_{in} \geq 1600$ . Dit toont aan dat de bovenkorst hier heterogener is en sterker uitdovend werkt dan de onderkorst en dat de typische schaallengte in dit gebied toeneemt met de diepte.



## Table des matières

Remerciements	v
Résumé en français	vii
Abstract in english	viii
Samenvatting in het nederlands	ix
<b>I Introduction générale</b>	<b>3</b>
<b>II Détection de signaux</b>	<b>7</b>
Résumé . . . . .	7
II.1 Introduction générale . . . . .	9
II.1.1 La reconnaissance supervisée . . . . .	9
II.1.2 Présentation du chapitre . . . . .	13
II.2 L'expérience Ecors 2.5-D dans les Pyrénées . . . . .	15
Résumé . . . . .	15
II.2.1 Eléments sur la géologie des Pyrénées . . . . .	15
II.2.2 Eléments sur le dispositif Ecors 2.5-D . . . . .	16
II.3 Classification par vraisemblance maximale . . . . .	21
Résumé . . . . .	21
<b>Recognition and reconstruction of coherent energy with application to         deep seismic reflection data . . . . .</b>	<b>23</b>
Abstract . . . . .	23
II.3.1 Introduction . . . . .	23
II.3.2 Recognition method . . . . .	25
II.3.3 Reconstruction method . . . . .	29
II.3.4 Data . . . . .	31
II.3.5 Data application: recognition . . . . .	31
II.3.6 Data application: reconstruction . . . . .	35
II.3.7 Data application: stacked sections . . . . .	37
II.3.8 Discussion . . . . .	39
II.3.9 Conclusions . . . . .	41
II.3.10 Acknowledgments . . . . .	42
II.4 Réseaux neuronaux . . . . .	43
Résumé . . . . .	43
II.4.1 Les Réseaux de neurones et la détection des signaux: Introduction . . . . .	43

**On the use of neural networks in geophysical applications** . . . . . 45

Abstract . . . . . 45

II.4.2 Introduction . . . . . 45

II.4.3 Neural Networks: Structure and behavior . . . . . 46

II.4.4 Network optimization . . . . . 50

II.4.5 Practice . . . . . 61

II.4.6 Discussion and Conclusions . . . . . 66

II.4.7 Acknowledgments . . . . . 67

II.4.8 Les réseaux de neurones et la détection de signaux: Conclusions . . . . . 68

II.5 Transformée Continue en Ondelette . . . . . 69

Résumé . . . . . 69

II.5.1 La Transformée Continue en Ondelette et la détection des signaux:  
Introduction . . . . . 69

**Recognition methods for S/N enhancement: A comparison** . . . . . 71

Abstract . . . . . 71

II.5.2 Introduction . . . . . 71

II.5.3 Maximum Likelihood Classification . . . . . 73

II.5.4 Wavelets . . . . . 73

II.5.5 A comparison . . . . . 79

II.5.6 Discussion . . . . . 81

II.5.7 Conclusions . . . . . 83

II.5.8 Acknowledgments . . . . . 84

II.5.9 Extraction simultanée de plusieurs types de signaux . . . . . 85

II.5.10 La Transformée Continue en Ondelette et la détection des signaux:  
Conclusions . . . . . 90

II.6 Les données Ecors 2.5-D sommées et migrées . . . . . 91

Résumé . . . . . 91

II.6.1 Introduction . . . . . 91

II.6.2 Une comparaison des résultats obtenus . . . . . 91

II.6.3 Discussion . . . . . 101

II.6.4 Conclusions . . . . . 103

II.7 Conclusions . . . . . 105

**III Localisation d'onde** . . . . . 109

III.1 Introduction . . . . . 111

III.2 Propagation d'ondes en milieu aléatoire . . . . . 113

Résumé . . . . . 113

**Acoustic wave propagation in 1-D random media: The wave localization  
approach** . . . . . 115

Abstract . . . . . 115

III.2.1 Introduction . . . . . 115

III.2.2 Theory . . . . . 118

III.2.3 Numerical simulations . . . . . 135

III.2.4 Discussion . . . . . 142

III.2.5 Conclusions . . . . . 144

III.2.6 Acknowledgements . . . . . 146

III.3 Essais d'inversions des caractéristiques statistiques d'un milieu . . . . . 147

Résumé . . . . . 147

III.3.1 Introduction . . . . . 147

III.3.2 Les facteurs influençant l'atténuation apparente . . . . . 148

III.3.3 Méthodes d'inversion . . . . . 153

III.3.4 Inversion de données synthétiques . . . . . 154

III.3.5 Inversions de données réelles . . . . . 172

III.3.6 Discussion . . . . . 177

III.3.7 Conclusions . . . . . 184

III.4 Conclusions . . . . . 187

III.A Méthode d'inversion locale . . . . . 189

III.B Méthode d'inversion globale . . . . . 193

III.B.1 Inversions Bayésiennes . . . . . 193

III.B.2 L'algorithme Vegas: Echantillonnage par importance . . . . . 194

**IV Conclusions finales** . . . . . 199

**Références** . . . . . 203



## Liste des abbréviations

<i>CSG</i>	<i>Common Shot Gather</i>
<i>CVM</i>	<i>Classification par Vraisemblance Maximale</i>
<i>CMP</i>	<i>Common MidPoint</i>
<i>DDA</i>	<i>Dynamic Decay Adjustment</i>
<i>FNP</i>	<i>Faille Nord Pyrénéenne</i>
<i>MLC</i>	<i>Maximum Likelihood Classification</i>
<i>MLP</i>	<i>Multi-Layer Perceptron</i>
<i>NCU</i>	<i>Non-Contributing Units</i>
<i>NN</i>	<i>Neural Networks</i>
<i>NPZ</i>	<i>Nord Pyrenean Zone</i>
<i>OBD</i>	<i>Optimal Brain Damage</i>
<i>OBS</i>	<i>Optimal Brain Surgeon</i>
<i>PCA</i>	<i>Principal Component Analysis</i>
<i>pdf</i>	<i>probability density function</i>
<i>RBF</i>	<i>Radial Basis Functions</i>
<i>RN</i>	<i>Réseaux de Neurones</i>
<i>SCG</i>	<i>Scaled Conjugate Gradient</i>
<i>STFT</i>	<i>Short Term Fourier Transform</i>
<i>SVD</i>	<i>Singular Value Decomposition</i>
<i>TCO</i>	<i>Transformée Continue en Ondelette</i>
<i>TFCT</i>	<i>Transformée Fourier à Court Terme</i>
<i>WT</i>	<i>Wavelet Transform</i>

## Table des figures

II.1.1	Méthode de classification par vraisemblance maximale. . . . .	11
II.1.2	Un réseau de neurones de base. . . . .	11
II.2.1	Les localisations des profils Ecors. 1 - Pyrénées Est, 2 - Ecors Arzacq et 3 - Ecors 2.5-D. D'après Siliqi (1994). . . . .	15
II.2.2	Les coupes-sommes migrées et interprétées des 3 profils Ecors Pyrénées. a) Pyrénées Est, b) Ecors Arzacq et c) l'interprétation du profil Ecors 2.5-D en traits gras) superposée sur b). La structure en éventail n'est visible que sur le premier profil, car les deux autres profils se trouvent au nord de la Zone Axiale (figure II.2.1). Par contre, le premier et le second profil montrent bien la différence de profondeur du Moho. D'après Roure et al. (1989), Daignières et al. (1994) et Lecerf (1996). . . . .	17
II.2.3	La localisation exacte du profil Ecors 2.5-D dans la géologie régionale. CSS - Chevauchement de Sainte-Suzanne; CSP - Chevauchement de Saint-Palais; CFNP - Chevauchement Frontal Nord Pyrénéen. Figure extraite de Lecerf (1996). . . . .	18
II.2.4	Dispositif du profil Ecors 2.5-D. Figure extraite de Marthelot et al. (1994). . . . .	19
II.3.1	CSG 8184. a) Raw data. b) Zoom-in showing the positions of the used picks (drawn line). . . . .	32
II.3.2	CSG 8184. a) Likelihood distribution of the presence of reflection energy using a principal component analysis to reduce the number of attributes. b) Extracted data. . . . .	33
II.3.3	CSG 8184. a) Likelihood distribution of reflection energy after incorporation of surrounding probabilities. b) Extracted data. . . . .	34
II.3.4	CSG 8184. Reconstructed data with semblance weighting after two passes (with 5, resp. 10 traces). Using a) raw data (figure 1a) and b) extracted data (figure 2 b). . . . .	35
II.3.5	CSG 8184. Reconstructed data using selection in $\tau$ - $p$ domain. One pass: 5 traces. . . . .	36
II.3.6	Stack. a) Raw data. b) Associated reconstruction using semblance weighting. . . . .	38
II.3.7	Stack. a) Pre-stack recognition and reconstruction on CSGs and a second application of recognition after stack. b) Associated reconstruction using semblance weighting. . . . .	39
II.4.1	The biological and the mathematical neuron. The mathematical neuron (b) mimics the behavior of the biological neuron (a). The weighted sum of the inputs are rescaled by an activation function (c) of which several examples are shown in (d). Adapted from Lippmann (1987), Hérault and Jutten (1994) and Romeo (1994). . . . .	47
II.4.2	a) Single perceptron layer and b) associated decision boundary. Adapted from Romeo (1994). . . . .	48
II.4.3	a) Single hidden perceptron layer and b) associated decision boundary. Adapted from Romeo (1994). . . . .	49



II.4.4	Generalization versus training error. Adapted from Moody (1994). . . . .	54
II.4.5	Common shot gather + 31 pick positions. a) Original data, b) 14 reflection picks, c) 3 pre-arrival noise + 6 ground roll picks, d) 6 picks on background noise + bad traces and e) 2 more picks on bad traces. . . . .	62
II.4.6	Extraction results using Gaussian statistics and a PCA for a) reflections, b) ground roll and c) other types of noise. To be compared with Figure II.4.5.a. . . . .	64
II.4.7	Extraction results using an MLP network for again a) reflections, b) ground roll and c) other types of noise. Notice the improvement of extraction results for the remaining types of noise. . . . .	65
II.5.1	Flow diagram of the MLC-technique using the STFT-technique for signal characterization. The first three steps are typical for any classification algorithm, whereas the fourth is required for signal extraction. . . . .	72
II.5.2	Trade-off of time versus frequency resolution for both an STFT and a WT. The STFT-technique employs a constant window length. Thus, a) a long period and b) a short period signal will have an identical resolution (width of peak) in frequency domain, whereas time and frequency resolution are inversely proportional for c) large scale and d) small scale analyzing wavelets. Hence, small scale wavelets have an excellent time, but nearly no frequency resolution, whereas large scale wavelets show an opposite behavior. The STFT-technique uses a Gaussian tapering window for this example which makes a) nearly identical to c), the imaginary part of a particular Slepian mother wavelet. After Bear and Pavlis (1997). . . . .	74
II.5.3	First 4 used Slepian wavelets in both time and frequency domain with $p = 3.0$ and $p_c = 5.0$ , resulting in time-concentrated wavelets. After Lilly and Park (1995). . . . .	77
II.5.4	Flow diagram for WT. To be compared with Figure II.5.1. . . . .	78
II.5.5	a) The employed common shot gather for testing the different recognition methods and b) the 14 picks of the Moho reflection. Identical to Figure II.3.1 (reproduced for convenience). . . . .	79
II.5.6	Extraction results of the MLC-technique using an STFT analysis with a) the scaled likelihood distribution and b) a principal component analysis. To be compared with the original CSG displayed in figure II.5.5.a. . . . .	80
II.5.7	As Figure II.5.6.a, but now using the first four Slepian wavelets (as shown in Figure II.5.3). . . . .	82
II.5.8	Résultats d'extraction en utilisant la méthode CVM et une TCO. A comparer avec la figure II.4.6 où une TFCT a été utilisé. a) Réflexions, b) ondes de surface et c) autres types de bruit. . . . .	86
II.5.9	Résultats d'extraction en utilisant un RN et une TCO. A comparer avec la figure II.4.7 où une TFCT a été utilisé. a) Réflexions, b) ondes de surface et c) autres types de bruit. . . . .	87
II.5.10	Résultats d'extraction de la méthode CVM et une TFCT pour une fenêtre d'analyse de 256 ms. Extraction des réflexions en utilisant a) une distribution Gaussienne ou b) une analyse de composants principaux. A comparer avec figure II.4.6.a. . . . .	88
II.5.11	Identique à la figure II.5.10. Extraction des ondes de surface par la méthode CVM et une TFCT pour une fenêtre d'analyse de 256 ms. a) Distribution Gaussienne et b) analyse de composants principaux. A comparer avec la figure II.4.6.b. . . . .	89

II.6.1	Partie supérieure de 2 lignes voisines de la coupe-somme des données Ecors 2.5-D. Résultats obtenus a) par nous et b) par Siliqi (1994). . . . .	93
II.6.2	Partie profonde de la ligne 132 de la coupe-somme obtenue a) par nous et b) par Siliqi (1994). . . . .	96
II.6.3	Migration à 3 dimensions des segments pointés d'après Siliqi (1994). a) Superposition de tous les segments pointés le long des lignes. b) Superposition de ces segments après migration le long des lignes. c) Superposition des segments restants après migration le long des colonnes et le troisième pointé. . . . .	98
II.6.4	Notre coupe-somme migrée. a) Coupe-somme finale résultant de la superposition des lignes 128 à 138 de la coupe-somme des données Ecors 2.5-D. b) Migration de la coupe-somme finale par la méthode de Gazdag (1978). . . . .	100
III.2.1	Localization length, apparent attenuation and dispersion as a function of $k'_0 a$ . Vertical scales have been normalized with $\langle \sigma_\kappa^2 \rangle / \cos^4(\phi_i)$ . In addition, units of the localization length are expressed in typical scale-lengths $a$ . The constant $\Gamma(0)$ in the Von Kármán function for $\nu = 0$ has been replaced by the constant 9.0 to display the general trend. Notice the strong dependence of all quantities on the angle of incidence, indicating e.g. the presence of anisotropy due to fine-scale layering. . . . .	128
III.2.2	Top parts of the used velocity profiles for a single realization of Gaussian, exponential and Von Kármán ( $\nu = 0$ ) autocorrelation functions. . . . .	138
III.2.3	Primaries for the random media shown in Figure III.2.2. Every second trace shows the case of $\phi_i = 30^\circ$ . First 2 traces display the homogeneous case, next 6 traces $\sigma_\kappa = 15\%$ and the last 6 traces $\sigma_\kappa = 30\%$ . Notice different scale on last 5 traces. G stands for Gaussian, E for exponential and VK for Von Kármán autocorrelation functions. . . . .	139
III.2.4	Measured localization length, apparent scattering coefficient and dispersion for 4 realizations of the 3 types of used correlation functions for both angles of incidence for $\sigma_\kappa = 15\%$ . Theoretical predictions are plotted too. The lowest theoretical curve always presents $\phi_i = 0$ and the other one $\phi_i = 30^\circ$ , except for the localization length where the order is inversed. The blocky character of $\delta U$ is due to the used sampling rate (1 ms). Notice, however, that the values of $\delta c$ have been calculated with a precision higher than the sampling rate. . . . .	140
III.2.5	Identical to Figure III.2.4 only now for $\sigma_\kappa = 30\%$ . Group and phase velocity measurements have not been displayed for $\phi_i = 30^\circ$ as phase unwrapping became completely unstable and the primary could not be automatically detected (main energy arrived after the primary). Notice changed vertical scales with respect to Figure III.2.4. . . . .	141
III.3.1	Sismogrammes synthétiques obtenus pour une réalisation particulière du milieu avec $\sigma_\kappa = 5\%$ . Chaque ensemble de 2 traces est composé d'un sismogramme pour un milieu homogène et ensuite pour un milieu hétérogène. Les ensembles correspondent de gauche à droite à un cas en transmission et plusieurs cas en réflexion où le coefficient de réflexion baisse (d'un réflecteur parfait jusqu'à $\hat{r}_{cr} = 0.20$ ). . . . .	156
III.3.2	Mesures de la longueur de localisation des sismogrammes montrés sur la figure III.3.1. . . . .	157
III.3.3	Mesures du rapport spectral pour $\sigma_\kappa = 5\%$ et a) un réflecteur parfait et b) $\hat{r}_{cr} = 0.20$ . . . . .	158



III.3.4	Comme la figure précédente, sauf que $\sigma_{\kappa} = 15\%$ . a) réflecteur parfait. b) $\hat{r}_{cr} = 0.20$ . . . . .	158
III.3.5	Fonctions marginales individuelles de la densité de probabilité pour $\sigma_{\kappa} = 15\%$ et un réflecteur parfait (figure III.3.4.a). L'«Ecart-type» est $\sigma_{\kappa}$ et les flèches indiquent les vraies valeurs des paramètres. . . . .	159
III.3.6	Fonctions marginales combinées de la densité de probabilité pour $\sigma_{\kappa} = 15\%$ et un réflecteur parfait (figure III.3.4.a). Les triangles blancs indiquent les vraies valeurs des paramètres. . . . .	160
III.3.7	Fonctions marginales individuelles de la densité de probabilité pour $\sigma_{\kappa} = 15\%$ et un réflecteur imparfait ( $\hat{r}_{cr} = 0.20$ , figure III.3.4.b). A comparer avec la figure III.3.5. . . . .	161
III.3.8	Fonctions marginales combinées de la densité de probabilité pour $\sigma_{\kappa} = 15\%$ et un réflecteur imparfait ( $\hat{r}_{cr} = 0.20$ , figure III.3.4.b). A comparer avec la figure III.3.6. . . . .	162
III.3.9	Fonctions marginales individuelles de la densité de probabilité pour $\sigma_{\kappa} = 15\%$ et respectivement un réflecteur parfait (reproduites de la figure III.3.5), un réflecteur parfait avec des erreurs doublées et quadruplées sur les mesures du rapport spectral d'un réflecteur parfait, puis le même rapport spectral mais avec les erreurs observées d'un réflecteur imparfait ( $\hat{r}_{cr} = 0.20$ , figure III.3.4.b) et enfin les résultats de $\hat{r}_{cr} = 0.20$ (reproduits de la figure III.3.7). . . . .	163
III.3.10	Fonctions marginales combinées de la densité de probabilité pour $\sigma_{\kappa} = 15\%$ et le rapport spectral d'un réflecteur parfait mais ayant des erreurs doublées. A comparer avec les figures III.3.6 et III.3.8. . . . .	164
III.3.11	Fonctions marginales combinées de la densité de probabilité pour $\sigma_{\kappa} = 15\%$ et le rapport spectral d'un réflecteur parfait mais ayant des erreurs quadruplées. A comparer avec les figures III.3.6, III.3.8 et III.3.10. . . . .	165
III.3.12	Spectres de bruit mesurés et prédits par les théories pour un réflecteur parfait. a) $\sigma_{\kappa} = 5\%$ et b) $\sigma_{\kappa} = 15\%$ . Notez les échelles verticales différentes. . . . .	167
III.3.13	Rapport spectral après corrections pour $\hat{r}_{cr} = 0.20$ . a) $\sigma_{\kappa} = 5\%$ et b) $\sigma_{\kappa} = 15\%$ . Notez les échelles verticales différentes. . . . .	168
III.3.14	Le rapport spectral mesuré (sans correction) après lissage pour $\hat{r}_{cr} = 0.20$ . a) $\sigma_{\kappa} = 5\%$ et b) $\sigma_{\kappa} = 15\%$ . A comparer avec les figures III.3.3.b et III.3.4.b qui sont présentées à la même échelle. Les moyennes sans lissage sont aussi montrées. . . . .	169
III.3.15	Rapport spectral de bruit mesuré. a) $\sigma_{\kappa} = 5\%$ et b) $\sigma_{\kappa} = 15\%$ . A comparer avec la figure III.3.12 qui est montrée à la même échelle. Les moyennes sans lissage sont aussi montrées. . . . .	169
III.3.16	Rapports spectraux corrigés pour $\hat{r}_{cr} = 0.20$ après lissage. a) $\sigma_{\kappa} = 5\%$ et b) $\sigma_{\kappa} = 15\%$ . A comparer à la figure III.3.13. . . . .	170
III.3.17	Résultats de l'inversion du rapport spectral mesuré pour $\hat{r}_{cr} = 0.20$ . a) $\sigma_{\kappa} = 5\%$ et b) $\sigma_{\kappa} = 15\%$ . . . . .	171
III.3.18	a) Collection de traces à point à tir commun 8184 et positions des pointés sur b) les premières arrivées, c) une réflexion dans la croûte supérieure et d) le Moho. . . . .	174
III.3.19	Identique à la figure III.3.18, mais pour le groupe de traces 8194. . . . .	175
III.3.20	Spectres mesurés (a et c) et rapports spectraux (b et d) pour les collections de traces 8184 (a et b) et 8194 (c et d). . . . .	176

III.3.21	Résultats des inversions pour la collection de trace 8184. a) et b) Rapports spectraux Moho sur premières arrivées, c) et d) rapports spectraux réflexion majeure sur premières arrivées et e) et f) rapports spectraux Moho sur réflexion majeure. Nous dénotons par Q le modèle n'ayant que de l'atténuation intrinsèque, D le modèle n'ayant que des diffractions multiples et D + Q les deux effets combinés. . . . .	178
III.3.22	Identique à la figure III.3.21, mais pour la collection de traces 8194. . . . .	179



## Liste des tableaux

III.2.1	Medium parameters with $L$ : thickness of random layer, $z_{source}$ : source depth, $c_0$ : average velocity, $\rho_0$ : density, $\Delta z$ : discrete layer thickness, $\Delta t$ : sample time, $f_{p,Ricker}$ : peak frequency of Ricker source wavelet. . . . .	135
III.3.1	Facteurs influençant les amplitudes et leurs abréviations . . . . .	149
III.3.2	Paramètres constants des simulations numériques. $L$ : épaisseur du milieu, $z_{source}$ : profondeur de la source, $c_0$ : vitesse moyenne, $\rho_0$ : densité, $\Delta z$ : épaisseur des couches, $\Delta t$ : pas d'échantillonnage, $f_{p,Ricker}$ : fréquence principale de l'ondelette de Ricker, $a$ : échelle caractéristique des hétérogénéités, $\sigma_k$ : magnitude caractéristique des hétérogénéités, $\phi_i$ : angle d'incidence, type: type de fonction d'autocorrélation et $Q_{in}$ : $Q$ intrinsèque. . . . .	154
III.3.3	Résultats des inversions des données synthétiques. . . . .	172
III.3.4	Résultats des inversions des données réelles des collections de traces 8184 et 8194. Les premières arrivées sont dénotées par l'étiquette 'src' et la réflexion majeure dans les sédiments par 'refl'. . . . .	180
III.3.4	Suite. . . . .	181
III.3.5	Moyennes des résultats des inversions des données réelles. . . . .	181

Introduction générale

## Chapitre I

### Introduction générale

L'objectif de toute acquisition de données géophysiques est d'obtenir une image plus précise de l'intérieur de la Terre. Afin de préciser cette image, on utilise principalement des méthodes d'inférence de type déterministe. Ce type de technique a comme objectif de fournir une description exacte de chaque détail particulier.

Par conséquent, la résolution de ces techniques est forcément limitée, puisqu'elles traitent chaque détail de façon individuelle. De plus, la complexité mathématique dépend souvent de la gamme d'échelles sur laquelle nous étudions le problème considéré. Ceci fait qu'elle est proportionnelle au nombre de détails. Le premier inconvénient nous empêche de décrire et de sonder la Terre à toutes les échelles, tandis que le second fait qu'il est très compliqué d'étudier et de comprendre certains processus physiques caractérisés par une certaine rigueur mathématique, puisqu'il nous faut décrire chaque détail et son influence sur le processus considéré.

Par contre, les méthodes statistiques ne dépendent que des caractéristiques principales de l'ensemble des détails. Ceci a pour conséquence qu'elles peuvent sonder la Terre sur une échelle plus petite augmentant ainsi la résolution possible. De plus, elles nous permettent de réduire la complexité des mathématiques impliquées et ainsi de mieux comprendre certains phénomènes physiques. Donc, une approche statistique peut nous aider à décrire et découvrir la structure de la Terre à une échelle plus petite que la résolution des méthodes d'inférence déterministe et, de plus, à mieux comprendre certains phénomènes physiques compliqués. Deux bons exemples de ce type de problèmes sont la résolution limitée de la tomographie sismologique et la complexité des mathématiques impliquées de la diffusion forte.

Un autre inconvénient des méthodes déterministes est qu'elles ne peuvent pas nous aider à prendre des décisions, puisqu'elles ne peuvent pas apprécier des probabilités ou nous fournir des vraisemblances. Donc, d'après elles, une décision est bonne ou mauvaise (noir ou blanc), mais des nuances (des teintes de gris) n'existent pas. Un bon exemple de ce type de problème est la détection de signaux ou des formes.

Contrairement aux méthodes déterministes, les approches statistiques n'ont pas ces inconvénients. Par conséquent, elles sont des alternatives intéressantes pour des applications géophysiques caractérisées par ces problèmes.

Dans cette thèse, nous illustrons quelques possibilités d'une approche statistique en considérant deux problèmes géophysiques différents, à savoir la détection de signaux et en particulier la détection automatique des réflexions dans une expérience de sismique réflexion profonde et la propagation d'onde dans des milieux aléatoires et la diffusion forte résultante.

Une approche statistique nous permet de détecter et d'extraire certains signaux en indiquant la vraisemblance de la présence de ce signal. Donc, il nous est maintenant possible de répondre à la question: 'S'agit-il d'une réflexion ou non?' en donnant la vraisemblance de sa présence



et la probabilité pour qu'il s'agisse d'un autre type de signal. Les vraisemblances peuvent être calculées par des méthodes dites de 'Classification par Vraisemblance Maximale' ou par des 'Réseaux de Neurones'.

Quant à la deuxième application considérée, la propagation d'onde dans des milieux aléatoires et la diffusion forte, une approche statistique nous permet de réduire considérablement le nombre de paramètres nécessaires. Par conséquent, les équations mathématiques décrivant les diffractions multiples sont fortement simplifiées. Ceci nous permet d'augmenter nos connaissances sur ce problème et de mieux comprendre le processus de la diffusion forte et les conséquences de la présence de certains types d'hétérogénéités. Afin de bien décrire l'influence des diffractions multiples sur l'atténuation et la dispersion d'une onde traversant un milieu aléatoire, nous utilisons la théorie de la localisation d'onde. Cette théorie a été développée dans le domaine de la mécanique quantique il y a 41 ans par Anderson et elle n'est connue que depuis une quinzaine d'années en physique acoustique.

Dans le **chapitre II**, nous décrivons la théorie de la détection de signaux et nous montrons comment elle peut être utilisée pour détecter et extraire de façon automatique les réflexions dans des données de sismique réflexion profonde. Les vraisemblances seront calculées en utilisant plusieurs méthodes et les signaux seront caractérisés par des transformations différentes.

Le **chapitre III** contient une description détaillée de la théorie de la localisation d'onde et de ses implications physiques. De plus, nous décrivons comment cette théorie pourrait être utilisée pour déterminer par inversion les caractéristiques statistiques d'un milieu stratifié et en particulier de la croûte terrestre en utilisant à nouveau des données de sismique réflexion profonde.

## Détection de signaux



## Chapitre II

# Détection de signaux

### Résumé

Les données d'expériences de sismique réflexion profonde n'ont souvent qu'une faible qualité. La raison principale est leur grande dimension d'acquisition, qui fait que la couverture multiple est très limitée et que l'influence des hétérogénéités locales augmente d'une façon considérable. Cependant, la qualité des données peut être améliorée en enlevant toute énergie non-cohérente et non-désirée. Ceci requiert une méthode de détection de signaux.

Dans cette étude, nous utilisons 2 méthodes de détection afin d'améliorer la qualité des données et de faciliter ainsi l'interprétation du profil considéré, à savoir la méthode de Classification par Vraisemblance Maximale et les réseaux de neurones. L'avantage de la première technique est qu'elle est facile à mettre en pratique, mais des problèmes ayant des distributions multi-modales et/ou asymétriques sont difficiles à gérer. Les réseaux de neurones sont plus difficiles à mettre en œuvre, mais des distributions multi-modales et/ou asymétriques ne posent aucun problème, car ces outils mathématiques peuvent approximer toute fonction continue et bornée de façon automatique.

Le choix des attributs utilisés est très important afin de caractériser au mieux les différents types de signaux. Nous recommandons d'utiliser la Transformée de Fourier à Court Terme, si l'extraction d'un seul type de signal est désirée. Autrement, une Transformée Continue en Ondelette est plus appropriée, car elle fait une analyse d'échelles des données et aucune spécification *a priori* de la longueur optimale de la fenêtre d'analyse n'est nécessaire.

L'inconvénient principal des techniques de détection est qu'elles ne peuvent reconstruire aucune énergie 'manquante'. Par conséquent, l'usage d'une technique de reconstruction est fortement conseillée après la détection des signaux. Surtout une implémentation à 3 dimensions produit de bonnes images de la géologie.

Nous avons appliqué toutes les méthodes de détection et de reconstruction développées sur les données de l'expérience de sismique réflexion profonde Ecors 2.5-D dans les Pyrénées occidentales. C'est un jeu de données difficile, car il a une faible qualité. De plus, les Pyrénées ont une géologie et tectonique très compliquées. Néanmoins, nos méthodes ont amélioré cette faible qualité et nous avons obtenu des bons résultats qui sont beaucoup plus faciles à interpréter.

## II.1 Introduction générale

### II.1.1 La reconnaissance supervisée

**Qualité des données** Dans ce chapitre nous étudions un problème couramment rencontré dans les profils de sismique réflexion profonde. Un faible rapport signal sur bruit rend souvent ces profils difficiles à interpréter. Ainsi, même les principales structures géologiques ne sont pas imagées avec fiabilité. Cela peut amener à des problèmes d'interprétation géologique et tectonique du profil.

Cette faible qualité des données est certainement évidente si nous faisons une comparaison avec les qualités des données régulièrement trouvées dans la sismique pétrolière. Plusieurs phénomènes contribuent à cette qualité inférieure. Premièrement, les couvertures multiples sont largement inférieures à celles de la sismique de prospection. Deuxièmement, la possibilité de l'existence de structures géologiques à 3 dimensions est souvent négligée dans les profils de sismique réflexion profonde. Troisièmement, ces profils s'intéressent surtout aux structures majeures de la croûte entière. Ceci pose plusieurs problèmes, puisque les enregistrements sont beaucoup plus longs et les distances entre les sources et les récepteurs atteignent plusieurs kilomètres. Même l'énergie émise par des charges jusqu'à 30 kilos d'explosifs a du mal à mettre en évidence des structures à plus de 10 km de profondeur. De plus, à cause de ces grandes distances entre les sources et les capteurs, l'influence des hétérogénéités locales sur la courbure des indicatrices apparente augmente d'une façon considérable. Ceci cause des perturbations statiques et dynamiques de la courbe de sommation qui sont difficilement corrigibles. Tous ces phénomènes font que la qualité des données avant et après sommation est très faible.

Plusieurs méthodes sont déjà disponibles pour améliorer le rapport signal sur bruit et ainsi la qualité des données après l'enregistrement. La méthode la plus simple est une sélection manuelle des 'bonnes' traces. Malheureusement, cette méthode est très coûteuse en temps à cause du grand volume des données de sismique réflexion profonde. Une autre approche est d'augmenter toute énergie latéralement cohérente, puisque c'est cette cohérence latérale qui caractérise les réflexions. Donc, cette approche fait 'ressortir' les réflexions. Cependant, le désavantage de cette approche est que non seulement les réflexions, mais également tous types de signaux sont latéralement cohérents comme par exemple les ondes de surface. Ceci a pour conséquence qu'une augmentation de toute énergie latéralement continue n'extrait pas que les réflexions mais possiblement aussi du bruit.

Afin d'éviter l'amplification de tous types de bruit latéralement cohérents, une extraction du signal désiré doit d'abord être effectuée, à savoir toute l'énergie réfléchie ou une réflexion particulière. Dans ce chapitre, nous étudions plusieurs méthodes pour extraire les réflexions dans des données de sismique réflexion profonde d'une façon semi-automatique afin d'augmenter le rapport signal sur bruit et ainsi de faciliter l'interprétation du profil considéré.

**Détection** Afin de pouvoir extraire un signal spécifique, il faut le détecter ou le reconnaître dans un enregistrement bruité. Même un interpréteur expérimenté ne sait pas dire sans aucune hésitation si un signal particulier est une réflexion ou du bruit dans des données de faible qualité. Donc, une approche statistique offre de nombreux avantages, car elle donne une estimation de la vraisemblance de la présence d'une réflexion.

Un problème de reconnaissance peut être considéré comme un problème de classification, c'est-à-dire un regroupement de certaines parties des données dans des catégories différentes représentant les différents types de signaux attendus. Dans le cas de la détection des réflexions,



les catégories différentes sont par exemple les réflexions, les ondes de surface et les autres types de bruit.

Dans cette thèse une approche supervisée est utilisée afin de pouvoir définir le nombre et les types de catégories. De plus, cette approche permet de choisir les exemples caractéristiques appartenant aux catégories définies. Par conséquent, une telle approche supervisée permet d'avoir une influence directe sur les résultats et la façon de les obtenir.

Une méthode de classification supervisée est composée de trois phases successives (Richards, 1993): la sélection, l'apprentissage et finalement la classification. Dans la phase de sélection, les diverses catégories sont définies et des exemples caractéristiques sont choisis en extrayant des fenêtres autour des pointés dans les données. Puis, dans la phase d'apprentissage, il faut apprendre les caractéristiques de chaque catégorie afin d'arriver à reconnaître et extraire les signaux différents dans la dernière phase.

**Mise en œuvre** Bien qu'une telle procédure soit très puissante, elle pose deux problèmes différents:

- Premièrement, quelle distribution statistique décrit le mieux la diversité dans une catégorie particulière? Une distribution Gaussienne est-elle suffisamment représentative ou faut-il chercher une autre distribution? Et doit-elle être différente pour chaque jeu de données?
- Deuxièmement, comment pouvons-nous traiter au mieux la diversité rencontrée dans des signaux appartenant à une même catégorie? C'est-à-dire, quel type de mesure (transformée de Fourier, etc. ...) permet d'obtenir les attributs distinguant au mieux entre les caractéristiques des catégories de signaux?

**Méthodes de classification** Quant à la distribution statistique spécifique, deux possibilités existent. On utilise soit une distribution définie *a priori* (en général une distribution Gaussienne), soit des réseaux neuronaux, car ils sont capables d'approximer la meilleure distribution d'une façon automatique.

Dans le cas où nous utilisons une distribution Gaussienne prédéfinie, la méthode est connue sous le nom de Classification par Vraisemblance Maximale (CVM ou MLC pour 'Maximum Likelihood Classification'). La méthode CVM calcule pour chaque catégorie individuelle les moyennes et les écarts-type de tous les attributs des exemples de référence, c'est-à-dire les exemples sélectionnés dans la première phase d'une classification supervisée. Ensuite, la vraisemblance de présence de chaque catégorie peut être calculée pour un signal donné en utilisant ces moyennes et écarts-type et une distribution Gaussienne. Le signal donné sera attribué à la catégorie de la vraisemblance maximale, d'où le nom de la méthode (figure II.1.1).

L'inconvénient principal de la méthode CVM est qu'elle ne donne de bons résultats que si la distribution statistique utilisée est proche de la vraie distribution sous-jacente des données. De plus, il n'existe souvent aucune connaissance *a priori* de cette distribution sous-jacente. Dans la plupart des cas, elle doit être déterminée par des approximations successives.

Les Réseaux Neuronaux (RN) sont, par contre, capables de trouver automatiquement cette distribution sous-jacente. Donc, ils peuvent constituer une alternative à la méthode CVM. Les réseaux neuronaux sont composés de plusieurs couches interconnectées, à savoir une couche d'entrée recevant les valeurs des attributs, une ou plusieurs couches cachées et une couche de sortie indiquant les vraisemblances de toutes les catégories. Une couche contient plusieurs neurones qui sont, en général, connectés avec tous les neurones de la couche suivante (figure II.1.2).

Chaque connexion a un poids différent et chaque neurone reçoit donc des signaux pondérés. Ensuite, il transmet un signal à tous les neurones de la couche suivante. Ce signal transmis est la

somme des signaux pondérés qui est ensuite transmise à une fonction d'activation non-linéaire. Cette fonction d'activation peut être une Gaussienne ou une fonction sigmoïde, c'est-à-dire, une fonction-échelon lisse. Cette fonction non-linéaire est nécessaire pour que le réseau puisse approximer des distributions non-linéaires comme par exemple des distributions asymétriques ou multimodales.

- Définition des catégories et choix des exemples de référence (pointés) **Sélection**
- Calcul des attributs des pointés et leurs moyennes et écarts-types **Apprentissage**
- Calcul des vraisemblances de toutes les catégories **Classification**
- Catégorie du signal = celle ayant la vraisemblance maximale

FIG. II.1.1 - Méthode de classification par vraisemblance maximale.

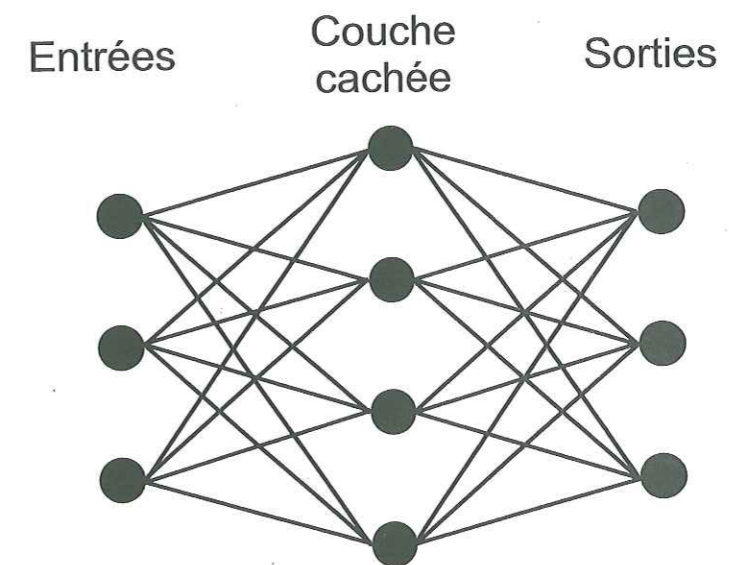


FIG. II.1.2 - Un réseau de neurones de base.

Pendant la phase d'apprentissage, le réseau neuronal essaie de trouver la meilleure distribution des poids internes en utilisant les exemples fournis. De nouveau, pendant la dernière phase de classification, un signal donné sera attribué à la catégorie ayant la vraisemblance maximale.

**Les attributs** La méthode CVM et les RN constituent deux techniques alternatives de reconnaissance qui sont capables de détecter et d'extraire un signal désiré et, ainsi, d'augmenter le



rapport de ce signal aux autres signaux (considérés comme du bruit). Maintenant, il nous reste à trouver les types d'attributs permettant une description optimale de la signature locale des données et, par conséquent, d'obtenir la meilleure caractérisation et les meilleurs résultats.

Si nous supposons que les données sont composées de réflexions, d'ondes de surface et d'autres types de bruit, une analyse du contenu fréquentiel local nous permettra de distinguer entre les différents types de signal. De plus, les ondes diffractées et les réflexions ne diffèrent que par un déphasage (Trorey, 1970). Les spectres d'amplitude d'une Transformée de Fourier à Court Terme (TFCT) et d'une Transformée Continue en Ondelette (TCO) sont insensibles aux distorsions de la phase et permettent ainsi à détecter à la fois les réflexions et les diffractions.

Un problème bien connu avec la TFCT est la difficulté d'estimer la longueur optimale de la fenêtre d'analyse glissante. Ceci pose surtout des problèmes si un enregistrement contient à la fois des signaux hautes fréquences et des signaux basses fréquences. Par contre, une TCO n'a pas ce problème, parce qu'elle analyse les données avec des fenêtres ayant des longueurs différentes. Malheureusement, ceci fait aussi qu'une TCO est plus lente qu'une TFCT.

Pour récapituler, quatre combinaisons seront utilisées afin de détecter et d'extraire les réflexions pour améliorer le rapport signal sur bruit, à savoir deux transformations locales de données (TFCT et TCO) et deux méthodes différentes de reconnaissance (CVM et RN). L'extraction du signal désiré se fait par une simple multiplication de la distribution de vraisemblance de présence de ce type de signal avec les données brutes.

**Reconstruction** Un inconvénient général de la stratégie proposée pour détecter les réflexions présentes est le fait qu'elle néglige toute information contenue dans les traces voisines. En effet, les vraisemblances de présence de réflexions sont calculées trace par trace, sans que les vraisemblances obtenues pour les traces voisines ne puissent influencer les résultats finaux de classification. Pourtant, on attend généralement une certaine homogénéité dans les distributions de types de signaux, puisque, par exemple, les réflexions et les ondes de surface ne sont pas aléatoirement mélangées, mais suivent des principes physiques bien définies. Donc, les méthodes proposées d'extraction de signaux ne permettent pas explicitement d'augmenter toute énergie latéralement cohérente. Cependant, ceci est un effet désiré, car il permet de faire 'ressortir' les réflexions et d'obtenir une estimation de toute énergie 'manquante'.

Deux méthodes différentes peuvent être envisagées pour incorporer les vraisemblances de différents types de signaux trouvées dans les traces voisines. Premièrement, les vraisemblances peuvent être pondérées avec la moyenne des vraisemblances autour de la partie de la trace considérée. Deuxièmement, après avoir fait une extraction du signal désiré, une technique d'interpolation de traces peut être appliquée pour corriger de l'énergie 'manquante'.

Les deux méthodes nécessitent un deuxième traitement des résultats, c'est-à-dire une deuxième itération. Elles ont bien l'effet désiré d'homogénéiser et, ainsi, d'améliorer les résultats finaux d'extraction. Par contre, la deuxième méthode peut plus facilement reconstruire les réflexions entières à cause de sa capacité d'interpolation.

Les méthodes décrites nous permettent d'améliorer, d'une façon semi-automatique, le rapport signal sur bruit et, ainsi, la qualité des données après l'enregistrement. Ceci est effectué sans qu'aucune amplification de tout type de bruit latéralement cohérent n'apparaisse. Cette qualité améliorée des données rendra l'interprétation des profils de sismique réflexion profonde plus facile et, ainsi, donnera plus de contraintes évidentes sur la géologie et la tectonique de la région considérée.

## II.1.2 Présentation du chapitre

Afin d'illustrer les possibilités des méthodes d'extraction de réflexions pour améliorer la qualité des données, nous utilisons les données de l'expérience de sismique réflexion profonde Ecors 2.5-D, qui a été effectuée dans la partie ouest de la Zone Nord Pyrénéenne (Marthélot et al., 1994). Cette région est caractérisée par une grande complexité tectonique. L'expérience est composée d'un seul andain d'une vraie configuration de sismique 3-D de prospection pétrolière. Elle a fourni un grand volume de données avec malheureusement un ordre de couverture très limité. Pour les raisons décrites auparavant, les résultats finaux du profil après sommation sont difficiles à interpréter. Seule une sélection manuelle des 'bonnes' traces a permis d'obtenir une section de qualité correcte (Siliqi, 1994).

Dans la section II.2 nous donnons une description brève de la géologie de la Zone Nord Pyrénéenne et plus de détails sur le dispositif et la localisation exacte du profil par rapport aux principaux structures géologiques de la région.

Ensuite, nous décrivons la théorie de la détection statistique de signaux dans plus de détails dans la section II.3 et nous montrons que la méthode CVM est capable d'augmenter le rapport signal sur bruit dans des collections de traces à point à tir commun et dans des coupes après sommation en couverture multiple. Parce que la détection automatique trace par trace ne donne que des images fragmentaires des réflecteurs et, ainsi, de la géologie, nous décrivons deux techniques de reconstruction qui permettent d'obtenir des bons résultats finaux beaucoup plus faciles à interpréter.

Les réseaux neuronaux sont beaucoup plus flexibles que la méthode CVM et n'exigent pas une spécification *a priori* de la distribution statistique des données. La section II.4 contient une description étendue de ces techniques. A cause du fait que les RN ne sont pas encore d'une utilisation courante dans le domaine de la géophysique appliquée, nous décrivons aussi des méthodes différentes permettant de contrôler et d'améliorer leurs performances. Il s'agit surtout des méthodes concernant l'apprentissage, la généralisation et la détermination de la structure optimale. Finalement, nous montrons que plusieurs RN différents peuvent donner des résultats similaires à ceux obtenus par la méthode CVM sur une collection de traces à point à tir commun.

Dans les sections II.3 et II.4 les signaux sont analysés et caractérisés par une TFCT. Bien que cette transformation donne des bons résultats, elle nécessite une spécification *a priori* de la longueur optimale de la fenêtre d'analyse et la qualité des résultats obtenus dépend directement de cette longueur. Contrairement à la TFCT, une TCO permet une analyse des données avec des fenêtres ayant des longueurs différentes et évite ainsi cette spécification *a priori*. La section II.5 contient une introduction à cette transformation particulière et des résultats obtenus en utilisant la méthode CVM et des RN.

Finalement, nous présentons dans la dernière section une des coupes sommées le long d'une ligne de récepteurs de l'andain Ecors 2.5-D. Nous montrons comment nous avons augmenté la qualité de cette coupe en utilisant les techniques décrites d'extraction et de reconstruction de l'énergie réfléchie. De plus, nous comparons la migration de cette coupe-somme avec les résultats obtenus par Siliqi (1994).



## II.2 L'expérience Ecors 2.5-D dans les Pyrénées

### Résumé

Deux grands profils de sismique réflexion profonde ont traversé les Pyrénées perpendiculairement à leur axe. Le premier se trouve dans les Pyrénées orientales et le second dans les Pyrénées occidentales. Dans cette étude, nous utilisons un troisième profil qui se trouve au milieu du second profil dans la Zone Nord Pyrénéenne. Ce profil est nommé Ecors 2.5-D, puisqu'il est composé d'un seul andain d'une configuration d'acquisition à 3 dimensions de la sismique de prospection.

La géologie des Pyrénées est très compliquée, mais les deux phénomènes caractéristiques sont sa forme en éventail et la différence en profondeur des Mohos ibérique et européen.

### II.2.1 Eléments sur la géologie des Pyrénées

Les Pyrénées sont une chaîne de montagnes séparant la France de l'Espagne sur une longueur d'environ 500 km. Leur largeur varie entre 50 et 150 km. Jusqu'à présent, deux profils de sismique réflexion profonde ont traversé les Pyrénées perpendiculairement à leur axe: le profil Ecors Pyrénées Est (Choukroune and Ecors-Team, 1989) et le profil Ecors Arzacq (Daignières et al., 1994) dans la partie ouest. Les localisations des deux profils sont montrées sur la figure II.2.1. Elle montre aussi un troisième profil dont nous parlerons plus tard. Le but de ces deux profils était d'obtenir plus de contraintes sur la structure crustale des Pyrénées et ses implications tectoniques.

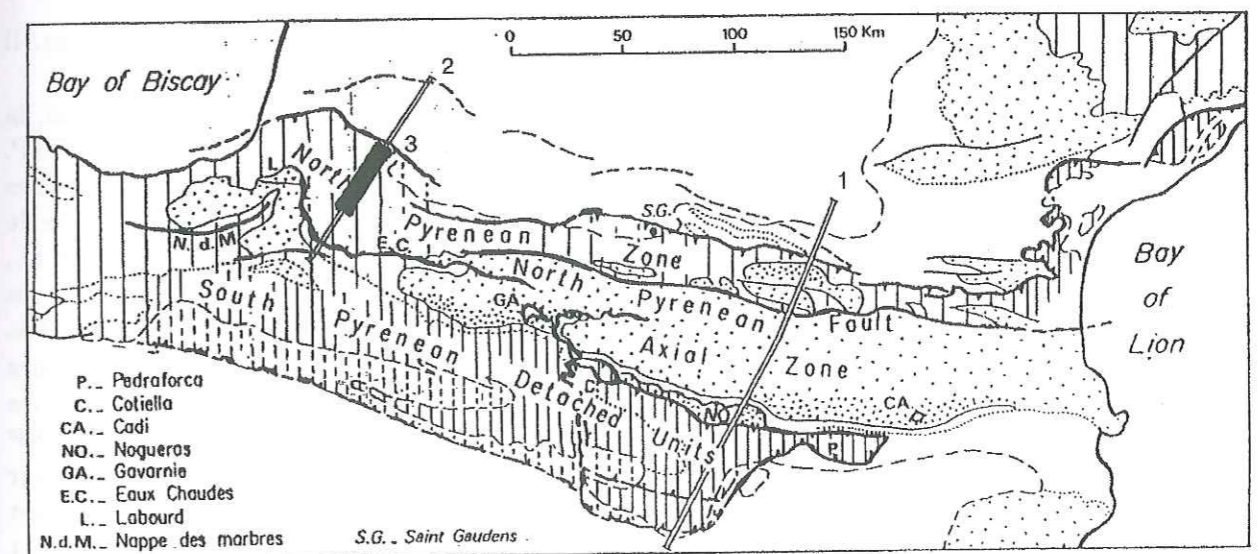


FIG. II.2.1 - Les localisations des profils Ecors. 1 - Pyrénées Est, 2 - Ecors Arzacq et 3 - Ecors 2.5-D. D'après Siliqi (1994).

L'histoire de la formation des Pyrénées est marquée par deux événements tectoniques principaux. Le premier est une phase d'extension liée au mouvement relatif des plaques ibérique et européenne et à la formation du golfe de Gascogne pendant le Crétacé Inférieur. Le second est une phase de compression pendant le Crétacé Supérieur et l'Eocène (Choukroune and Ecors-Team, 1989; Roure et al., 1989).



Généralement les Pyrénées sont divisées en plusieurs zones principales (figure II.2.1), à savoir (du nord au sud) le bassin d'avant pays européen (le Bassin Aquitain), la Zone Nord Pyrénéenne, la Zone Axiale, la Zone Sud Pyrénéenne et le bassin d'avant pays ibérique (le Bassin d'Ebre). La structure géométrique la plus remarquable des Pyrénées est sa forme en éventail avec des chevauchements vers le nord dans la Zone Nord Pyrénéenne, des structures verticales autour de la Zone Axiale et des chevauchements vers le sud dans la Zone Sud Pyrénéenne. De plus, le Moho ibérique est beaucoup plus profond que le Moho européen (respectivement entre 14 à 15 s et 9 à 10 s temps double). Ces deux phénomènes sont clairement visibles sur la figure II.2.2 qui montre les coupes-sommes migrées et interprétées des trois profils Ecors Pyrénées (Roure et al., 1989; Daignières et al., 1994; Siliqi, 1994).

Les deux bassins d'avant pays sont asymétriques et d'origine flexurale. Le Bassin Aquitain est composé d'une épaisse couche sédimentaire d'âge Mésozoïque sur un socle Paléozoïque. Par exemple, dans le domaine Arzacq du Bassin Aquitain (Pyrénées ouest) qui nous intéresse ici, les sédiments font plus de 10 km d'épaisseur.

La Zone Nord Pyrénéenne est limitée au sud par la Faille Nord Pyrénéenne (FNP) et au nord par le Chevauchement Frontal Nord Pyrénéen, qui chevauche la partie sud du Bassin Aquitain. Dans les Pyrénées occidentales, la Zone Nord Pyrénéenne est composée des Massifs Nord Pyrénéens et du bassin de Mauléon. Ce bassin est constitué de sédiments Mésozoïques d'épaisseur variable sur un socle Paléozoïque. Il est divisé en deux par le Chevauchement de Saint-Palais dont la semelle est constituée d'évaporites Triassiques (figure II.2.2.b). Le bassin de Mauléon est limité au nord par le Chevauchement de Sainte-Suzanne qui est un segment du Chevauchement Frontal Nord Pyrénéen et repose ainsi sur le Bassin Aquitain (Daignières et al., 1994).

Les Massifs Nord Pyrénéens et la Zone Axiale ont un noyau Hercynien (Paléozoïque). Il existe même quelques affleurements pré-cambriens (Protérozoïques) reliés au manteau.

La FNP constitue la limite entre les Massifs Nord Pyrénéens et la Zone Axiale. De plus, la transition entre les Mohos ibérique et européen se trouve toujours dans les environs de la FNP. Certains auteurs pensent même que la FNP est la manifestation en surface de la suture des plaques ibérique et européen (Choukroune and Ecors-Team, 1989). Par contre, la géométrie de la FNP en profondeur est mal connue. Selon les hypothèses, elle est plus concave (Roure et al., 1989) ou plus verticale (Mattauer, 1990). Ça forme exacte a des grandes implications sur les structures géologiques et les mécanismes tectoniques juste avant la vraie orogénèse des Pyrénées.

Malgré les grandes similarités dans les structures géologiques en surface, il existe plusieurs différences entre les Pyrénées orientales et occidentales. A l'ouest, la Zone Axiale est encore couverte par des dépôts Crétacés et Eocènes. De plus, la manifestation en surface de la FNP n'est pas tout à fait claire à cause des Massifs Nord Pyrénéens qui sont chevauchés vers le sud sur la couverture Mésozoïque de la Zone Axiale. La Zone Nord Pyrénéenne est plus large dans l'ouest que dans l'est des Pyrénées (60 km au lieu de 20 km) et la métamorphose à basses pressions et hautes températures du Crétacé manque dans la zone autour de la FNP dans l'ouest.

## II.2.2 Eléments sur le dispositif Ecors 2.5-D

Après avoir acquis deux profils sismiques traversant les Pyrénées, l'équipe Ecors a décidé de réaliser une nouvelle expérience sismique afin d'essayer de mettre en évidence des structures à 3 dimensions dans la croûte entière. Dans ce but, une adaptation des géométries d'acquisition à 3-D de la sismique de prospection a été utilisée. Vu que seulement un seul andain a été utilisé, nous ne pouvons pas obtenir une image de la croûte ayant la même qualité et autant de détails comme nous la rencontrons souvent dans la sismique de prospection à 3-D. A cause de l'existence

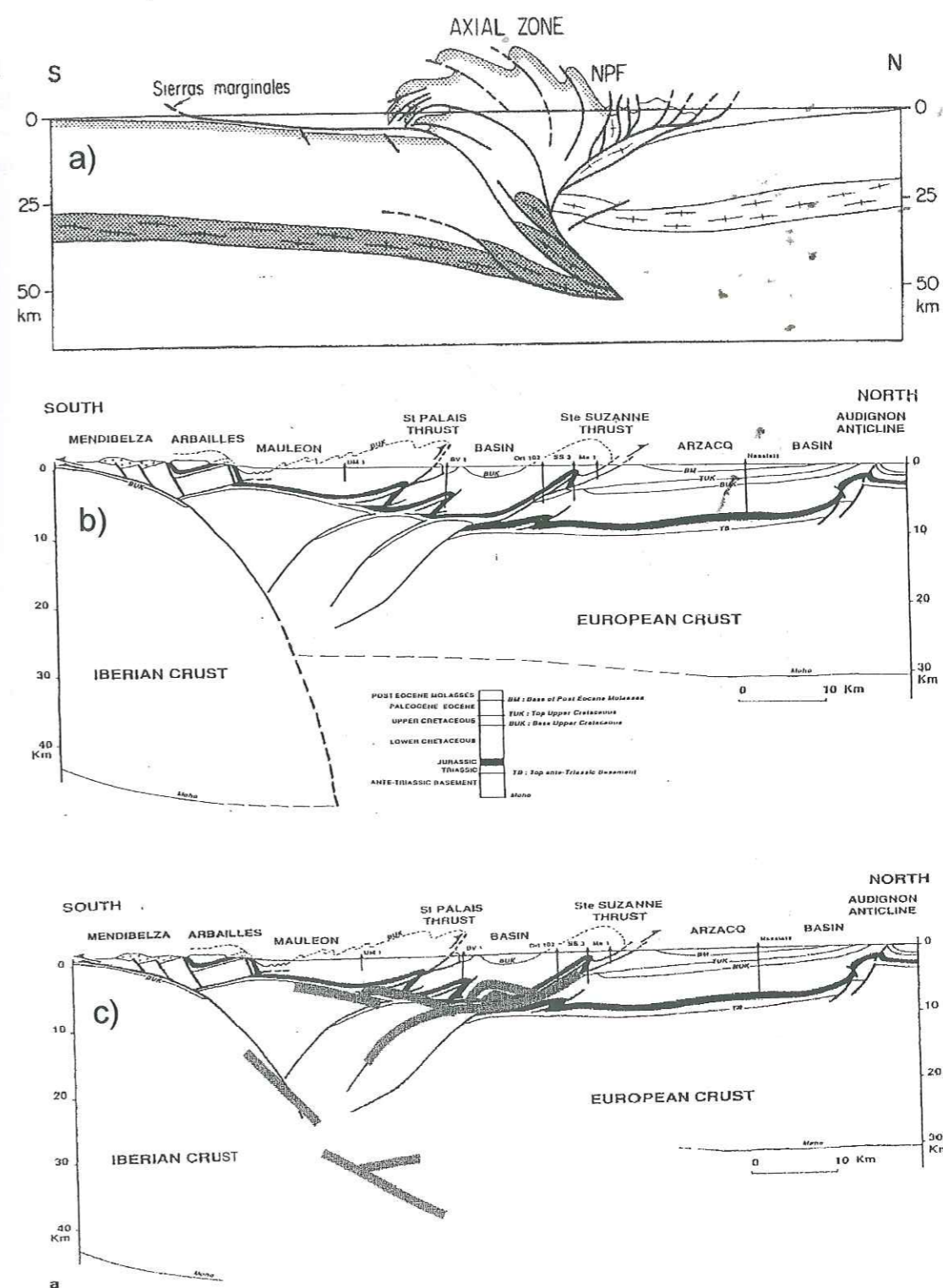


FIG. II.2.2 - Les coupes-sommes migrées et interprétées des 3 profils Ecors Pyrénées. a) Pyrénées Est, b) Ecors Arzacq et c) l'interprétation du profil Ecors 2.5-D en traits gras) superposée sur b). La structure en éventail n'est visible que sur le premier profil, car les deux autres profils se trouvent au nord de la Zone Axiale (figure II.2.1). Par contre, le premier et le second profil montrent bien la différence de profondeur du Moho. D'après Roure et al. (1989), Daignières et al. (1994) et Lecerf (1996).



de ce seul andain, l'expérience était nommée Ecors 2.5-D (Marthelot et al., 1994).

Les figures II.2.1 et II.2.3 nous montrent la localisation exacte de l'expérience Ecors 2.5-D dans la Zone Nord Pyrénéenne des Pyrénées occidentales. Elle est entièrement située dans le Bassin de Mauléon. La partie nord se trouve dans le synclinal d'Orriule entre le Chevauchement de Sainte-Suzanne et le Chevauchement de Saint-Palais et la partie sud à quelques kilomètres du massif des Arbailles (un des Massifs Nord Pyrénéens).

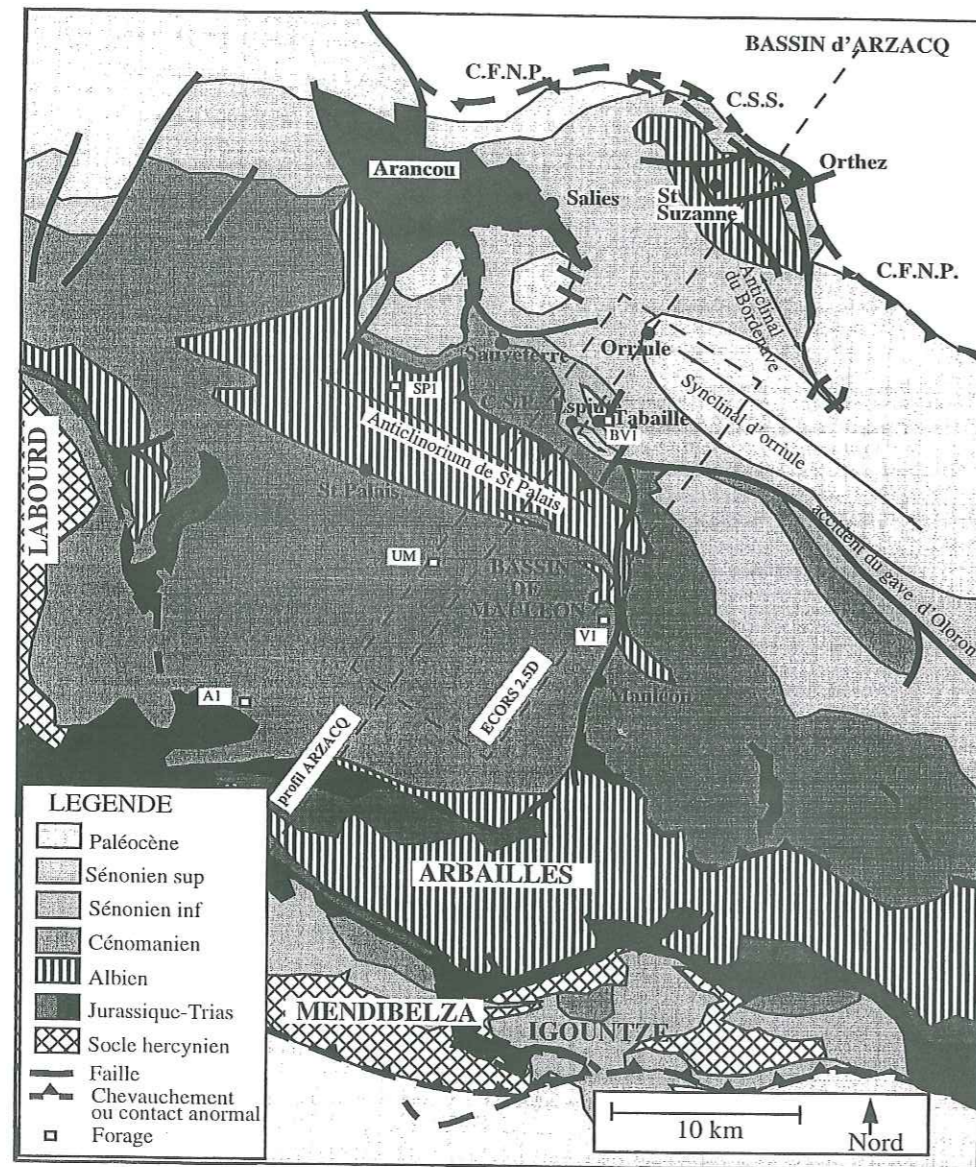


FIG. II.2.3 – La localisation exacte du profil Ecors 2.5-D dans la géologie régionale. CSS - Chevauchement de Sainte-Suzanne; CSP - Chevauchement de Saint-Palais; CFNP - Chevauchement Frontal Nord Pyrénéen. Figure extraite de Lecerf (1996).

La figure II.2.4 nous donne une idée du dispositif utilisé. L'expérience utilisait 6 lignes parallèles de capteurs. Chaque ligne était composée de 120 groupes de capteurs et chaque groupe individuel de 18 géophones distribués selon une simple droite. La distance entre deux groupes de capteurs le long d'une ligne était de 100 m, alors que la distance entre deux lignes faisait 1600

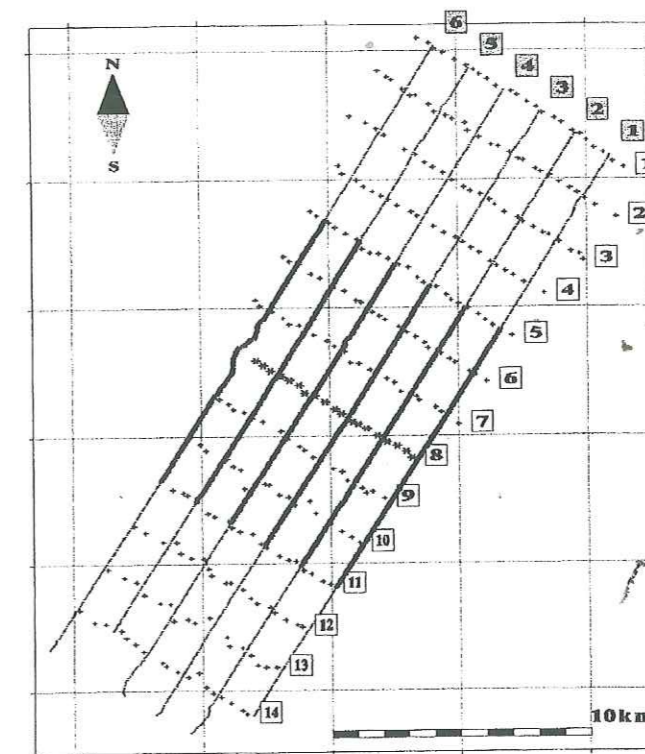


FIG. II.2.4 – Dispositif du profil Ecors 2.5-D. Figure extraite de Marthelot et al. (1994).

m. Chaque tir était ainsi enregistré sur une surface de 96 km<sup>2</sup>. Les tirs étaient localisés sur 14 colonnes perpendiculaires aux lignes de capteurs (figure II.2.4). Il y avait entre 20 et 28 tirs dans une colonne. La distance entre deux tirs dans une colonne était de 400 m. Entre chaque série de tirs, les lignes et les colonnes étaient déplacées de 2 km dans la direction des lignes. Ainsi, une surface totale de 8 x 27 km<sup>2</sup> a été couverte par les 14 colonnes et les 292 tirs. La figure II.2.4 montre qu'un dispositif avec tir au centre a été utilisé sauf aux extrémités du profil. Les charges explosives (10 à 30 kg) étaient placées dans des puits avec une profondeur comprise entre 25 et 30 m (Marthelot et al., 1994).

Siliqi (1994) a sélectionné les bonnes parties des données avec des techniques automatiques et manuelles. Son interprétation de la coupe-somme migrée est montrée sur la figure II.2.2.c (traits épais). Cette figure nous montre qu'il y a des fortes indications que la FNP existe dans l'ouest des Pyrénées et qu'elle est vraiment la limite entre les deux plaques. Le Moho semble d'être bifurqué sous le bassin de Mauléon indiquant possiblement le contact entre les deux plaques.



## II.3 Classification par vraisemblance maximale

### Résumé

Les réflexions dans des données de sismique réflexion profonde ne sont souvent que visibles sur quelques traces voisines d'une collection à point milieu commun. Afin d'éviter la dégradation de la coupe somme à cause de cette énergie non-cohérente, il faut l'enlever.

Dans cette étude, nous présentons une technique de classification d'utilisation courante dans le domaine de la télédétection et nous montrons comment cette technique peut être appliquée afin d'améliorer la qualité des données. Elle utilise une méthode de reconnaissance pour détecter et extraire toute énergie cohérente à la fois dans des collections de traces à point à tir commun et dans des coupes-sommes finales. La technique calcule la vraisemblance de la présence d'un signal sismique en utilisant la signature statistique fournie par des pointés de référence. Ensuite, nous obtenons une coupe 'propre' en multipliant cette distribution de vraisemblance avec les données originales.

L'application de la technique proposée renforce toutes les réflexions présentes dans les données d'une expérience de sismique réflexion profonde. La technique décrite de reconnaissance ne peut détecter que des signaux présents, mais elle ne peut reconstruire l'énergie 'manquante'. Afin de pouvoir reconstruire cette énergie, nous proposons 2 méthodes de reconstruction. La première est une analyse  $\tau$ - $p$  locale pondérée avec la semblance à appliquer après la technique de reconnaissance. La seconde consiste à appliquer directement une technique de reconnaissance dans le domaine  $\tau$ - $p$ . Dans tous les cas, nous améliorons la qualité des données en ne sommant que des signaux sismiques spécifiques. Nous montrons que si nous utilisons à la fois la reconnaissance et la reconstruction, nous obtenons des meilleurs résultats que si nous n'appliquons qu'une seule des 2 techniques.



## Recognition and reconstruction of coherent energy with application to deep seismic reflection data<sup>1</sup>

Mirko van der Baan and Anne Paul

*Laboratoire de Géophysique Interne et Tectonophysique,  
Université Joseph Fourier and CNRS, BP 53, 38041 Grenoble-Cedex 9, France*

### Abstract

Reflections in deep seismic reflection data tend to be visible on only a limited number of traces in a common midpoint gather. To prevent stack degeneration, any non-coherent reflection energy has to be removed.

In this paper, a standard classification technique in remote sensing is presented to enhance data quality. It consists of a recognition technique to detect and extract coherent energy in both common shot gathers and final stacks. This technique uses the statistics of a picked seismic phase to obtain the likelihood distribution of its presence. Multiplication of this likelihood distribution with the original data results in a 'cleaned up' section. Application of the technique on data of a deep seismic reflection experiment enhanced the visibility of all reflectors considerably.

Because the recognition technique cannot produce an estimate of 'missing' data, it is extended with a reconstruction method. Two methods are proposed: application of semblance weighted local slant stacks after recognition or direct recognition in the linear  $\tau$ - $p$  domain. In both cases, the power of the stacking process to increase the signal-to-noise ratio is combined with the direct selection of only specific seismic phases. The joint application of recognition and reconstruction resulted in data images which showed reflectors that were present even more clearly than mere application of a single technique.

### II.3.1 Introduction

As in exploration seismics, remote sensing or teledetection suffers from the fact that vast amounts of data are gathered that require intensive manual labor before they are processed and interpreted (Richards, 1993). In addition, the obtained data may be distorted and may not always have the desired high quality. Therefore, attention in remote sensing has been focussed for a long time on the questions of how to reduce processing and interpretation time and how to enhance the obtained images using semi-automatic or interactive techniques. For example, a frequently encountered problem in remote sensing is the need of classification of each pixel of all images into several classes to monitor, e.g., crop growth or damages in forests due to summer fires. For this specific task, very often recognition and classification techniques are used.

In this paper, a standard classification technique in remote sensing is presented and it is shown how this technique can be used to reduce the problem of low signal-to-noise ratios in deep seismic reflection data. In conventional processing, stack quality is decreased notably because of two phenomena. Firstly, reflections tend to be visible on only a limited number of adjacent traces, and secondly, due to the large offsets – half-offsets of several kilometers are not unusual – move-outs very often exhibit static and even dynamic undulations on both a large and a small scale caused by the presence of overlying heterogeneities. Stack degeneration can readily be reduced by removing non-coherent energy along the stacking trajectories, thereby attacking the first problem. The second problem is beyond the scope of this paper.

<sup>1</sup>Article accepté par Geophysics en 1999.



Many different methods to increase the signal-to-noise ratio are available in exploration geophysics. Among these are the skeletonization method (Le and Nyland, 1990; Lu and Cheng, 1990), the statistical method of Hansen et al. (1988), methods using local  $\tau$ - $p$  analyses (Harlan et al., 1984; Kong et al., 1985) and spatial prediction filtering (Hornbostel, 1991; Abma and Claerbout, 1995).

The simplest method would be to calculate a variable describing the local coherence between traces as is e.g. done in some attribute maps like the coherence cube (Bahorich and Farmer, 1995; Marfurt et al., 1998). A multiplication of the local coherence with the original traces will emphasize the coherent parts. However, the skeletonization method (Le and Nyland, 1990; Lu and Cheng, 1990) will probably represent a better alternative. This method reduces a seismic section to a line drawing by correlating waveforms of neighboring traces, thereby automatically tracking seismic reflectors. However, for these methods to work properly, a good signal-to-noise ratio must already be present in the data.

The method of Hansen et al. (1988) scans subsections over a range of slownesses and intercept times and simply counts the number of positive and negative amplitudes in a line segment. Only if either number exceeds a certain threshold, the examined line segment is passed as signal. Moreover, the method can assert both the probability that a line is passed containing only noise and the total number of noise-only segments.

The methods actually computing the local  $\tau$ - $p$  transforms use it either as a focusing measure (Harlan et al., 1984) or in combination with semblance (Kong et al., 1985). The first method uses the  $\tau$ - $p$  transform to obtain the desired amplitude distributions of noise and signals needed for the recognition and extraction of samples containing the highest percentage of signal, whereas in the second method semblance weighted slant stacks are used to emphasize and extract laterally coherent signals.

Finally, spatial prediction filtering is widely used in industry to improve data quality. This approach replaces samples by their predicted values based on information contained in adjacent traces. The laterally coherent signal energy can be extracted using two different techniques, namely by means of least-mean-square adaptive filtering in the  $t$ - $x$  domain (Hornbostel, 1991) or using a Wiener-filter in either the  $f$ - $x$  or  $t$ - $x$  domains (Hornbostel, 1991; Abma and Claerbout, 1995).

Unfortunately, all of the above-mentioned methods emphasize any laterally coherent energy including coherent noise. Hence, even if such noise has a signature which is completely different from the desired reflections, it will be extracted. Therefore, a technique extracting preselected phases only is needed.

Moreover, most of the above-described methods assume that reflectors are at least locally linear, i.e. represented by a line in 2-D or a plane in 3-D data volumes, whereas it is well known that this is rarely the case in real data. Thus, in addition, a method is required which is still able to distinguish signal from noise, but does not impose these restrictions.

With these aims in mind, a standard supervised classification technique from remote sensing will be introduced. Supervised classification is done using some specific attributes describing a local part of the data (i.e. its signature) and consists basically of three steps (Richards, 1993): the selection, training or learning and classification stage. In the first stage the different classes are defined and a training set is selected for each class. Then the characteristics of each class are estimated (the training stage) and finally classification or clustering takes place. Each class represents a different category or type of signal (e.g. burnt and non-burnt woods, urban soil and water in remote sensing or the different types of signals in seismic data: reflections, diffractions, ground roll and noise). The characteristics of the training set are defined as the averages and variances of the attributes. With these the statistical likelihoods of the presence of each class can

be calculated and the data can be clustered by means of a simple comparison. After clustering, i.e. classification, the desired signals are extracted, thereby enhancing the visibility of that specific signal.

What attributes best discriminate between different signals? If, in a first simplification, data is assumed to be composed of ground roll, reflections and noise, then discrimination can be done using frequency content. Since reflections and diffractions differ only by a phase shift (Torey, 1970), the most logical signature would be the local amplitude spectrum. Hence, the use of the local amplitude spectrum allows the classification technique to recognize signals displaying phase perturbations including e.g. diffractions. Most other techniques, however, demand for phase consistency notably along the stacking trajectories, thereby excluding the detection of diffractions. This additional ability to detect diffracted signals sets the recognition method even further apart from the other methods.

On the other hand, the methods incorporating adjacent traces have the possibility to reconstruct 'missing' energy, thereby improving data quality over mere classification. Fortunately, it is not difficult to adapt the recognition technique to estimate for missing data. We propose two methods: a) the application of the reconstruction method of Kong et al. (1985) using semblance weighted slant stacks after recognition and b) direct application of the recognition technique in the  $\tau$ - $p$  domain. Unfortunately, these methods require again the assumption of locally linear events.

First the recognition method is described and it is shown how this method can be used to extract the 'coherent' features in a seismic section. Then the topic of reconstruction is addressed. The effectiveness of the recognition technique and some of the possibilities of reconstruction are both shown using the data of the Ecors 2.5-D deep seismic reflection experiment (Marthelot et al., 1994). Finally, the combined recognition and reconstruction techniques are discussed in some more detail, some possible modifications are given of the recognition method and several further applications are indicated.

## II.3.2 Recognition method

### II.3.2.1 Unconditional maximum likelihood classification

The most common supervised classification method used for remote sensing image data is maximum likelihood classification (Richards, 1993). Assuming that we have a vector  $\mathbf{x}$  containing the attributes describing a local part of the data, classification only requires knowledge of the conditional probability density functions (pdf)  $p(\omega_i|\mathbf{x})$ , where  $\omega_i$  are the  $I$  (predefined) spectral classes. Namely,

$$\mathbf{x} \in \omega_i \text{ if } p(\omega_i|\mathbf{x}) > p(\omega_k|\mathbf{x}) \quad \forall k \neq i, \quad (\text{II.3.1})$$

since the conditional pdf  $p(\omega_i|\mathbf{x})$  represents the likelihood that  $\omega_i$  is the correct class given the attribute vector  $\mathbf{x}$ . The spectral classes  $\omega_i$  are mutually exclusive and span the spectrum  $\Omega$ , such that always one and only one class must occur. Thus, mathematically,  $\omega_1 \cup \omega_2 \cup \dots \cup \omega_I = \Omega$  and  $\omega_i \cap \omega_k = \emptyset \quad \forall i \neq k$ . These conditions are necessary to justify the use of Bayes' rule hereafter.

To be more specific, the spectral classes  $\omega_i$  can be seen as the different types of anticipated signals and the attribute vector  $\mathbf{x}$  as representing the local amplitude spectrum for a specific time and offset. In that case, the conditional pdf  $p(\omega_i|\mathbf{x})$  will represent the probability that a particular signal type is present at a specific time and offset given the local amplitude spectrum.

The, yet, unknown conditional pdf  $p(\omega_i|\mathbf{x})$  can be estimated using Bayes' rule, which states that

$$p(\omega_i|\mathbf{x}) = p(\mathbf{x}|\omega_i)p(\omega_i)/p(\mathbf{x}), \quad (\text{II.3.2})$$



thus relating the conditional pdf  $p(\omega_i|\mathbf{x})$  to the *a priori* pdf  $p(\omega_i)$ , to  $p(\mathbf{x})$  and to the other conditional pdf  $p(\mathbf{x}|\omega_i)$ . The denominator  $p(\mathbf{x})$  represents the probability distribution that any class is present for the attribute vector  $\mathbf{x}$ . Since it is independent of the classes  $\omega_i$ , it is either neglected or treated as a normalization constant. The pdf  $p(\omega_i)$  reflects our *a priori* estimation of the occurrence of class  $\omega_i$  and is independent of the data. Finally, the conditional pdf  $p(\mathbf{x}|\omega_i)$  describes the likelihood that class  $\omega_i$  will be present at the data point with attribute vector  $\mathbf{x}$  given the characteristics of class  $\omega_i$ . The learning phase of the supervised classification method mainly consists of estimating this distribution function.

If we assume that the conditional pdf  $p(\mathbf{x}|\omega_i)$  has a normal or Gaussian distribution then equation (II.3.2) can be written as

$$p(\omega_i|\mathbf{x}) = cp(\omega_i)(2\pi)^{-J/2}|\mathbf{C}^{(i)}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{m}^{(i)})^t \mathbf{C}^{(i)-1} (\mathbf{x} - \mathbf{m}^{(i)})\right\} \quad (\text{II.3.3})$$

with  $c$  the normalization constant,  
 $|\mathbf{C}^{(i)}|$  the determinant of the covariance matrix  $\mathbf{C}^{(i)}$  for class  $\omega_i$ ,  
 $\mathbf{m}^{(i)}$  the mean of the attributes for class  $\omega_i$ ,  
 $t$  vector transpose,  
 $J$  size of vector  $\mathbf{x}$ .

If either the covariances are assumed to be negligible or not enough samples can be selected for each spectral class to justify their estimation then equation (II.3.3) can be approximated by using only the variances  $\sigma_j^{(i)2}$  for each attribute  $j$  of class  $\omega_i$ , i.e.

$$p(\omega_i|\mathbf{x}) = cp(\omega_i)(2\pi)^{-J/2} \prod_{j=1}^J \sigma_j^{(i)-1} \exp\left\{-\frac{(x_j - m_j^{(i)})^2}{2\sigma_j^{(i)2}}\right\}. \quad (\text{II.3.4})$$

Equation (II.3.4) constitutes a close approximation to the exact equation (II.3.3) if covariances are negligible. However, if no accurate estimation of the covariance matrix can be obtained then equation (II.3.4) may be a bad substitution for equation (II.3.3), since no guarantee exists that attributes are not highly correlated. In the next subsection a standard technique will be described to reduce existing correlations between attributes.

Therefore, to perform supervised classification of the data, in the first stage a representative set of  $K^{(i)}$  samples is selected and their attribute vectors  $\mathbf{x}^{(i,k)}$  ( $k = 1, K^{(i)}$ ) are calculated for each class  $\omega_i$ . Then in the learning stage the means and covariance matrices or variances are calculated using

$$\mathbf{m}^{(i)} = \frac{1}{K^{(i)}} \sum_{k=1}^{K^{(i)}} \mathbf{x}^{(i,k)} \quad (\text{II.3.5})$$

and

$$\mathbf{C}^{(i)} = \frac{1}{K^{(i)} - 1} \sum_{k=1}^{K^{(i)}} \left\{ (\mathbf{x}^{(i,k)} - \mathbf{m}^{(i)}) (\mathbf{x}^{(i,k)} - \mathbf{m}^{(i)})^t \right\} \quad (\text{II.3.6})$$

or

$$\sigma_j^{(i)2} = \frac{1}{K^{(i)} - 1} \sum_{k=1}^{K^{(i)}} (x_j^{(i,k)} - m_j^{(i)})^2, \quad (\text{II.3.7})$$

where  $j$  ranges over the elements of the attribute vector  $\mathbf{x}$  (i.e. from 1 to  $J^{(i)}$ , the number of attributes selected for the  $i^{\text{th}}$  class).

Finally, all conditional pdfs are calculated with either equation (II.3.3) or (II.3.4) and classified using equation (II.3.1).<sup>2</sup>

As was already mentioned above, in our case the attribute vector contains the amplitude spectrum of a local part of the data and the characteristics are then calculated from representative picks of each spectral class, i.e. each signal type. To prevent detection problems due to amplitude fluctuations of reflection, energy contained in all amplitude spectra is normalized. Since classification occurs using a local  $f$ - $t$  analysis, basically a flexible frequency analysis is used to detect coherent energy. With the amplitude spectra as attributes, equation (II.3.5) and (II.3.7) therefore represent the mean spectral amplitude and variance, respectively.

### II.3.2.2 Correcting for correlations between attributes

A fundamental problem arises if equation (II.3.3) has to be approximated by equation (II.3.4) due to an insufficient number of picks to guarantee a close estimation of the covariance matrix  $\mathbf{C}$ . Namely, non-anticipated correlations between attributes may reduce the detection ability of the recognition technique. To give an idea of the number of picks required, it is often assumed that at least 10 times more training samples are needed than the number of used attributes (Richards, 1993).

A principal component analysis is a standard technique in clustering or classification problems to decrease existing dependencies between attributes and thereby reducing the number of required attributes for an accurate performance. This analysis is also known as the Karhunen-Loève transform and is most often used for image reconstructions and statistical analyses (Richards, 1993). It is closely related to the Singular Value Decomposition (SVD) as shown by Freire and Ulrych (1988). In geophysics, it is most often used for wave-noise separation and wave-wave separation (Hemon and Mace, 1978; Freire and Ulrych, 1988; Glangaud and Mari, 1994) or (multi-variate) statistical analyses in reservoir characterization studies (Dumay and Fournier, 1988).

It is well known that a non-square matrix  $\mathbf{X}$  can be rewritten in a semi-diagonal form by means of an SVD (Lanczos, 1961), i.e.

$$\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^t. \quad (\text{II.3.8})$$

The columns of both matrices  $\mathbf{U}$  and  $\mathbf{V}$  contain orthogonal eigenvectors and the matrix  $\mathbf{\Lambda}$  is a diagonal matrix of which the elements  $\lambda_j$  are the singular values of  $\mathbf{X}$ . These singular values are the positive square roots of the eigenvalues of both  $\mathbf{X}\mathbf{X}^t$  and  $\mathbf{X}^t\mathbf{X}$ . The Karhunen-Loève transform or principal component analysis consists of projecting data upon the eigenvectors contained in  $\mathbf{U}$ . Thus, e.g., the principal component matrix  $\mathbf{Y}$  is defined as (Freire and Ulrych, 1988; Richards, 1993):

$$\mathbf{Y} = \mathbf{U}^t\mathbf{X} = \mathbf{\Lambda}\mathbf{V}^t. \quad (\text{II.3.9})$$

Redundancy in data can now be decreased by using only a limited number of eigenvectors  $\mathbf{u}_j$ , where  $\mathbf{u}_j$  is defined as the  $j^{\text{th}}$  column of  $\mathbf{U}$  and singular values are sorted in decreasing order. Therefore, a vector  $\mathbf{x}$  can be transformed to a vector  $\mathbf{x}'$  of reduced length using only the first  $p$  eigenvectors  $\mathbf{u}_j$ :

$$x'_j = \mathbf{u}_j \cdot \mathbf{x}, \quad j = 1, p. \quad (\text{II.3.10})$$

<sup>2</sup>La figure II.5.1 nous montre le schéma proposé.



Hence, a vector is projected upon a new set of axes, displaying a decreasing order of correlations between the training attributes. Furthermore, a variable is needed to measure the perfectness of the projection, i.e. the information still contained in the new vectors. This can be done using the energy  $E_p$  contained in the first  $p$  eigenvectors, i.e.

$$E_p = \sum_{j=1}^p \lambda_j^2 / \sum_{j=1}^J \lambda_j^2. \quad (\text{II.3.11})$$

Thus, to decrease redundancy in the training samples and existing correlations between attributes, the columns of the matrix  $\mathbf{X}$  should contain the amplitude spectra of the training samples. The eigenvectors  $\mathbf{u}_j$  will then be used as a new set of axes upon which the attributes  $\mathbf{x}$  can be projected using equation (II.3.10). Finally, note that the computation of the SVD is not strictly required, since  $\mathbf{\Lambda}$  and  $\mathbf{U}$  can be directly computed from  $\mathbf{X}\mathbf{X}^t$ . In our application, normally one or, in some rare cases, two eigenvectors will represent virtually 100 % of the total energy, since the amplitude spectra of the picks are highly similar.

Classification can then proceed in the following way. First, examples for each single class are picked and the needed eigenvectors  $\mathbf{u}_j$  are calculated. Then, the samples are projected upon the eigenvectors with equation (II.3.10), and the averages and variances of the transformed attributes are calculated using equations (II.3.5) and (II.3.7). The number of required eigenvectors can be estimated using their energy distribution (equation (II.3.11)). Finally, all samples to be classified are projected upon the same set of eigenvectors and all conditional pdfs are calculated using equation (II.3.4) and classified using equation (II.3.1).

Unfortunately, this method requires the construction of a matrix  $\mathbf{X}^{(i)}$  for each class  $\omega_i$  and projection has to be done upon the different sets of eigenvectors  $\mathbf{u}_j^{(i)}$ . If this is unwanted then a different approach can be used, namely by means of a canonical or factor analysis. This method calculates the covariance matrices of training samples within an identical class and between different classes. It then tries to find a projection which minimizes distances within classes, while at the same time maximizing distances between classes, thus requiring only a single projection. For more details we refer to Richards (1993) and for an application to Dumay and Fournier (1988).

### II.3.2.3 Conditional maximum likelihood classification: Bayes' optimum rule

The classifier equation (II.3.1) labels each data part based on its spectral components only, without any consideration for the classification of the elements in its neighborhood. In general, however, it can be assumed that some correlation exists between neighboring points, i.e. there will be some homogeneity in the spatial distribution of the different signal types. In a second iteration this correlation can be incorporated into the classification algorithm.

The simplest method would be to use some spatial averaging filter on all conditional pdf  $p(\omega_i|\mathbf{x})$  and to reapply equation (II.3.1) on the new distributions. However, we choose a different, though related, approach using Bayes' optimum rule. Contrary to Richards (1993), we propose a method which allows only for spatial correlations between identical phases, thus neglecting the fact that some relationship may exist between the different spectral classes (e.g. a reflection is more easily misclassified as a refraction than as ground roll). By quantifying these correlations, this effect can be corrected for.

Thus, after the conditional pdf  $p(\omega_i|\mathbf{x})$  have been calculated, they are multiplied with a weighting function incorporating the relations between different classes,  $p(\omega_k|\omega_i)$ , or the influence of a (predefined) neighborhood  $A_{\mathbf{x}}$  surrounding the data point,  $p(\omega_i|A_{\mathbf{x}})$ . Since we neglect any

relation between the different spectral classes, the optimum conditional pdf  $L_{\mathbf{x}}(\omega_i)$  are obtained from Bayes' rule, equation (II.3.2), using

$$L_{\mathbf{x}}(\omega_i) = p(\omega_i|A_{\mathbf{x}})p(\omega_i|\mathbf{x}) = p(\omega_i|A_{\mathbf{x}})p(\mathbf{x}|\omega_i)p(\omega_i)/p(\mathbf{x}), \quad (\text{II.3.12})$$

where we define

$$p(\omega_i|A_{\mathbf{x}}) = E\{p(\omega_i|\mathbf{y}), \mathbf{y} \in A_{\mathbf{x}}\}, \quad (\text{II.3.13})$$

yielding that the conditional pdf  $p(\omega_i|A_{\mathbf{x}})$  is the expectancy or mean of the likelihoods of class  $\omega_i$  for all data points  $\mathbf{y}$  in the neighborhood  $A_{\mathbf{x}}$ . If necessary, some weighting can be incorporated or several iterations can be used to update the likelihood distributions, since after each iteration the neighboring likelihood distributions have changed.

To reclassify each data point, equation (II.3.1) changes to

$$\mathbf{x} \in \omega_i \text{ if } L_{\mathbf{x}}(\omega_i) > L_{\mathbf{x}}(\omega_k) \quad \forall k \neq i. \quad (\text{II.3.14})$$

### II.3.2.4 Extraction

After classification has taken place, 'cleaned up' versions of the seismic section  $\phi(x, t)$  can be obtained by extracting the different classes. To extract these features a representation  $R(x, t)$  of their likelihoods for each data point (i.e. for each offset  $x$  and time  $t$ ) has to be obtained, which is then multiplied with the original section, yielding

$$\phi'(x, t) = \phi(x, t)R(x, t). \quad (\text{II.3.15})$$

Depending on the desired result this representation may yield either the conditional pdf of equation (II.3.2) or (II.3.12) for a particular class only, i.e. for class  $\omega_i$

$$R_i(x, t) = p(\omega_i|\mathbf{x}(x, t)) \text{ or } L_{\mathbf{x}(x, t)}(\omega_i). \quad (\text{II.3.16})$$

In addition, the classifier conditions (II.3.1) or (II.3.14) can be used as thresholds. In that case the likelihood of a specific class remains unchanged if the condition is satisfied, but is set to zero if not, i.e. for class  $\omega_i$

$$R_i(x, t) = \begin{cases} p(\omega_i|\mathbf{x}(x, t)) \text{ or } L_{\mathbf{x}(x, t)}(\omega_i) & \text{if } \omega_i \text{ is selected class} \\ 0 & \text{otherwise,} \end{cases} \quad (\text{II.3.17})$$

where  $\mathbf{x}(x, t)$  is the attribute vector obtained at a specific time and offset.

### II.3.3 Reconstruction method

Although many reconstruction techniques exist in remote sensing or more accurately in image processing (see e.g. Stark (1987)), we resorted to a technique which has already been used for a long time in geophysics and is both easy to implement and relatively fast. The local  $\tau$ - $p$  (intercept time - apparent dip or slowness) transform has already been used several times to extract laterally coherent signals (see e.g. Harlan et al. (1984) and Kong et al. (1985)).

The  $\tau$ - $p$  transform or slant stack is in mathematics known as the Radon transform and consists of a summation along 'slant lines' described by a specific slowness and intercept time. Thus, adjacent traces are incorporated in the extraction process to increase and facilitate detectability. This in contrast to the recognition method which does not make use of the information contained in the adjacent traces, but treats each trace separately. The drawback of the use of



the  $\tau$ - $p$  transform for reconstruction is, however, that no phase changes are allowed for, i.e. it is assumed that a signal has constant amplitude and phase along the slant lines.

To extract and reconstruct the coherent signals, the locally  $\tau$ - $p$  transformed section is multiplied with a coherency measure along the same slant lines, before being inverted back to the offset - time domain. This method allows for the interpolation and extrapolation of the coherent signals.

Following Kong et al. (1985) the seismic section  $\phi(x, t)$  is divided into  $N$  partly overlapping subsections  $\phi_n(x, t)$  of  $K$  traces each, which are transformed to the  $\tau$ - $p$  domain:

$$\psi_n(p_i, \tau_j) = \sum_{k=1}^K \phi_n(x_k, \tau_j + p_i x_k). \quad (\text{II.3.18})$$

Since the value  $\tau_j + p_i x_k$  will not necessarily be represented by a discrete time point, a 2-point linear interpolator in time will be used for its approximation. Moreover, edges may introduce artifacts in the transformed section (Yilmaz, 1987). Therefore, all edges are progressively weighted down. The exact weighting is of minor importance for the performance of the recognition and reconstruction algorithms.

Similar to equation (II.3.15), extraction of the desired features consists of the multiplication with some representation  $R(p_i, \tau_j)$  of coherency:

$$\psi'_n(p_i, \tau_j) = \psi_n(p_i, \tau_j) R_n(p_i, \tau_j). \quad (\text{II.3.19})$$

Applying the inverse  $\tau$ - $p$  transform on the summation of all subsections  $\psi'_n(p_i, \tau_j)$  then yields the 'cleaned up' seismic section  $\phi'(x_k, t_j)$ :

$$\phi'(x_k, t_j) = \frac{d}{dt} H \left[ \sum_{i=1}^{NP} \sum_{n=1}^N F_w(p_i, t_j - p_i x_k) \psi'_n(p_i, t_j - p_i x_k) \right]. \quad (\text{II.3.20})$$

The time derivative and the Hilbert transform  $H$  are required to prevent spectral distortion of the signal.  $NP$  represents the total number of slowness traces in the  $\tau$ - $p$  domain. The weighting function  $F_w(p_i, \tau_j)$  equals the reciprocal of the number of times a specific  $(p_i, \tau_j)$  has been calculated. It is used to calibrate the different subsections  $\psi'_n(p_i, \tau_j)$  and to correct for artifacts due to partial overlap. However, it is only of importance at the edges, since it is otherwise uniform. Moreover, to prevent the loss of important information at the edges, very often non-relevant traces and time samples are added at the edges.

What measure can now be best used to represent or quantify coherency? Kong et al. (1985) used smoothed semblance. Semblance is equivalent to the ratio of the coherent over the total energy:

$$S_n(p_i, \tau_j) = \frac{\left[ \sum_{k=1}^K \phi_n(x_k, \tau_j + p_i x_k) \right]^2}{K \sum_{k=1}^K \phi_n^2(x_k, \tau_j + p_i x_k)}. \quad (\text{II.3.21})$$

Smoothing is necessary to stabilize the semblance map and can be achieved by a rolling average along the  $\tau$ -axis of length  $W$ , representing the longest principal period present in the data. The smoothed semblance is then used as the representation of coherency in equation (II.3.19). Kong et al.'s method therefore uses semblance weighted slant stacks for the reconstruction of laterally

coherent signals. For more details, examples and optimization of parameter settings, the reader is referred to Kong et al. (1985). Yilmaz (1987) provides background on the  $\tau$ - $p$  transform.

It is, however, also possible to use different measures of coherency. In fact, it is possible to perform recognition directly in the  $\tau$ - $p$  domain. In that case, the attribute vector in the conditional pdf  $p(\omega_i | \mathbf{x}(\tau, p))$  or  $L_{\mathbf{x}}(\tau, p)(\omega_i)$  will represent the attributes (i.e. the local amplitude spectrum) of a data point in the  $\tau$ - $p$  domain. The representation  $R_n(p_i, \tau_j)$  will then be equal to the conditional pdfs for a specific spectral class for all intercept times and slownesses with possibly a conditional threshold, analogous to the extraction method in the  $x$ - $t$  domain discussed above, equations (II.3.16) and (II.3.17), respectively, where  $\mathbf{x}_{\text{now}}$  equals  $\mathbf{x}(\tau, p)$ . In this case, the recognition method can be seen as a flexible frequency analysis performed in the  $\tau$ - $p$  domain instead of in the  $x$ - $t$  domain.

### II.3.4 Data

The data used to show the effectiveness of the new method comes from the Ecos 2.5-D deep seismic reflection experiment (Marthelot et al., 1994) conducted within the western North Pyrenean thrust zone, an area of complex surface tectonics. The experiment was performed to test the feasibility of determining the 3-D geometry of deep structures within a crustal volume using current acquisition techniques from exploration geophysics. It is called 2.5-D, since it only consists of a single swath of a true 3-D exploration configuration.

The experiment included 6 parallel receiver lines of 120 receiver groups each with 14 shot lines perpendicular to the receiver lines using a symmetrical split spread configuration. The distances between lines, receivers and shots in a column were respectively 1600, 100 and 400 m. The lines and columns were moved by steps of 2 km in the direction of the lines. In total 292 shots were recorded (Marthelot et al., 1994; Siliqi, 1994).

The North Pyrenean Zone is limited to the north by the North Pyrenean Frontal Thrust and to the south by the steep North Pyrenean Fault. This zone extends along the whole length of the Pyrenean Range with widths from less than 20 km (central and eastern part) to 60 km (western Pyrenees). During the Early Cretaceous the North Pyrenean Zone was thinned and affected by significant subsidence, whereas in the Late Cretaceous and Tertiary it was compressed in response to the opening of the Bay of Biscay, the relative motion between Iberia and Europe and the formation of the Pyrenees. More geological background and maps are provided in Marthelot et al. (1994), Siliqi (1994) and Lecerf (1996).<sup>3</sup>

The experiment resulted therefore in a large dataset recorded in a region of complex tectonics. The size of the dataset made manual processing techniques nearly impossible or at least very expensive. As a consequence, any technique to increase the signal-to-noise ratio should be automatic or at least semi-automatic.

### II.3.5 Data application: recognition

#### II.3.5.1 Strategy

To detect and extract the coherent energy present in the seismic sections the following strategy was employed. Since we were mostly interested in reflections from the deeper parts of the crust, just a single spectral class was used, namely a picked reflection from the Moho.

Thus, in the first stage, the Moho-reflection is selected on several adjoining traces of a high quality common shot gather. Then, in the second (learning) stage, the characteristics (i.e. the means and variances of the attributes) of this specific phase are calculated using equations (II.3.5)

<sup>3</sup>et dans la section II.2.



and (II.3.7). We do not use the covariance matrix, since it will rarely happen that sufficient picks can be done to justify its estimation and thus its use. The attribute vectors contained the amplitude spectra of a 25 % cosine-tapered window centered around the picks (i.e. 25 % of the total window length is tapered). Energy contained in the amplitude spectra was normalized to correct for possible different amplitudes of signals. To prevent a degradation of detection results due to unaccounted for correlations between attributes, a principal component analysis has been used following equation (II.3.10). Since the amplitude spectra of the picked reflection displayed a very high similarity, only a single eigenvector was needed.

Next likelihood estimation or 'classification' takes place, i.e. the likelihood of the presence of the picked phase is calculated with equation (II.3.4) for all times and offsets using a sliding window to obtain the local amplitude spectra. In a second iteration the contributions of a predefined neighborhood is incorporated using equation (II.3.12) to readapt the likelihoods. This neighborhood has the shape of the wings of a butterfly centered around the considered sample to allow for static shifts and time dips.

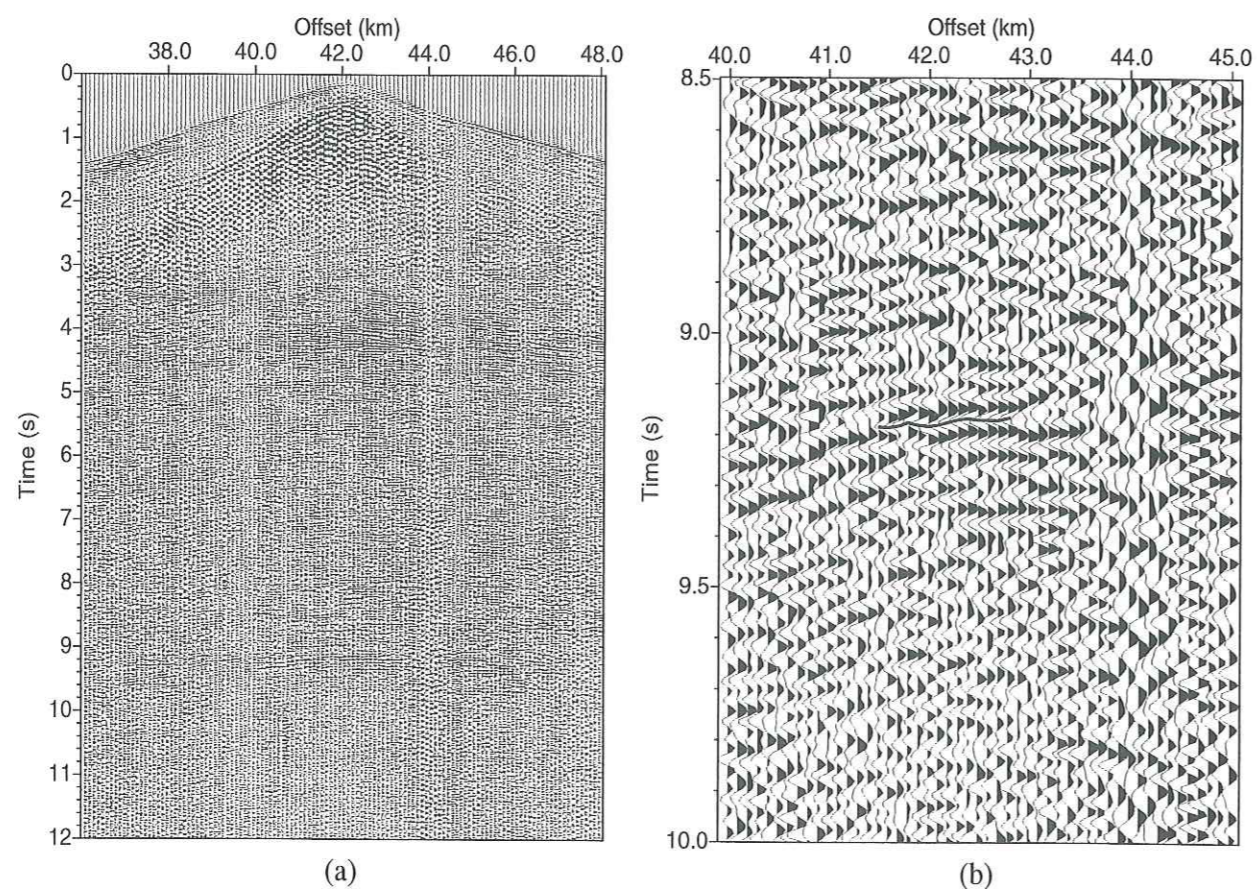


Figure II.3.1: CSG 8184. a) Raw data. b) Zoom-in showing the positions of the used picks (drawn line).

Finally, the obtained likelihoods are used as a representation of coherency in equation (II.3.15) to 'clean up' the seismic section and thus to extract the coherent energy. Before extraction the likelihood distributions are corrected for the employed windowing taper by means of a convolution of this taper with the likelihood distributions. Since only a single spectral class was used, all amplitude factors in equation (II.3.4) were neglected (including the *a priori*

distribution  $p(\omega_{refl}) = 1$ ) and only the product of exponentials was retained.

### II.3.5.2 Results

Figure II.3.1.a shows common shot gather (CSG) 8184<sup>4</sup> after frequency filtering and application of automatic gain control to increase visibility. Clearly visible are the top of the basement at 4 s and the Moho just below 9 s. Figure II.3.1.b shows a zoom-in on the same CSG from 8.5 to 10 s and offsets 40.0 to 45.0 km. It displays the positions of the 14 picks used to describe the Moho. A 128 ms (16 points) window is used for the calculation of the means and variances of the frequencies contained in the amplitude spectrum.

Picks were done on the Moho-reflection, because our initial aim was to extract it automatically in the whole dataset. However, the computation of the unconditional likelihood distributions of equation (II.3.4) without a principal component analysis had two unexpected results. Firstly, the method seemed to work too rigorously, since only the picked parts of the Moho-reflection were detected. Secondly, small parts of overlying reflections showed up in addition to large parts of the refraction branches. Seemingly, the spectral signature of deep reflections and refraction arrivals have a large similarity.

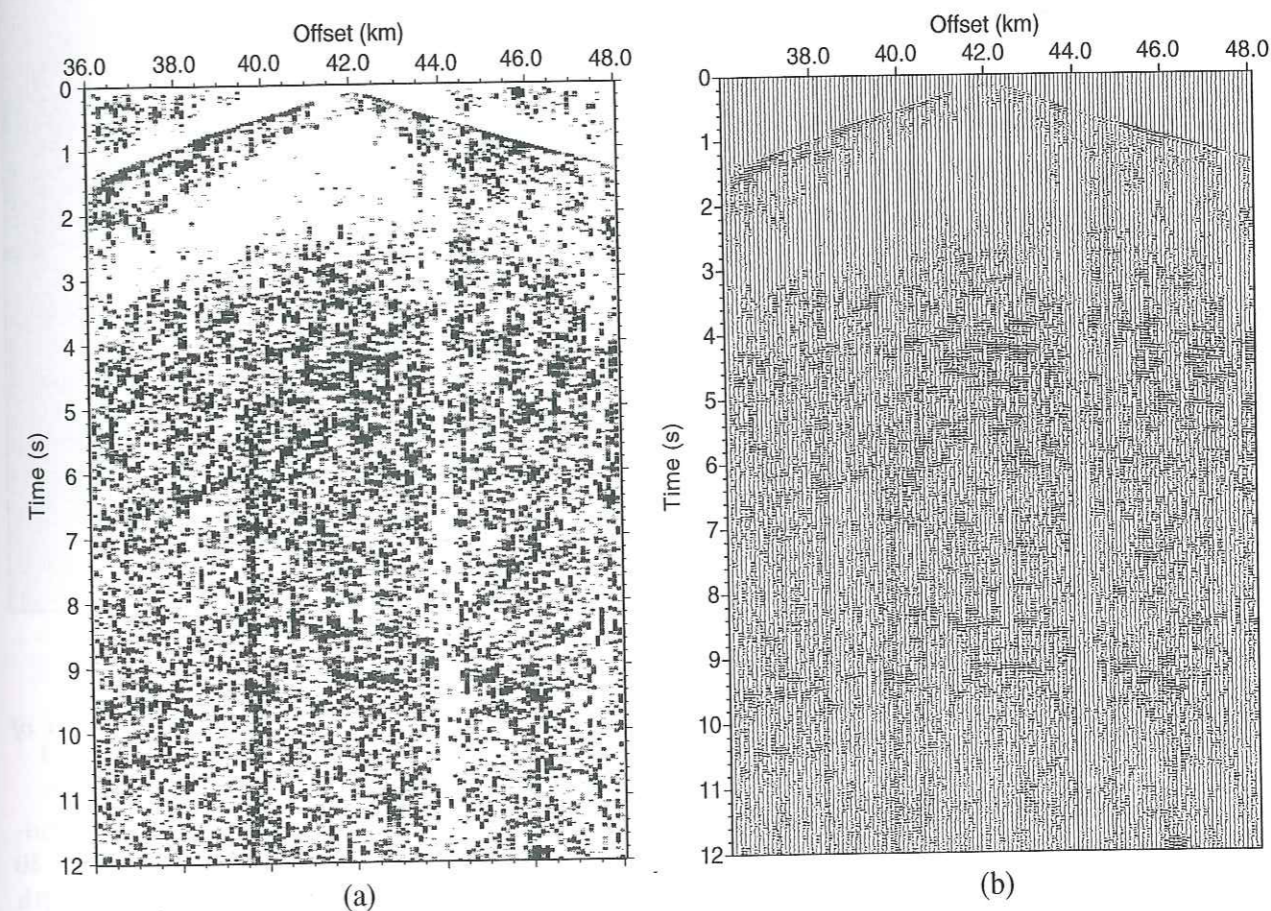


Figure II.3.2: CSG 8184. a) Likelihood distribution of the presence of reflection energy using a principal component analysis to reduce the number of attributes. b) Extracted data.

<sup>4</sup>La code 8184 signifie que ligne 4 des capteurs a enregistré point à tir 18 dans la 8<sup>e</sup> colonne de sources. Voir figure II.2.4.



Adapting the parameters of window length and tapering did not change the results significantly. To detect more energy, more, but questionable, picks of the Moho had to be included, though this was not the way we wanted to proceed.

To relax the strictness of the method, we applied a principal component analysis, thereby reducing the number of frequencies in the amplitude spectra to a single attribute, since the first eigenvector represented virtually 100 % of the total energy. After projection of the local amplitude spectra of the Moho-picks upon the first eigenvector and recalculating the mean and variance, this single attribute was used for likelihood estimation or 'classification' and thereafter extraction. Figures II.3.2.a and II.3.2.b show the results. Much more energy is now passed as being 'coherent'. In the extracted section nearly all reflections are visible.

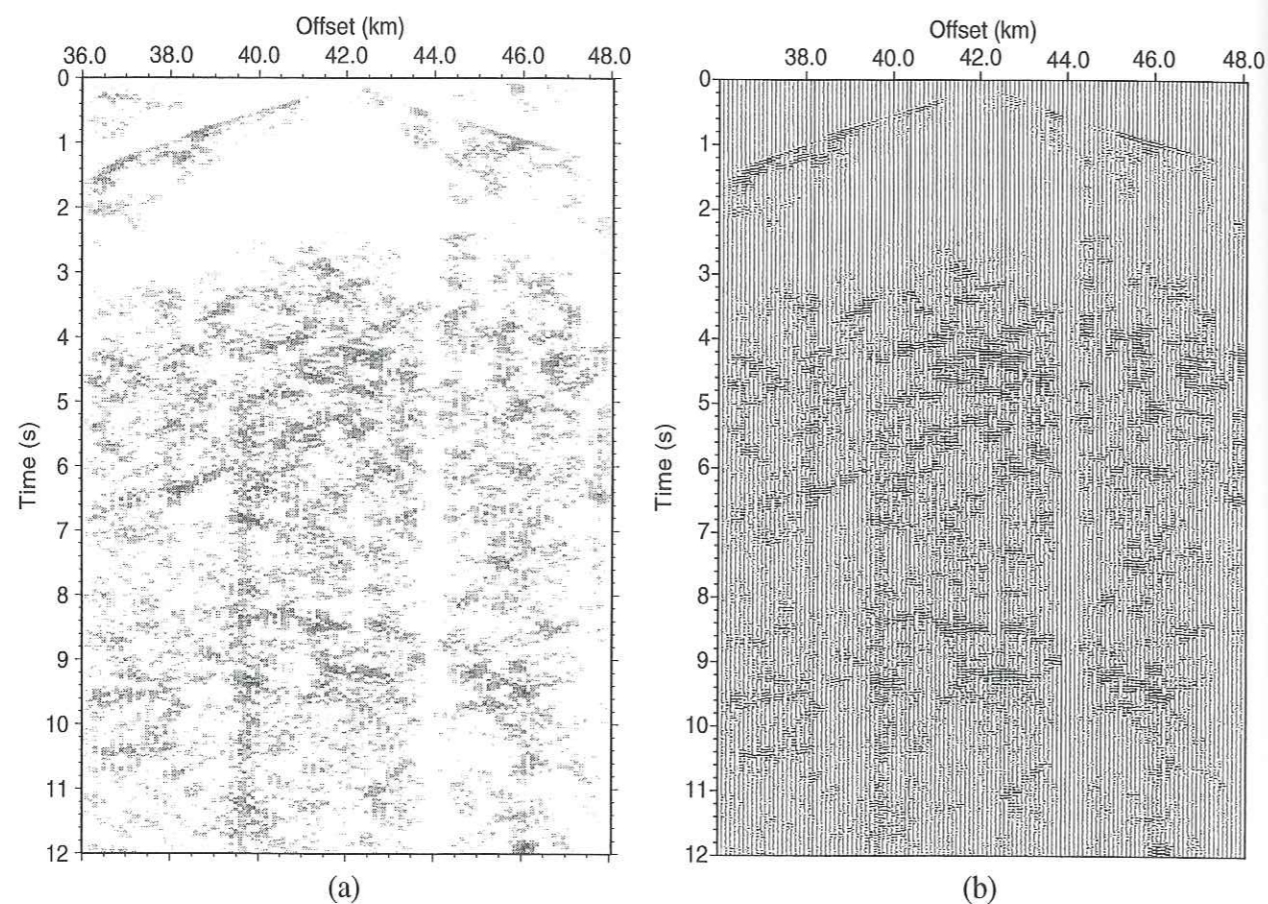


Figure II.3.3: CSG 8184. a) Likelihood distribution of reflection energy after incorporation of surrounding probabilities. b) Extracted data.

In Figures II.3.3.a and II.3.3.b the results are displayed after incorporation of the surrounding likelihoods. A neighborhood of twice two adjacent traces was used. Static shifts of  $\pm 40$  ms and linear move outs up to  $\pm 0.25$  s/km were allowed for, resulting in a neighborhood with the shape of the wings of a butterfly. Any lateral coherency is now more evident in both the likelihood distribution and the extracted section.

If the method has to be semi-automatic then the selected training set has to be representative not only for a single CSG, but also for a smaller or larger part of the stack profile. From extensive testing using CSGs up to several kilometers apart several conclusions could be drawn.

First of all, the calculated characteristics proved to be very robust, i.e. there is not much need to repick a phase to extract the coherent energy of other CSGs. However, if the coherent signals are already hard to detect with the eye, the algorithm will not detect them either, due to the fact that the trained eye of an interpreter will always be better able to detect coherent signals.

Nevertheless, trial and error tests have to be performed to check whether or not another set of picks can be obtained which results in better signal extractions and to verify to what extent the training set can extract signals from neighboring CMPs or CSGs.

Finally, extracted reflectors may exhibit a more or less fragmentary character, since no reconstruction of 'missing' energy can be done by the recognition technique. This topic will be addressed in the next section.

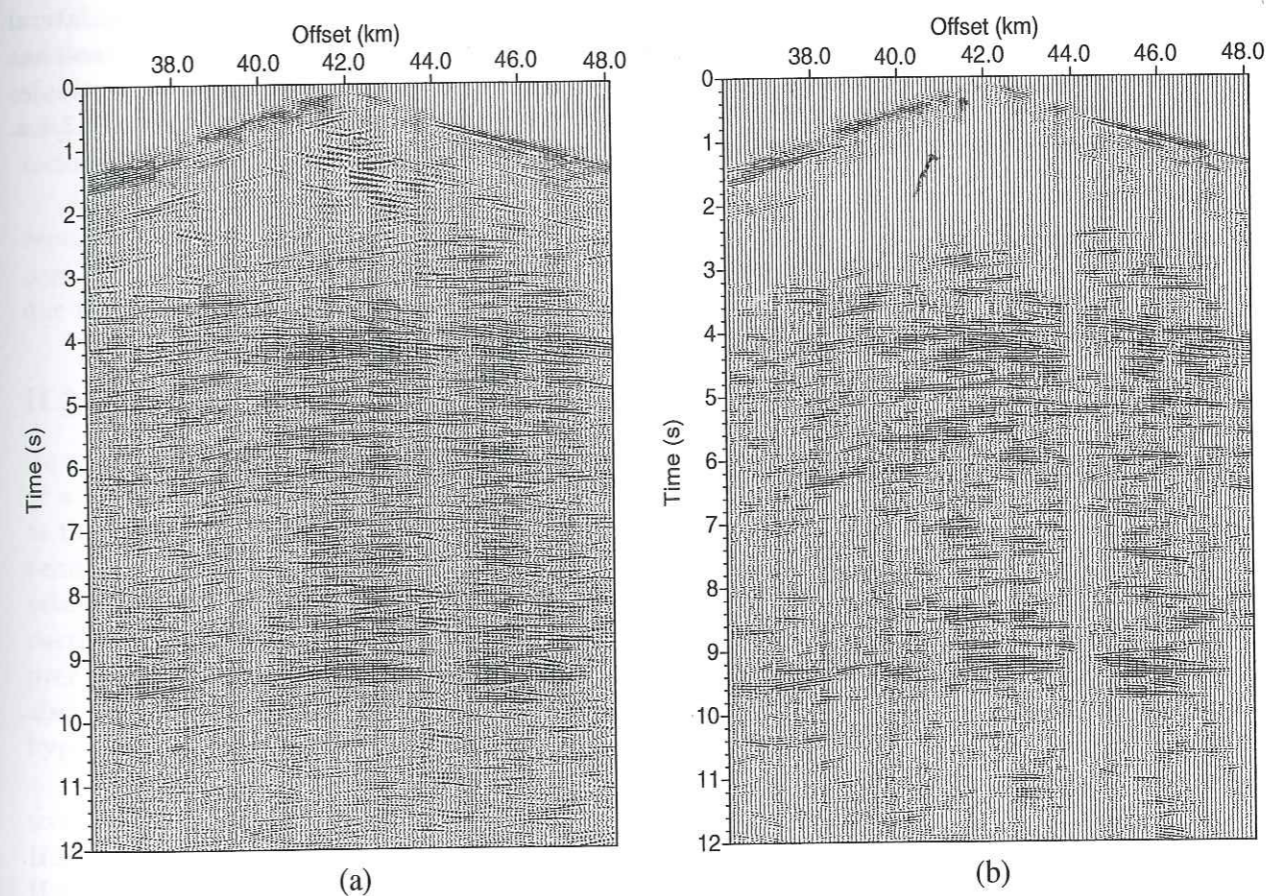


Figure II.3.4: CSG 8184. Reconstructed data with semblance weighting after two passes (with 5, resp. 10 traces). Using a) raw data (figure II.3.1.a) and b) extracted data (figure II.3.2.b).

### II.3.6 Data application: reconstruction

After having detected and extracted coherent energy, we want to explore some of the possibilities of reconstruction. As a strategy we adopt the philosophy that it is not correct to reconstruct energy directly over long offsets (i.e. many traces), but that an iterative approach is recommended. Thus, the number of traces will be gradually increased with iteration to interpolate small gaps first. Because of the enormous number of possible combinations only a few selected



examples will be shown. However, which combination or strategy works best for a specific data set can only be found out by trial and error.

To start with we show the results of the algorithm of Kong et al. (1985) using smoothed semblance as a representation of lateral coherence. This technique will emphasize any energy which is spatially coherent. The only tunable parameters are the range of slownesses, the number of traces used in the local  $\tau$ - $p$  analysis and the width  $W$  of the averaging filter. To prevent reconstructions of ground roll energy, slowness ranges of  $\pm 0.2$  s/km were used and  $W$  was put to 64 ms, although  $W$  has only a minor influence. Two consecutive passes were performed using respectively 5 and 10 traces in the  $\tau$ - $p$  transformation. Subsections were taken with overlaps of 50 %.

Figure II.3.4.a shows the result of the second pass. It is clearly visible that any lateral coherency has been stressed. Unfortunately, much undesired energy has remained or worse has been created. For example, some low frequency ground roll has remained. The worst examples of false reconstruction, however, can be found near the refraction branches in Figure II.3.4.a. There, energy has been created before the actual first arrivals.

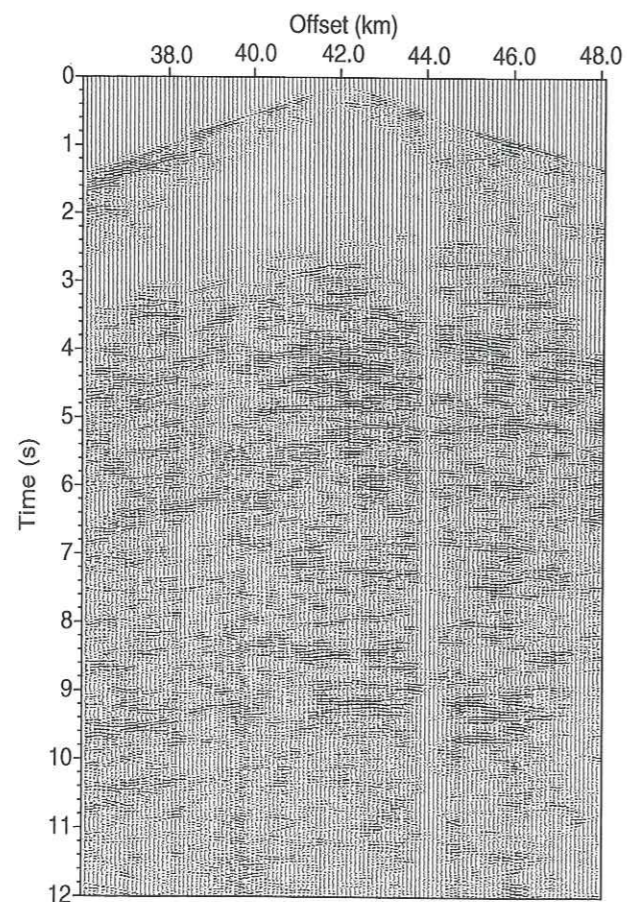


Figure II.3.5: CSG 8184. Reconstructed data using selection in  $\tau$ - $p$  domain. One pass: 5 traces.

Nevertheless, many of the artifacts can be removed if reconstruction is done after recognition. Figure II.3.4.b contains the reconstructed image of the extracted section displayed in Figure II.3.2.b using the same parameter settings as before. Clearly, the occurrence of many artifacts has been prevented, though some remain around the refraction branches. Applying the

reconstruction technique on the extracted section of Figure II.3.3.b, in which the influence of the surrounding likelihoods has been accounted for, gave even better results, since even more background noise has been filtered out of the data (not shown). However, changing the order and applying recognition after reconstruction resulted in considerably worse images. Although it removed some of the artifacts, many remained.

As an alternative, recognition and reconstruction can be performed in a single step by directly calculating the local amplitude spectra and the likelihood distributions in the  $\tau$ - $p$  domain and regarding these as the desired representation of coherency. Multiplying the corrected likelihoods with the  $\tau$ - $p$  transformed subsection and inverse-transforming to the  $x$ - $t$  domain resulted in Figure II.3.5, where subsections consisting of 5 traces were used.

This method produces high quality reconstructions which remain close to the original seismic section. Thus, nearly no artifacts have been introduced. A second pass using 10 instead of 5 traces per subsection, did not produce significantly different results. The image had become only slightly clearer. However, further reconstructions can be done using again the semblance technique of Kong et al. (1985).

A last conclusion which could be drawn from all reconstructions is the tendency of the repeated passes to emphasize more laterally coherent structures in the data than just a single pass. On the other hand, results from these consecutive passes tend to exhibit more artifacts due to the exaggerated extrapolations, like those displayed around the refraction branches.

### II.3.7 Data application: stacked sections

As a second example, we will show how the combined methods enhance the signal-to-noise ratio of a stacked section significantly. The stacked section is situated at exactly the same position as the common midpoints of CSG 8184 (Figure II.3.1.a). In general, stacked sections from deep seismic reflection data have a lower quality than those frequently encountered in exploration seismics. However, the Ecors 2.5-D case had an extraordinarily low stack quality. This can be partly attributed to the particular acquisition configuration, which resulted in low fold stacks over large offsets (in excess of 4 km). In addition, the complexity of the surface tectonics has also contributed to the stack degeneration (Lecerf, 1996). This led to fragmentary and non-hyperbolic move-outs of the reflections.

To increase stacking quality, half-offsets were limited to 3000 m and the bin size perpendicular to the receiver lines was reduced (Siliqi, 1994). Nevertheless, reflections are barely visible (Figure II.3.6.a). Though a semblance weighted reconstruction did improve the stacked profile (Figure II.3.6.b), it did not result in an easily interpretable section, since details remain unclear. Only a very global idea of the existing geology can be obtained. Slant stacks were done in a single pass on subsections consisting of 20 traces with 50 % overlap to emphasize the large structures in the data.

To increase the quality of the stacked section, recognition was applied before stack on all CMPs contributing to the stacked profile. To characterize the desired reflections, the same Moho-picks (Figure II.3.1.b) were used throughout the whole stacking analysis, showing again the stability of the recognition method. The resulting profile displayed, however, a very fragmentary image of the reflector positions. This can easily be explained by the fact already discussed that the extracted reflections in CSGs tend to have a more or less fragmentary character (see e.g. Figure II.3.2.b). Although already a clearer image of the reflector positions was obtained by applying reconstruction after stack on the stacked section, the clearest results were obtained using both recognition and reconstruction before stack.



Thus, recognition and thereafter reconstruction were applied on all CSGs contributing to the stacked profile. Pre-stack reconstructions were done in a single pass using subsections consisting of 10 traces with 50 % overlap. Since the stack tended to reintroduce non-coherent energy, the recognition technique was re-applied after the stacked section had been obtained (see Figure II.3.7.a for the resulting profile). If required, the reconstruction technique can also be re-applied on the obtained profile with the same parameter settings as used for Figure II.3.6.b, which resulted in Figure II.3.7.b.

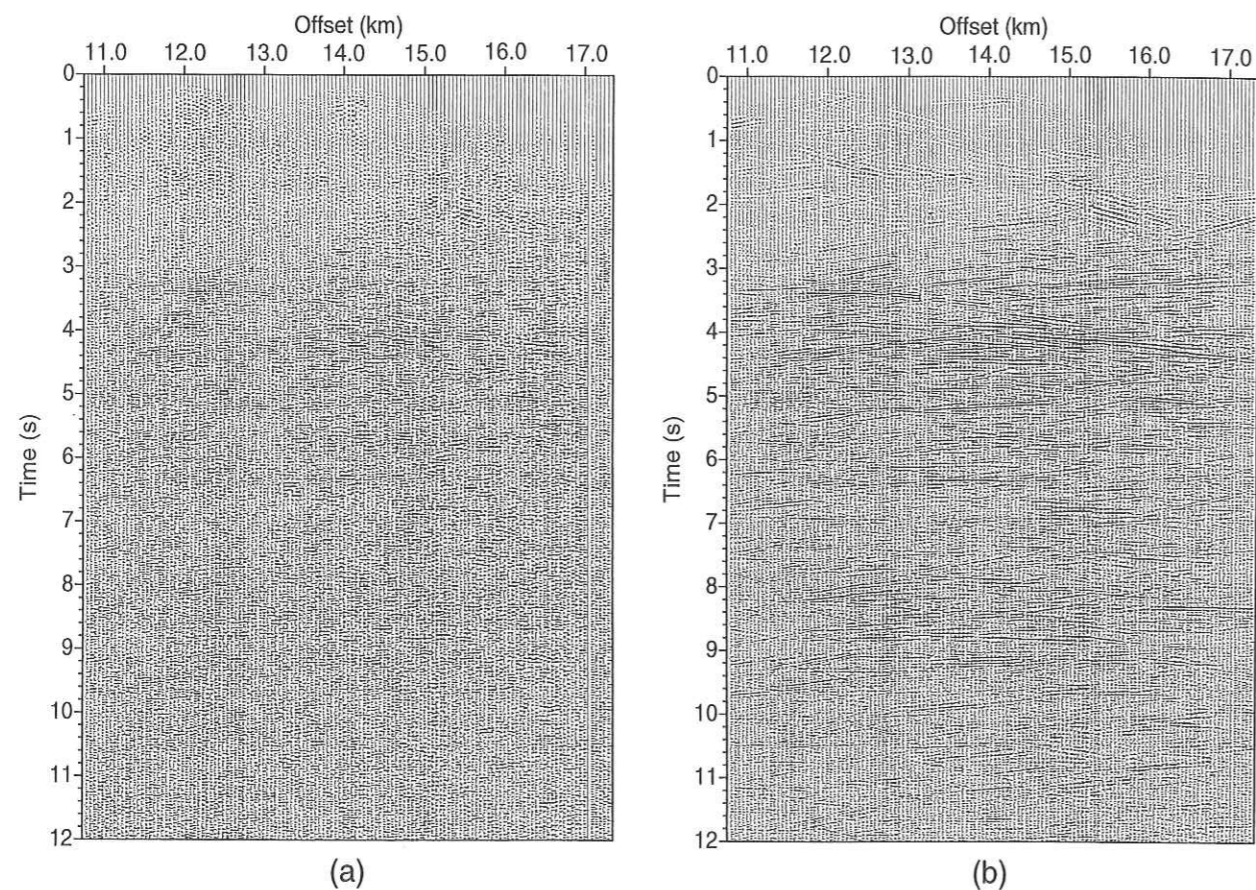


Figure II.3.6: Stack. a) Raw data. b) Associated reconstruction using semblance weighting.

Figures II.3.7.a and II.3.7.b give clear indications of the locations of the reflectors, thereby rendering interpretation much easier compared with the raw stacked sections (Figures II.3.6.a and II.3.6.b). It leaves no doubt that the signal-to-noise ratio has been increased significantly.

It should be noted that applying reconstruction both before and after stacking without the application of any recognition did also improve results, but to a much lesser extent as a large amount of background noise was left in the data. This conclusion could already be expected from a comparison of Figures II.3.4.a and II.3.4.b.

Finally, as a last test, reconstructions were done pre-stack on the CSGs using recognition in the  $\tau$ - $p$  domain (analogous to Figure II.3.5). Although final reconstruction results were better than Figure II.3.6.b, more background noise remained than in Figure II.3.7.b and some unexplained strongly dipping features could be seen interfering with the Moho-reflection (just below 9 s). It remained unclear if these strongly dipping features were actually present in the data and caused by geology or due to non-removed noise interference.

As a final remark, the strong resemblance between Figures II.3.7.b and Figure II.3.4.b should be noted, indicating the excellent reconstruction of CSG 8184.

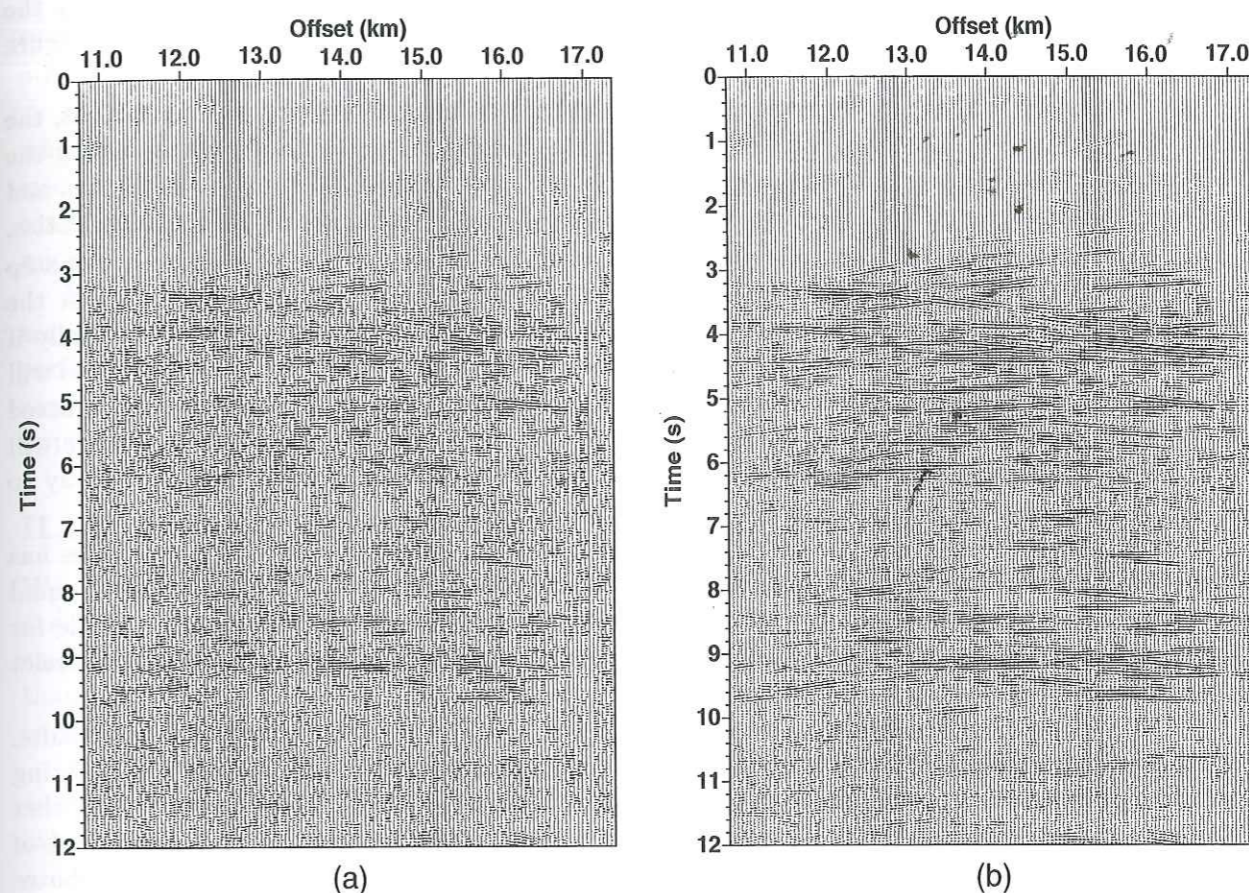


Figure II.3.7: Stack. a) Pre-stack recognition and reconstruction on CSGs and a second application of recognition after stack. b) Associated reconstruction using semblance weighting.

### II.3.8 Discussion

The combined recognition and reconstruction methods result in a very powerful technique, since it combines the power of the stacking process to increase the signal-to-noise ratio with the possibility to directly select a specific seismic phase.

It is this selection possibility that sets the method apart from nearly all other techniques to increase the signal-to-noise ratio. The skeletonization technique (Le and Nyland, 1990; Lu and Cheng, 1990) has nearly no selection possibilities at all. It will simply try to correlate waveforms laterally. Both the statistical method of Hansen et al. (1988) and the methods using the local  $\tau$ - $p$  transform (Harlan et al., 1984; Kong et al., 1985) can only select by changing the range and length of the slant lines. However, this will not always be adequate as shown in Figure II.3.4.a, where for example some surface wave energy remains. On the other hand, the performance of all techniques can be enhanced by means of accurate data processing.

At a first glance, it may seem remarkable that amplitude spectra of reflections picked at 9 s display a large similarity with those of reflections at 3 or 4 s. However, existing estimates of  $Q$  values in the region suggest that attenuation is mainly due to the upper sedimentary cover



with values of  $Q$  of 300 and less in the subsurface and 700 and higher in the underlying crust (Rappin et al., 1991; Rappin, 1992). Hence, spectra of reflections from the lower part of the sedimentary cover and below it do not evolve much with depth. Moreover, it explains why the refraction branches are better detected than the direct arrivals close to the source (e.g. Figure II.3.2.b).

An interesting side effect will be that in regions with strongly varying attenuation values, the recognition method will be able to detect and extract a specific reflection only, thus enabling the extracting of specific reflection horizons only. Otherwise, to extract reflections at all depths and offsets, several different reflections have to be picked to correct for changes in the characteristics.

The influence of some parameters has not yet been accounted for. The most important step in any supervised classification method is the selection phase and only to a lesser extent the learning phase. The selection of a representative set of picks is of vital importance, especially if it also has to represent the coherent energy of other CSG or CMP gathers. Clearly, there will be a trade-off between the internal consistency of the picked set and the amount of extracted energy. Or to speak in the terms of the statistical method of Hansen et al. (1988): there will be a trade-off between the 'false alarm rate' and the probability of a 'miss'. The best way to obtain a representative set is simply by trial and error.

The length of the sliding window in the local  $f$ - $t$  analysis for calculation of the attributes has to represent the characteristic scale-length of the selected seismic phase, which is at least equal to its largest principal period. We found, however, that the results remained quite stable for larger windows. As an alternative, a change of attributes can be considered, namely a wavelet transform may provide the desired characteristic scale-lengths.

The application of a principal component analysis greatly improved the detection results. It reduced the number of attributes to a single one, thereby correcting for possibly existing correlations between the attributes. Moreover, it produced superior detection results over other than Gaussian probability distributions which we tested. The simplest of these consisted of dividing each term within the exponentials of equation (II.3.4) with the number of used attributes  $J$  to correct for the influence of different window lengths and thus the number of attributes. This method may be applicable if more than a single eigenvector is used in the principal component analysis.

Combined recognition and reconstruction in the  $x$ - $t$  domain gives the best results. However, the reconstruction technique explicitly demands phase consistency along the stacking trajectory, whereas the recognition technique allows for random perturbations of the phase. Thus, the reconstruction technique seems to limit the possibilities of the more general recognition technique to some extent.

If reflection energy is not distributed on a straight line (e.g. due to statics), then semblance weighting may not be the optimum method, since a large amount of energy will be suppressed. In that case, the method of Harlan et al. (1984) as implemented by Bano (1989) may be a better solution. This method estimates the standard deviation of noise from histograms of a  $\tau$ - $p$  transformed part of the data containing noise only and uses thereafter twice this standard deviation as a threshold for all  $\tau$ - $p$  transformed data.

As an alternative to semblance weighted slant stacks, spatial prediction filtering (Hornbostel, 1991; Abma and Claerbout, 1995) in either  $t$ - $x$  or  $f$ - $x$  domain can be considered for signal reconstruction, since these techniques are also able to give an estimate for 'missing' energy. However, it remains explicitly assumed that events are at least locally linear and, furthermore, it is unclear if these techniques are able to deal with signals displaying phase perturbations. Thus, final results will probably display large similarities.

Some of the limitations implicit to the recognition method have to be thoroughly realized.

First of all, the recognition algorithm is basically a classification technique. This implies that, in the case of superposed signals, the algorithm will only be able to recognize the strongest signal or none at all and, thus, no reflections hidden by ground roll or excessive noise. Also very close to overlapping superpositions of two reflections may pose problems, e.g. in pinch-outs. On the other hand, reconstruction methods will probably solve some of these problems.

In addition, amplitudes are both adjusted by the recognition and reconstruction techniques, which may pose a problem if amplitude information becomes of fundamental importance for interpretation (e.g. AVO). Moreover, reconstructions may remove some of the existing small-scale features of reflections. On the other hand, the original sections always remain available for a joint interpretation.

Finally, we want to state that many of the artifacts visible in the reconstructed seismic sections can be prevented if additional processing is done, e.g. muting of traces and removal of surface waves. The results, however, show that the recognition technique is very stable even on raw data, because any processing was only done to increase visibility. Nevertheless, results will improve with some fine tuning.

### II.3.9 Conclusions

We have presented a new method to enhance the signal-to-noise ratio of a seismic section in a semi-automatic way using a supervised classification technique from remote sensing. The employed maximum likelihood technique allows for the detection and extraction of any energy that has the same spectral attributes as a selected seismic phase and should therefore be seen as a recognition technique.

Because the method is insensitive to phase perturbations of a signal, both diffractions and waves with polarity changes can be detected by the technique. Moreover, the recognition method can be applied both pre-stack and post-stack.

A shortcoming of the recognition method is the fact that it does not use the information contained in adjacent traces, but treats each trace separately. Thus, it cannot give an estimation of 'missing' energy. Therefore, the new method was combined with two reconstruction techniques using local  $\tau$ - $p$  transforms to combine the signal-to-noise ratio enhancement of the stacking process with the possibility to directly select only particular phases.

The first reconstruction method, using semblance weighted slant stacks, gave good results for both stack profiles and common shot gathers after extraction had been done. The second method, applying recognition directly in the  $\tau$ - $p$  domain, resulted in particular good reconstructions for common shot gathers, but worked less well for stacks. Both reconstruction techniques demand, however, phase consistency of the signal energy, thereby sacrificing some of the possibilities of the recognition technique.

To test the method, we applied it to the Ecors 2.5-D dataset which has a very low data quality as far as the lateral coherency of reflections and their ability to be stacked correctly is concerned. Our method succeeded in enhancing the signal-to-noise ratio in both individual shot gathers and in the stacked section. Figure II.3.7.b shows much more clearly than the brute stack of Figure II.3.6.a that the only laterally coherent reflections correspond to the fold and thrust system of the sedimentary cover (3 to 5 s) and the base of the crust at 9 s. In between, the whole crust (sedimentary cover excepted) is characterized by a large number of short reflection segments indicative of a strong heterogeneity at small scale.

The potential applications of the described method are manifold:

The recognition technique can be used as an interactive picking tool. That is, after picking only a part of a reflection, the interpreter will be able to pick more easily the remaining part



and other reflections in both common shot gathers and final stacks with the aid of the extracted sections.

Interpretation of stacked seismic sections can be facilitated, since the coherent energy (i.e. the reflectors) will be identified more easily in noisy data. In addition, it can be used for data segmentation (i.e. reduction of the data volume for storage).

Finally, the removal of non-coherent energy in common midpoint gathers prevents stack degeneration due to noise interference and in stacked sections it will prevent the occurrence of migration artifacts, like e.g. smiles (Kong *et al.*, 1985; Yilmaz, 1987). A limitation of the method is its amplitude fidelity. Care should be taken in using the method if the interpretation of true amplitude data is needed, since the recognition technique adjusts the amplitudes, and after reconstruction small-scale features of reflectors may have disappeared.

### II.3.10 Acknowledgments

The first author wishes to thank Kabir Roy-Chowdhury for discussions and Frédéric Pivot and Jean-Michel Marthelot for giving him an overview of all the difficulties associated with the dataset and for allowing him free access to their software. In addition, especially Georges Poupinet is thanked for urging each time to find a more generally applicable method and numerous discussions. Finally, we are grateful for the constructive reviews of Michel Dietrich, Jérôme Mars, Clark Trantham and especially Warren S. Ross.

## II.4 Réseaux neuronaux

### Résumé

Au cours de ces dernières années, les réseaux de neurones ont bénéficié d'une popularité croissante auprès des géophysiciens. En effet, ces outils mathématiques permettent d'obtenir une approximation d'une fonction continue et bornée avec une précision arbitraire. Par conséquent, ils peuvent apporter une contribution significative à la résolution de problèmes appartenant à une large gamme d'applications géophysiques.

Néanmoins, ils ne sont pas encore d'une utilisation courante dans le domaine de la géophysique. Ceci a pour conséquence que beaucoup des techniques développées récemment afin d'en faciliter la mise en œuvre et d'en augmenter les performances ne sont pas bien connues dans ce domaine-ci. Dans cette étude, nous décrivons des techniques permettant un conditionnement (apprentissage) plus rapide, une meilleure performance générale ainsi que des méthodes permettant une estimation automatique de la meilleure configuration et architecture du réseau.

La détection et l'extraction de 3 catégories différentes de signaux, à savoir les réflexions, les ondes de surface et les autres types de bruit, nous ont montré que les réseaux de neurones ont certains avantages sur une technique classique comme la méthode de Classification par Vraisemblance Maximale (CVM). La méthode CVM ne donne de bons résultats que si toutes les catégories de signaux sont caractérisées par une distribution Gaussienne (ou connue). Par contre, si la distribution sous-jacente est multi-modale ou asymétrique, des approximations successives sont nécessaires afin d'estimer la vraie distribution et d'éviter des mauvais résultats.

D'autre part, les réseaux de neurones sont beaucoup plus flexibles, car ils sont capables d'estimer la distribution sous-jacente de façon automatique, mais leur mise en œuvre est plus difficile.

### II.4.1 Les Réseaux de neurones et la détection des signaux: Introduction

Dans la section précédente, nous avons vu comment une technique classique de classification appartenant au domaine de la télédétection peut augmenter le rapport signal sur bruit. Bien que nous ayons obtenu de bons résultats, cette méthode a quelques inconvénients. Premièrement, la distribution statistique sous-jacente des signaux doit être connue ou déterminée en utilisant des approximations successives. Deuxièmement, elle utilise une transformée de Fourier à Courte Terme afin de caractériser au mieux les signaux. Malheureusement, la longueur optimale de la fenêtre d'analyse de cette transformée doit être spécifiée *a priori*, puisqu'elle n'est plus adaptée. Ce problème-ci sera traité dans la section suivante (II.5) où nous faisons appel à la Transformée Continue en Ondelette.

Dans cette section-ci, nous introduisons les Réseaux de Neurones (RN), puisqu'ils sont capables d'obtenir une approximation de la distribution statistique sous-jacente des signaux de façon automatique. Par conséquent, ils constituent une alternative intéressante à la méthode de Classification par Vraisemblance Maximale (CVM) dans le cas où il est très difficile de trouver cette distribution désirée.

Puisque les RN ne sont pas encore d'une utilisation courante, cette section-ci contient principalement une description étendue de ce type d'outils mathématiques et de plusieurs méthodes permettant d'améliorer leurs performances. Par conséquent, cette étude est aussi intéressante pour les lecteurs voulant utiliser les RN pour d'autres applications que la détection de signaux.

En une première lecture rapide et pour les lecteurs qui ne désirent pas directement utiliser des RN, la section II.4.3 donne une description courte du fonctionnement des RN et dans la section



II.4.5.2 les résultats obtenus avec les RN sont comparés avec ceux obtenus par la méthode CVM. Les autres sections sont écrites pour des utilisateurs actuels et futurs des RN. La section II.4.2 contient une introduction générale à l'article et la section II.4.4 décrit des techniques pour améliorer les performances des RN et pour faciliter leur usage. Finalement, II.4.5.1 donne une liste de points généraux de considération pour toute application des RN. De plus, le paragraphe 'Méthodes de classification' dans l'introduction générale II.1 contient aussi une description courte des RN et une comparaison entre la méthode CVM et les RN.

## On the use of neural networks in geophysical applications<sup>5</sup>

Mirko van der Baan<sup>a</sup> and Christian Jutten<sup>b</sup>

<sup>a</sup>Laboratoire de Géophysique Interne et Tectonophysique,  
Université Joseph Fourier and CNRS, BP 53, 38041 Grenoble Cedex 9, France.

<sup>b</sup>Laboratoire des Images et des Signaux, Institut National Polytechnique,  
46 av. Félix Viallet, 38031 Grenoble Cedex, France.

### Abstract

Neural networks enjoy an increasing popularity in geophysics. As they are universal approximators, these tools can approximate any continuous function with an arbitrary precision. Hence, they may yield important contributions to finding solutions to a large variety of geophysical applications.

However, knowledge of many methods and techniques recently developed to increase the performances and to facilitate the use of neural networks does not seem to be widespread in the geophysical community. Therefore, the power of these tools has not yet been explored to their full extent. In this paper, techniques are described for faster training, better overall performance, i.e. generalization, and the automatic estimation of network size and architecture.

### II.4.2 Introduction

Neural Networks (NN) have gained a large popularity this last decade in geophysics. They have been successfully applied to a large variety of problems. To name some of these restricted to the geophysical domain, neural networks have been used for waveform recognition and first break picking (Murat and Rudman, 1992; McCormack et al., 1993), electromagnetic (Poulton et al., 1992), magnetotelluric (Zhang and Paulson, 1997) and seismic inversion purposes (Röth and Tarantola, 1994; Langer et al., 1996; Calderón-Macías et al., 1998), shear wave splitting (Dai and MacBeth, 1994), well log analysis (Huang et al., 1996), trace editing (McCormack et al., 1993), seismic deconvolution (Wang and Mendel, 1992; Calderón-Macías et al., 1997), event classification (Dowla et al., 1990; Romeo, 1994) and in many other problems.

Nevertheless, the majority of these applications does not make use of more recently developed techniques which facilitate their use. Hence, expressions of the type 'designing and training a network is still more an art than a science' are not rare. The objective of this paper is to write a short introduction to these new techniques. However, it is not the idea to write a complete review covering the whole domain of NN types, since excellent reviews have already been written by Lippmann (1987), Hush and Horne (1993) and for readers fluent in French Hérault and Jutten (1994) and Chentouf (1997).

The above-mentioned statement that 'designing and training a network is still more an art than a science' is mainly due to several well known difficulties related to NN. Among these, the problem of the determination of the optimal net configuration (i.e. its structure), the optimal weight distribution of a specific network and the guarantee of a good overall performance (i.e. good generalization) are most eminent. In this paper, techniques will be described to tackle most of these well known difficulties.

A large variety of types of NN exist. Some of these have already been applied to geophysical problems. However, we will limit this tutorial to static, feedforward networks. Static implies that

5. Article soumis à Geophysics en 1999.



the weights, once determined, remain fixed and do not evolve with time, whereas feedforward indicates that the output is not feed back, i.e. refed, to the network. Thus, this type of network does not iterate to a final solution, but directly translates the input signals to an output independent of previous input.

Moreover, only supervised NN will be considered and in particular those suited for classification problems. Nevertheless, it should be noted that the same types of NN can also be used for function approximation and inversion problems (Poulton et al., 1992; Röth and Tarantola, 1994). Supervised classification mainly consists of three different stages (Richards, 1993), namely the selection, learning or training and classification stage. In the first stage, the number and nature of the different classes are defined and representative examples for each class are selected. Next, in the learning phase, the characteristics of each individual class have to be extracted from the training examples. Finally, all data can be classified using these characteristics.

Nevertheless, many other interesting networks do exist, which unfortunately have to fall beyond the scope of this paper. Other interesting NN include: the self-organizing map of Kohonen (1989), the Adaptive Resonance Theory (ART) of Carpenter and Grossberg (1987) and the Hopfield network (Hopfield, 1984) and other recurrent networks. See Lippmann (1987) and Hush and Horne (1993) for a partial taxonomy.

This paper starts with a short introduction to two types of static, feedforward NN and explains their general way of working. It then proceeds with a description of new techniques to increase performance and to facilitate their use. Next, a general strategy is described to tackle geophysical problems. Finally, some of these techniques are illustrated on a real data example, namely the detection and extraction of reflections, ground roll and other types of noise in a very noisy common shot gather of a deep seismic reflection experiment.

### II.4.3 Neural Networks: Structure and behavior

The mathematical perceptron was conceived some 55 years ago by McCulloch and Pitts (1943) to mimic the behavior of a biological neuron (Figure II.4.1.a). The biological neuron is mainly composed of 3 parts, namely the dendrites, the soma and the axon. A neuron receives an input signal from other neurons which are connected to its dendrites by the so-called synapses. These input signals are attenuated with an increasing distance from the synapses to the soma. The soma 'integrates' its received input (over time and space) and thereafter activates an output depending on the total input. The output signal is transmitted by the axon and distributed to other neurons by the synapses located at the tree structure at the end of the axon (Hérault and Jutten, 1994).

The mathematical neuron proceeds in a similar but simpler way (Figure II.4.1.b) as integration takes only place over space. The weighted sum of its inputs is fed to a non-linear transfer function (i.e. the activation function) to rescale the sum (Figure II.4.1.c). A constant bias  $\theta$  is applied to shift the position of the activation function independent of the signal input. Several examples of such activation functions are displayed in Figure II.4.1.d.

Historically, the Heaviside or hard-limiting function was used. However, this particular activation function gives only a binary output (i.e. 1 or 0 meaning yes or no). Moreover, it was found to be very difficult to estimate the optimum weights, since this particular function is not continuously differentiable. Thus, e.g. first order perturbation theory cannot be used. Nowadays, the sigmoid is mostly used. This is a continuously differentiable, monotonically increasing function that can be best described as a smooth step function (see Figure II.4.1.d). It is expressed by:  $f_s(\alpha) = (1 + e^{-\alpha})^{-1}$ .

To gain some insight in the working of static feedforward networks and their ability to deal

with classification problems, two of such networks will be considered: one composed of a single neuron and a second with a single layer of hidden neurons. Both networks will use a hard-limiting function for simplicity.

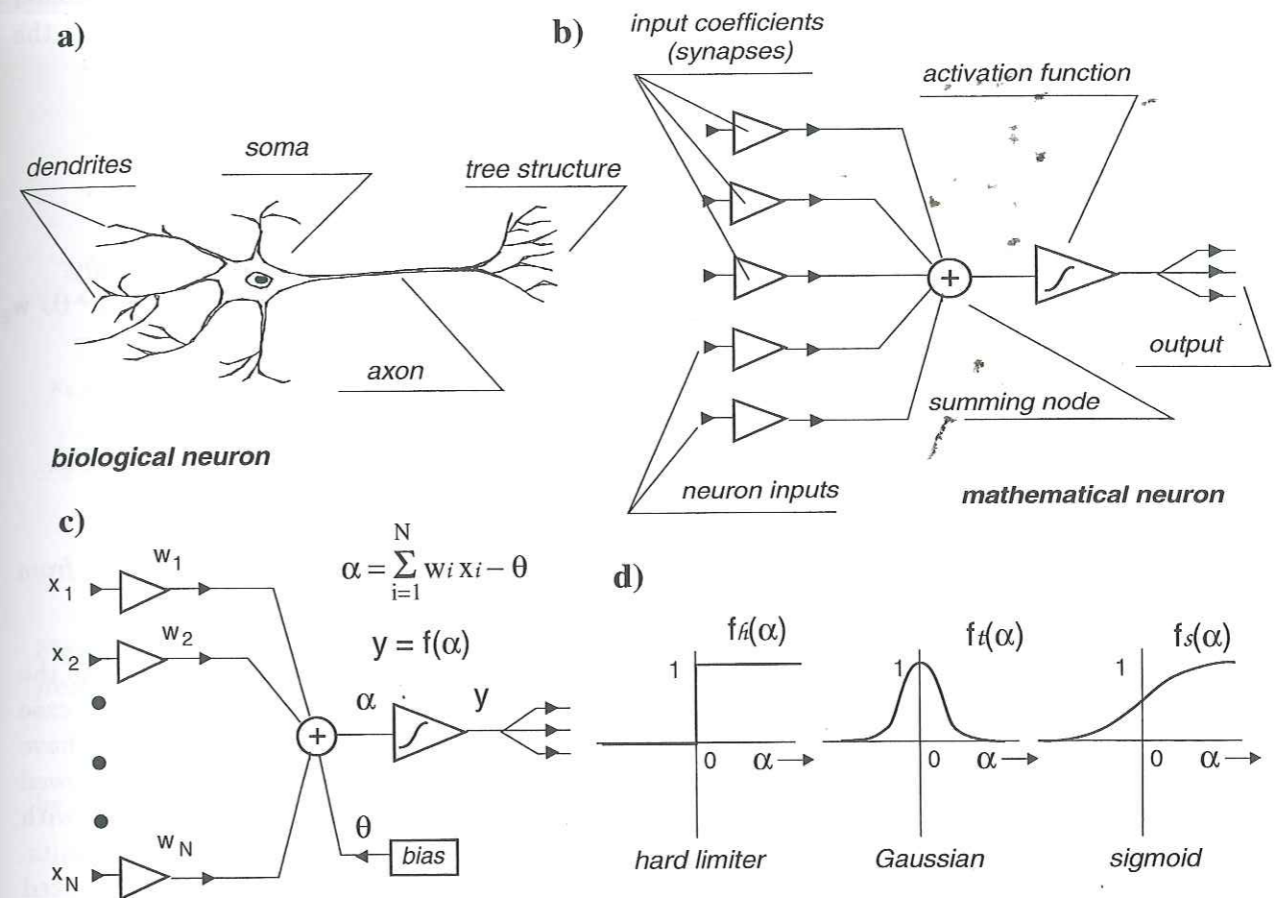


Figure II.4.1: The biological and the mathematical neuron. The mathematical neuron (b) mimics the behavior of the biological neuron (a). The weighted sum of the inputs are rescaled by an activation function (c) of which several examples are shown in (d). Adapted from Lippmann (1987), Hérault and Jutten (1994) and Romeo (1994).

Figure II.4.2.a displays a single neuron layer. Such a network will be able to classify data in two classes. For a 2-dimensional input this is done by separating the two distributions with a line (Figure II.4.2.b). In general, the two classes will be separated by an  $(n - 1)$ -dimensional hyperplane for an  $n$ -dimensional input.

More complex distributions can be handled if a 'hidden' layer of neurons is added. Such layers lie between the input and output layer, connecting them indirectly. However, the general way of working does not change at all as is shown in Figures II.4.3.a and II.4.3.b. Again, each neuron in the hidden layer divides the input space in two halfspaces. Finally, the last neuron combines these to form a closed shape or subspace. With the addition of a second hidden layer, quite complex shapes can be formed (Romeo, 1994). See also Figure 14 in Lippmann (1987).

Using a sigmoidal instead of a hard-limiting function does not change the general picture, the transitions between classes are smoothed. On the other hand, the use of a Gaussian activation function does implicate major changes, since it has a localized response. Hence, the sample



space will be divided in two parts. The part close to the center of the Gaussian with large outputs is enveloped by the subspace at its tails showing small output values. Thus, only a single neuron with a Gaussian activation function and constant variance is needed to describe the gray class in Figure II.4.3 instead of the depicted 3 neurons with hard-limiting or sigmoidal activation functions. Moreover, the Gaussian will place a perfect circle around the class in the middle (if a common variance is used for all input parameters).

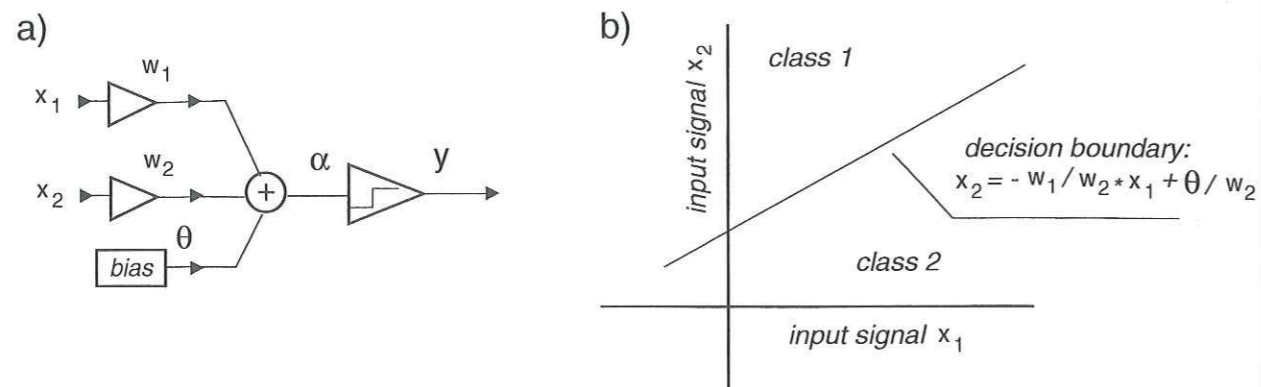


Figure II.4.2: a) Single perceptron layer and b) associated decision boundary. Adapted from Romeo (1994).

This insight into the general way NN proceed to solve classification problems will enable the user to obtain a first notion of the structure required for his particular application. In the case of very complicated problems with e.g. skewed, multi-modal distributions, he will probably have to choose for NN structures with 2 hidden layers. Notice, however, that Cybenko (1989) showed that NN using sigmoids are able to approximate asymptotically any continuous function with an arbitrary close precision using only a single non-linear, hidden layer and linear output units. Similarly, Park and Sandberg (1991) showed that, under mild conditions, NN with localized activation functions (such as Gaussians) are also universal approximators. Unfortunately, neither theorem is able to predict the exact number of neurons needed as these are asymptotic results. Moreover, applications exist where NN with 2 hidden layers produce similar results as a single hidden layer NN with a strongly reduced number of links and, therefore, a less complicated weight optimization problem, i.e. training will be much easier (Chentouf, 1997).

Note that in Figure II.4.1.d two types of activation functions have been used. The hard-limiter and the sigmoid are monotonically increasing functions, whereas the Gaussian has a localized activation. Both types are commonly used in NN applications. In general, NN with monotonically increasing activation functions are called Multi-Layer Perceptrons (MLP) and NN with localized activation functions Radial Basis Functions (RBF).

Hence, MLP networks with one output perceptron and a single hidden layer are described by

$$f_{MLP}(\mathbf{x}) = \sigma \left( \sum_{k=1}^{n_{h1}} w_k \sigma(\mathbf{w}^{(k)} \cdot \mathbf{x} - \theta^{(k)}) - \theta \right) \quad (\text{II.4.1})$$

with  $\sigma(\cdot)$  the sigmoidal activation function,  $w_k$  the weight of link  $k$  to the output node,  $n_{h1}$  the number of nodes in the hidden layer,  $\mathbf{w}^{(k)}$  the weights of all links to node  $k$  in the hidden layer and  $\theta$  the biases. Equation (II.4.1) can easily be extended to contain several output nodes and more hidden layers.

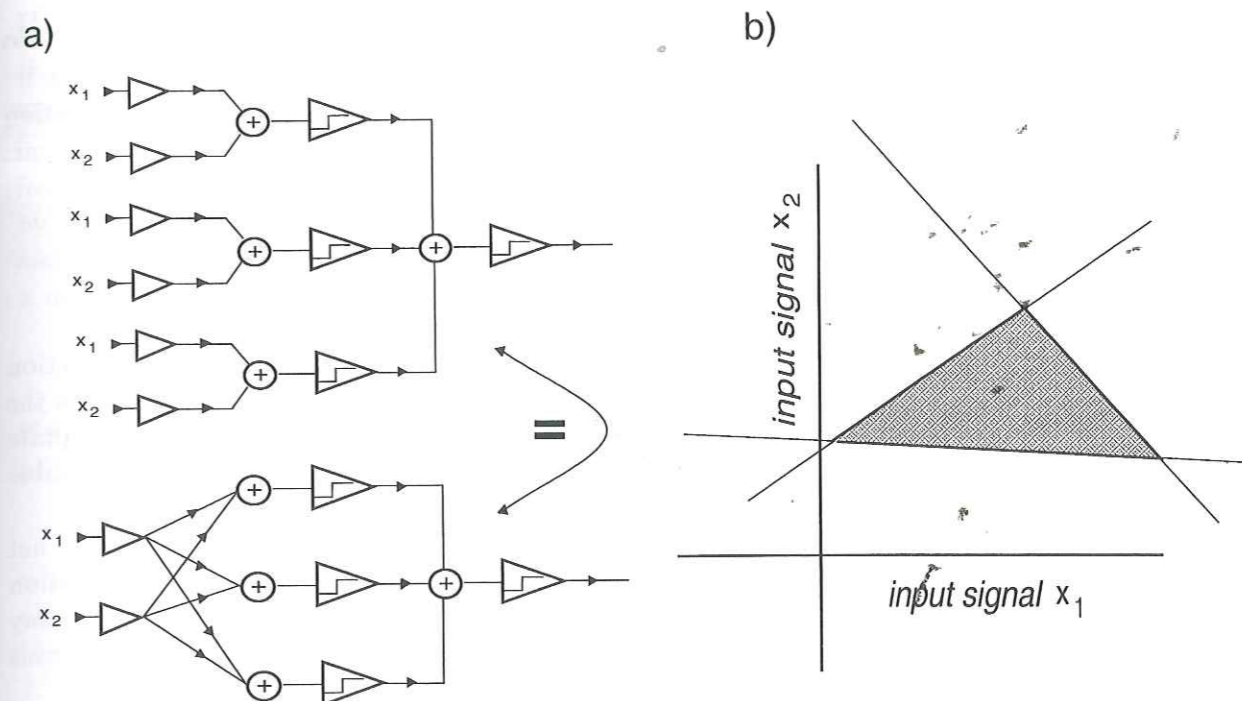


Figure II.4.3: a) Single hidden perceptron layer and b) associated decision boundary. Adapted from Romeo (1994).

Likewise, RBF networks with a single hidden layer and one output perceptron are described by

$$f_{RBF}(\mathbf{x}) = \sigma \left( \sum_{k=1}^{n_{h1}} w_k \mathcal{K}(s_k \| \mathbf{x} - \mathbf{c}^{(k)} \|) - \theta \right) \quad (\text{II.4.2})$$

with  $\mathcal{K}(\cdot)$  the localized activation function,  $\| \cdot \|$  a (distance) norm,  $\mathbf{c}^{(k)}$  the center of the localized activation function in hidden node  $k$  and  $s_k$  its associated width (spread).

It is important to be aware of the total number  $n_{tot}$  of internal variables determining the behavior of the used NN structure as we will show hereafter. Fortunately, this number is very easy to calculate from equations (II.4.1) and (II.4.2). For MLP networks it is composed of the number of links plus the number of perceptrons to incorporate the number of biases. If  $n_i$  denotes the number of input variables,  $n_{hi}$  the number of perceptrons in the  $i^{\text{th}}$  hidden layer and  $n_o$  the number of output perceptrons then  $n_{tot}$  is given by

$$n_{tot} = (n_i + 1) * n_{h1} + (n_{h1} + 1) * n_o \quad (\text{II.4.3})$$

for an MLP with a single hidden layer and

$$n_{tot} = (n_i + 1) * n_{h1} + (n_{h1} + 1) * n_{h2} + (n_{h2} + 1) * n_o \quad (\text{II.4.4})$$

for an MLP with 2 hidden layers. The number of internal variables is exactly equal for isotropic RBF networks, since each Gaussian is described by  $n_i + 1$  variables, namely its position and variance, i.e. width. Moreover, in this article only RBF networks with a single hidden layer will be considered. In addition, only RBF neurons in the hidden layer will have Gaussian activation



functions. The output neurons will have sigmoids as activation functions. Hence,  $n_{tot}$  is also given by equation (II.4.3).

As we will see hereafter, the ratio  $n_{tot}/m$  determines if an adequate network optimization can be hoped for, where  $m$  defines the number of training samples.

## II.4.4 Network optimization

### II.4.4.1 Known problems

The two most important steps in the application of NN to recognition problems are the selection and learning stages, since these are of direct influence on the overall performance and thus the results obtained. Three reasons can cause a bad performance (Romeo, 1994): an inadequate net configuration, the training algorithm being trapped in a local minimum or an unsuitable learning set.

Let us start with the net configuration. As was shown in Figures II.4.2 and II.4.3, the net configuration should allow for an adequate description of the underlying statistical distribution of the spread in the data. Since the number of input and output neurons is fixed in many applications, our main concern is with the number of hidden layers and the number of neurons therein.

No rules exist for determining the exact number of neurons in a hidden layer. However, Huang and Huang (1991) showed that the upper bound of number of neurons needed to reproduce exactly the desired outputs of the training samples is on the order of  $m$ , the number of training samples. Thus, the number of neurons in the hidden layer should never exceed the number of training samples. Moreover, to keep the training problem overconstrained, the number of training samples should always be larger than the number of internal weights. In practice,  $m \approx 10n_{tot}$  is considered to be a good choice. Hence, care should be taken that the number of neurons is kept limited. Otherwise the danger exists that the training set is simply 'memorized' by the net (overfitting). Classically, the best configuration is found by trial-and-error, starting with a small number of nodes.

A second reason why the net may not obtain the desired results, is that it may become trapped in a local minimum. The misfit function is very often extremely complex (Hush et al., 1992). Thus, the net can easily be trapped in a local minimum instead of attaining the searched for global one. In that case even the training set cannot be properly fit.

Remedies are simple. Either several minimization attempts have to be done using each time a different (random or non-random) initialization of the weights or other inversion algorithms have to be considered as e.g. global search.

Finally, problems can occur with the selected training set. The two most frequent problems are overtraining and a bad, i.e. not representative, learning set. In the latter case, either too many bad patterns are selected (i.e. patterns attributed to the wrong class) or the training set does not allow for a good generalization. For instance, the sample space may be incomplete, i.e. samples needed for an adequate training of the network are simply missing.

Overtraining of the learning set may also pose a problem. Overtraining means that the selected training set is memorized, such that performance is only excellent on this set, but not on other data. To circumvent this problem, the selected set of examples is often split in a training and a validation set. Weights are optimized using the training set. However, cross-validation with the second set ensures an overall good performance.

In the following subsections, all of these problems will be considered in more detail and several techniques will be described to facilitate the use of NN and to enhance their performance.

### II.4.4.2 Network training/Weight estimation: An optimization problem

If a network configuration has been chosen, an optimal weight distribution has to be estimated. This is an inversion or optimization problem. The most common procedure is a so-called localized inversion approach. In such an approach it is first assumed that the output  $\mathbf{y}$  can be calculated from the input  $\mathbf{x}$  using some kind of function  $\mathbf{f}$ , i.e.  $\mathbf{y} = \mathbf{f}(\mathbf{x})$ . Output may be contaminated by noise, which is assumed to be uncorrelated to the data and to have zero mean. Next, it is assumed that the function can be linearized around some initial estimate  $\mathbf{x}_0$  of the input vector  $\mathbf{x}$  using a first order Taylor expansion, i.e.

$$\mathbf{y} = \mathbf{f}(\mathbf{x}_0) + \frac{\partial \mathbf{f}(\mathbf{x}_0)}{\partial \mathbf{x}} \Delta \mathbf{x}. \quad (\text{II.4.5})$$

If we write  $\mathbf{y}_0 = \mathbf{f}(\mathbf{x}_0)$ ,  $\Delta \mathbf{y} = \mathbf{y} - \mathbf{y}_0$  and  $\mathbf{A}^{(x)} = \partial \mathbf{f} / \partial \mathbf{x}$ , equation (II.4.5) can also be formulated as

$$\Delta \mathbf{y} = \mathbf{A}^{(x)} \Delta \mathbf{x}, \quad (\text{II.4.6})$$

where the Jacobian  $\mathbf{A}^{(x)} = \nabla_{\mathbf{x}} \mathbf{f}$  contains the first partial derivatives with respect to  $\mathbf{x}$ . To draw an analogy with a better known inversion problem, in a tomography application  $\mathbf{y}$  would contain the observed travel times,  $\mathbf{x}$  the desired slowness model and  $A_{ij}$  the path lengths of ray  $i$  in cell  $j$ .

However, there exists a fundamental difference with a tomography problem. Namely, in an NN application, both the output  $\mathbf{y}$  and the input  $\mathbf{x}$  are known, since  $\mathbf{y}^{(i)}$  represents the desired output for training sample  $\mathbf{x}^{(i)}$ . Hence, the problem is not the construction of a model  $\mathbf{x}$  explaining the observations, but the construction of the approximation function  $\mathbf{f}$ . As this function is described by its internal variables, it is another linear system which has to be solved, namely

$$\Delta \mathbf{y} = \mathbf{A}^{(w)} \Delta \mathbf{w}, \quad (\text{II.4.7})$$

where the Jacobian  $\mathbf{A}^{(w)} = \nabla_{\mathbf{w}} \mathbf{f}$  contains the first partial derivatives with respect to the internal variables  $\mathbf{w}$ . The vector  $\mathbf{w}$  will contain the biases and weights for MLP networks and the weights, variances and centers for RBF networks. For the exact expression of  $\mathbf{A}^{(w)}$ , we refer to Hush and Horne (1993) and Héroult and Jutten (1994). Nevertheless, it should be noted that all expressions can be analytically calculated. Moreover, both the sigmoid and the Gaussian are continuously differentiable, which is the ultimate reason for their use. Thus, no first order perturbation theory has to be applied to obtain estimates of the desired partial derivatives, implying a significant gain in computation time for large NN.

In general, the optimization problem will be ill posed, since  $\mathbf{A}^{(w)}$  suffers from rank deficiency, i.e.  $\text{rank}(\mathbf{A}^{(w)}) \leq n_{tot}$ . Thus, system (II.4.7) is underdetermined. However, at the same time, any well formulated inversion problem will be overconstrained, because  $m \gg n_{tot}$ , yielding that there are more training samples than internal variables.

Due to the fact that system (II.4.7) is ill posed, a nullspace will exist. Hence, the internal variables cannot be uniquely determined. If, in addition,  $n_{tot} \gg m$  then the danger of overtraining, i.e. memorization, will be considerably increased, resulting in non-optimal performance. Two reasons cause  $\mathbf{A}$  to be rank deficient. First, the sample space may be incomplete, i.e. some samples needed for an accurate optimization are simply missing and some training samples may be erroneously attributed to a wrong class. Second, noise contamination will prevent a perfect fit of both provided and non-provided data. For example, in a tomographic problem, rank deficiency will already occur if none visited cells are present, making a correct estimate of the true velocities in these cells impossible.



To give an idea of the number of training samples required, the theoretical study of Baum and Haussler (1989) showed that for a desired accuracy level of  $(1 - \epsilon)$ , at least  $n_{tot}/\epsilon$  examples must be provided, i.e.  $m \geq n_{tot}/\epsilon$ . Thus, to classify 90 % of the data correctly, at least 10 times more samples have to be provided than internal variables are present, i.e.  $m \geq 10n_{tot}$ .

How can we solve equation (II.4.7)? A possible method of estimating the optimal  $\mathbf{w}$  is by minimizing the sum of the squared differences between the desired and the actual output of the network. This leads to the least-mean-squares solution, i.e. the weights are determined by solving the normal equations

$$\Delta \mathbf{w} = (\mathbf{A}^t \mathbf{A})^{-1} \mathbf{A}^t \Delta \mathbf{y}, \quad (\text{II.4.8})$$

where the superscript ( $w$ ) has been dropped for clarity.

This method, however, has the well known disadvantage that singularities in  $\mathbf{A}^t \mathbf{A}$  will cause the divergence of the Euclidean norm  $|\Delta \mathbf{w}|$  of the weights, since this norm is inversely proportional to the smallest singular value of  $\mathbf{A}$ . Moreover, if  $\mathbf{A}$  is rank deficient then this singular value will be zero or at least effectively zero due to a finite machine precision. The squared norm  $|\Delta \mathbf{w}|^2$  is also often called the variance of the solution.

To prevent divergence of the solution variance, very often a constrained version of equation (II.4.8) is constructed using a positive damping variable  $\beta$ . This method is also known as Levenberg-Marquardt or Tikhonov regularization, i.e. system (II.4.8) is replaced by

$$\Delta \mathbf{w} = (\mathbf{A}^t \mathbf{A} + \beta \mathbf{I})^{-1} \mathbf{A}^t \Delta \mathbf{y}, \quad (\text{II.4.9})$$

with  $\mathbf{I}$  the identity matrix (Lines and Treitel, 1984; Van der Sluis and Van der Vorst, 1987).

The matrix  $\mathbf{A}^t \mathbf{A} + \beta \mathbf{I}$  is not rank deficient in contrast to  $\mathbf{A}^t \mathbf{A}$ . Hence, the solution variance does not diverge, but remains constrained. Nevertheless, the method comes at an expense. Namely, the solution will be biased due to the regularization parameter  $\beta$ . Therefore, it does not provide the optimal solution in a least-mean-squares sense. The exact value of  $\beta$  has to be judiciously chosen to optimize the trade-off between variance and bias (*cf.* Van der Sluis and Van der Vorst, 1987; Geman et al., 1992).

More complex regularization can be used on both  $\Delta \mathbf{w}$  and  $\Delta \mathbf{y}$ . For instance, if uncertainty bounds on the output are known (e.g. their variances), then these can be used to rescale the output. A similar rescaling can also be applied on the input and/or weights. This method allows for the incorporation of any *a priori* information available. Hence, a complete Bayesian inversion problem can be formulated. See Tarantola (1987) for more details on this approach in general.

Just like in tomography problems, the normal equations (II.4.8) and (II.4.9) are rarely directly solved. More often an iterative approach is applied. The best known method in NN applications is the gradient backpropagation method of Rumelhart et al. (1986) with or without a momentum term, i.e. a term analogous to the function of the damping factor  $\beta$ . It is a so-called first order optimization method which approximates  $(\mathbf{A}^t \mathbf{A})^{-1}$  in equations (II.4.8) and (II.4.9) by  $\alpha \mathbf{I}$  with  $\beta = 0$ .

This method is basically a steepest descent algorithm. Hence, all disadvantages of such gradient descent techniques apply. For instance, in the case of curved misfit surfaces, the gradient will not always point to the desired global minimum. Therefore, convergence may be slow (*cf.* Lines and Treitel, 1984). To accelerate convergence, the calculated gradients are multiplied with a constant factor  $\alpha$ ,  $0 < \alpha < 2$ . However, a judicious choice of  $\alpha$  is required, since non-optimal choices will have exactly the opposite effect, i.e. convergence is slowed down even further. For instance, if  $\alpha$  is chosen too large, strongly oscillating misfits are obtained which do not converge

to a minimum, whereas choosing too small a value will not only slow down convergence, but may also hinder the escape from very small local minima. Furthermore, convergence is not guaranteed within a certain number of iterations. In addition, previous ameliorations in the misfit can be partly undone by the next iterations.

Although several improvements have been proposed concerning adaptive modifications of both  $\alpha$  (Dahl, 1987; Jacobs, 1988; Riedmiller and Braun, 1993) and more complex regularization terms (Hanson and Pratt, 1989; Weigend et al., 1991; Williams, 1995), the basic algorithm remains identical. Fortunately, other algorithms can be applied to solve the inversion problem. As a matter of fact, any method can be used which solves the normal equations (II.4.8) or (II.4.9), like e.g. Gauss-Newton methods. Particularly suited are Scaled Conjugate Gradient (SCG) methods which are proven to converge within  $\min(m, n_{tot})$  iterations, which automatically estimate  $(\mathbf{A}^t \mathbf{A} + \beta \mathbf{I})^{-1}$  without an explicit calculation and which have a 'memory' of previous search directions, since the present gradient is always conjugate to all previously computed (Møller, 1993; Masters, 1995).

Furthermore, in the case of strongly non-linear error surfaces with e.g. several local minima, both genetic algorithms and simulated annealing (Goldberg, 1989; Hertz et al., 1991; Masters, 1995) offer interesting alternatives and hybrid techniques can be considered (Masters, 1995). For instance, simulated annealing can be used to obtain several good initial weight distributions, which can then be optimized by an SCG-method. A review of learning algorithms including second order methods can be found in Battiti (1992). Reed (1993) gives an overview of regularization methods.

A last remark concerning the initialization of the weights. Equation (II.4.5) clearly shows the need to start with a good initial guess of these weights, as otherwise training may become very slow and the risk of falling in local minima increases significantly. Nevertheless, the most commonly used procedure is to apply a random initialization, i.e.  $w_i \in [-r, r]$ . Even some optimum bounds for  $r$  have been established (see for example Nguyen and Widrow (1990)).

As was already mentioned above, an alternative procedure is to use a global training scheme first to obtain several good initial guesses to start a localized optimization with. However, several theoretical methods have also been developed. The interested reader is referred to the articles of Nguyen and Widrow (1990) who use a linearization by parts of the produced output of the hidden neurons, Dencœur and Lengellé (1993) who use prototypes (selected training examples) for an adequate initialization and Sethi (1990, 1995) who uses decision trees to implement a 4-layer NN. Another interesting method is given by Karouia et al. (1994) using the theoretical results of Gallinari et al. (1991) who showed that a formal equivalence exists between linear NN and discriminant or factor analyses. Hence, they initialize their NN so that such an analysis is performed and start training from there on.

All of these initialization methods make use of the fact that although linear methods may not be capable of solving all considered applications, they do constitute a good starting point for NN. Hence, a 'linear' initialization is better than a random initialization of weights.

#### II.4.4.3 Generalization

Now that we are able to train a network, a new question arises. When should training be stopped? At first sight, it would seem to be a good idea to stop training when a local minimum is attained or when the convergence rate has become very small, i.e. improvement of iteration to iteration is zero or minimal. However, Geman et al. (1992) have shown that this leads to overtraining, i.e. memorization of the training set, as it is now the noise which is fitted and not any longer the global trend. Hence, the obtained weight distribution will be optimal for



the training samples, but it will result in a bad performance in general. A similar phenomenon occurs in tomography problems, where it is known as overfit (Scales and Snieder, 1998).

Overtraining is caused by the fact that system (II.4.7) is ill posed, i.e. a nullspace exists. The least-mean-squares solution of system (II.4.7), equation (II.4.8), will result in an optimal performance only if a perfect and complete training set is used without any noise contamination. Otherwise, any solution is non-unique due to the existence of this nullspace. Regularization with equation (II.4.9) reduces the influence of the nullspace, but does also result in a biased solution as was already mentioned above.

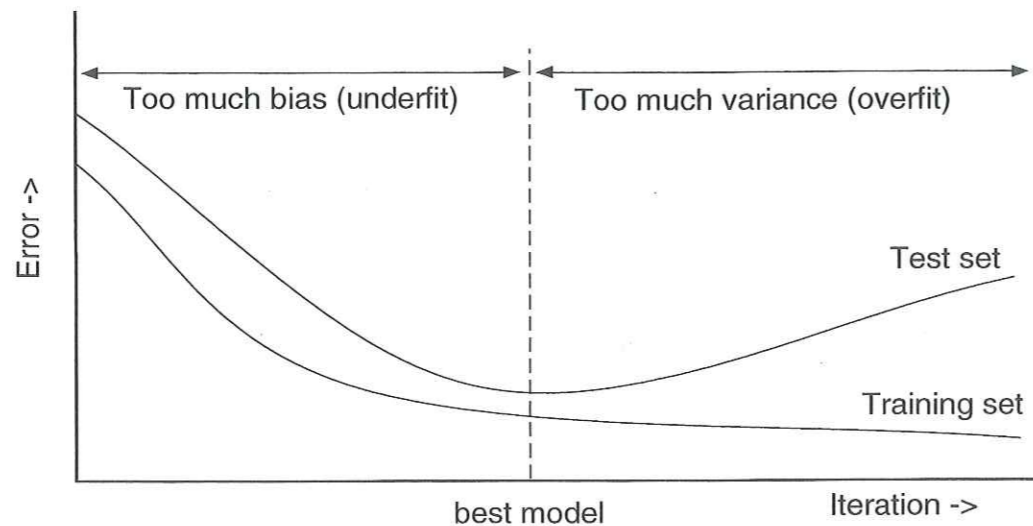


Figure II.4.4: Generalization versus training error. Adapted from Moody (1994).

The classical solution to this dilemma is the use of a split set of examples. One part is used for training, whereas the other part is used as a reference set to quantify the general performance (Figure II.4.4). Training is stopped when the misfit of the reference set has reached a minimum. This method is known as hold-out cross-validation.

Although this method is generally known to produce good results, it does result in a reduced training set which may pose a problem if only a limited amount of examples is available. Namely, as this method requires subdivision of the number of existing examples, the final number of used training samples will be reduced even further. Hence, the information contained in the selected examples is not optimally used and an increased risk of underconstrained training occurs.

It is possible to artificially increase the number of training samples  $m$  by using noise injection or synthetic modeling to generate noise free data. However, caution should be used when applying such artificial methods. In the former, small random perturbations are superposed on the existing training data. Mathematically, it can be shown that this corresponds to weight regularization (Matsuoka, 1992; Bishop, 1995; Grandvalet and Canu, 1995), thereby only reducing the number of effective weights. Moreover, the noise parameters must be chosen judiciously to optimize again the bias/variance trade-off. In addition, a bad noise model could introduce systematic errors. In the latter case, the underlying model may inadequately represent the real situation, thus discarding or misinterpreting important mechanisms.

To circumvent the problem of split data sets, some other techniques do exist, namely generalized cross-validation methods, residual analysis and theoretical measures which examine both the obtained output and the network complexity.

The problem of hold-out cross-validation is that information contained in some examples is

left out of the training process. Hence, this information is partly lost, since it is only used to measure the general performance, but not to extract the fundamentals of the considered process. As an alternative  $v$ -fold cross-validation can be considered (Moody, 1994; Chentouf, 1997).

In this method, the examples are divided in  $v$  sets of (roughly) equal size. Training is then done  $v$  times on  $v - 1$  sets in which each time another set is excluded. The individual misfit is defined as the misfit of the excluded set, whereas the total misfit is defined as the average of the  $v$  individual misfits. Training is stopped when the minimum of the total misfit is reached or convergence has become very slow. In the limit of  $v = m$ , the method is called 'leave-one-out'. In that case, training is done on  $m - 1$  examples and each individual misfit is calculated on the excluded example.

The advantage of  $v$ -fold cross-validation is that no examples are ultimately excluded in the learning process. Therefore, all available information contained in the training samples is used. Moreover, training is performed on a large part of the data, namely on  $(m - m/v)$  examples. Hence, the optimization problem is more easily kept overconstrained. On the other hand, training is considerably slower due to the repeated cross-validations. For further details we refer to Stone (1974) and Wahba and Wold (1975). Other statistical methods can also be considered like the Jackknife or the Bootstrap (Efron, 1979; Efron and Tibshirani, 1993; Masters, 1995), two statistical techniques which try to obtain the true underlying statistical distribution from the finite amount of available data without posing *a priori* assumptions on this distribution. Moody (1994) also described a method called non-linear cross-validation.

Another possible way to avoid a split training set is to minimize theoretical criteria relating the network complexity and misfit to the general performance (Chentouf, 1997). Such criteria are based on certain theoretical considerations which have to be satisfied. Some well known measures are the AIC and BIC criteria of Akaike (1970). Others can be found in Judge et al. (1980). For instance, the BIC criterion is given by

$$BIC = \ln \left( \frac{\sigma_r^2}{m} \right) + n_{tot} \frac{\ln m}{m}, \quad (\text{II.4.10})$$

where  $\sigma_r^2$  denotes the variance of the error residuals (misfits). The first term is clearly related to the misfit, whereas the second is related to the network complexity.

These criteria, however, have been developed for linear systems and are not particularly suited to NN due to their non-linear activation functions. Hence, several theoretical criteria had to be developed for such non-linear systems (MacKay, 1992; Moody, 1992; Murata et al., 1994). Like their predecessors, they are composed of a term related to the misfit and a term describing the complexity of the net. Hence, these criteria will also try to minimize both the misfit and the complexity of a network simultaneously.

However, it should be realized that these criteria are extremely powerful if the underlying theoretical assumptions are satisfied and in the limit of an infinite training set, i.e.  $m \gg n_{tot}$ . Otherwise they may yield erroneous predictions which can decrease the general performance of the obtained network. Moreover, these criteria can only be used if the NN is trained and its structure is adapted simultaneously.

A third method has recently been proposed in the NN literature by Jutten and Chentouf (1995) inspired by statistical optimization methods. It consists of a statistical analysis of the error residuals, i.e. an analysis of the misfit for all output values of all training samples is performed. It states that an optimally trained network has been obtained if the residuals and the noise have the same characteristics. For example, if noise is assumed to be white, training is stopped if the residuals have zero mean and exhibit no correlations (as measured by a statistical test). The method can be extended to compensate for non-white noise (Hosseini



and Jutten, 1999). The main drawback of this method is that *a priori* assumptions have to be made concerning the characteristics of the noise.

#### II.4.4.4 Configuration optimization: Preprocessing and weight regularization

The last remaining problem concerns the construction of a network configuration yielding optimal results. Insight in the way NN proceed to tackle classification problems will already allow for a notion of the required number of hidden layers and the type of NN. Nevertheless, in most cases, only vague ideas of the needed number of neurons per hidden layer will exist.

Classically, this problem is solved by trial-and-error, i.e. simply several structures are trained and their performances are examined. Finally, the best configuration is retained. The main problem of this approach is its need of extensive manual labor which may be very costly, although automatic scripts can be written for construction, training and performance testing.

In addition, the specific application and its complexity are not the only factors of influence. As was shown above, also the ratio of the number of total internal variables to the number of training samples is of direct importance to prevent an underconstrained optimization problem. This problem is of immediate concern for applications disposing of large input vectors, i.e.  $n_i$  is large, although regularization may help to limit the number of effective weights (Hush and Horne, 1993). Very often the number of required links and nodes can be easily reduced using preprocessing techniques to highlight the important information contained in the input or by use of local connections and weight sharing.

Many different preprocessing techniques are available. However, one of the best known is Principal Component Analysis or the Karhunen-Loève Transform. In this approach, training samples are placed as column vectors in a matrix  $\mathbf{X}$ . The covariance matrix  $\mathbf{X}\mathbf{X}^t$  is then decomposed in its eigenvalues and eigenvectors. Finally, training samples and, later, data are projected upon the eigenvectors of the  $p$  largest eigenvalues ( $p < m$ ). These eigenvectors span a new set of axes displaying a decreasing order of linear correlation between the training samples. In this way, any abundance in the input may be reduced. Moreover, only similarities are extracted which may reduce noise contamination. The ratio of the sum of the  $p$  largest eigenvalues (squared) over the total sum of squared eigenvalues yields an accurate estimate of the information contained in the projected data. More background is provided in Richards (1993) and Van der Baan and Paul (1999)<sup>6</sup>.

The matrix  $\mathbf{X}$  may contain either all training samples, the samples of only a single class or individual matrices may be formed for each existing class. In the latter case, each class will have its own network and particular preprocessing of the data. The individual networks are then often called expert systems which are only able to detect a single class, requiring therefore repeated data processing to extract all classes.

The use of the Karhunen-Loève Transform may pose problems if many different classes exist, as it will become more difficult to distinguish between classes using their common features. As an alternative, a factor or canonical analysis may be considered. This method separates the covariance matrix of all data samples into two covariance matrices of training samples within classes and between different classes. Next, a projection is searched which simultaneously yields minimal distances within classes and maximal distances between classes. Hence, only a single projection is required. A more detailed description can be found in Richards (1993).

The reason why principal component and factor analyses may increase the performance of NN is easy to explain. Gallinari et al. (1991) showed that a formal equivalence exists between linear NN (i.e. with linear activation functions) and discriminant or factor analyses. Strong

<sup>6</sup>Section II.3.

indications exist that non-linear NN (such as MLP and RBF networks) are also closely related to discriminant analyses. Hence, the use of a principal component or a factor analysis allows for a simplified network structure, since a part of the discrimination and data handling has already been performed. Therefore, local minima are less likely to occur.

Other interesting preprocessing techniques to reduce input can be found in Almeida (1994). All of these are cast in the form of NN structures. Notice, however, that nearly always the individual components of the input are scaled to lie within well defined ranges (e.g. between -1 and 1) to put the dynamic range of the input values within the most sensitive part of the activation functions. This often results in a more optimal use of the input. Hence, it may reduce the number of hidden neurons. For instance, le Cun et al. (1991) showed that correcting each individual input value for the mean and standard deviation of this component in the training set will increase the learning speed. Furthermore, for data displaying a large dynamic range, often the use of  $\log(x)$  instead of  $x$  is recommended.

Another possible way to limit the number of internal variables is to make *a priori* assumptions about the NN structure and, in particular, about the links between the input and the first hidden layer. For instance, instead of using a fully connected input and hidden layer, only local connections may be allowed for, i.e. it is assumed that only neighboring input components are related. Hence, links between these input nodes and a few hidden neurons will be sufficient. The disadvantage is that this method may force the number of hidden neurons to increase for an adequate description of the problem.

However, if the use of local connections is combined with weight sharing then a considerable decrease of  $n_{tot}$  may be achieved. Thus, grouped input links to a hidden node will have identical weights. Even grouped input links to several nodes may be forced to have identical weights. For large networks, this method may considerably decrease the total number of free internal variables. See e.g. le Cun et al. (1989). Unfortunately, results depend heavily on the exact NN structure and no indications exist for the optimal architecture.

The soft weight sharing technique of Nowlan and Hinton (1992) constitutes an interesting alternative. In this method it is assumed that weights may be clustered in different groups exhibiting Gaussian distributions. During training, simultaneously network performance, centers and variances of the Gaussian weight distributions and their relative occurrences are optimized. As one of the Gaussians is often centered around zero, the method combines weight sharing with Tikhonov regularization. One of the disadvantages of the method is its strong *a priori* assumption concerning weight distributions. Moreover, no method exists for determining the optimal number of Gaussians, again yielding an architecture problem.

#### II.4.4.5 Configuration optimization: Simplification methods

This incessant architecture problem can nowadays be solved in two different ways using either constructive or destructive, i.e. simplification, methods. The first method starts with a small network and simultaneously adds and trains neurons, whereas the second method starts with a large, trained network and progressively removes redundant nodes and links. First some simplification methods will be described. These simplification methods can be divided in two categories, those which remove only links and those which remove whole nodes. All simplification methods are also referred to as pruning techniques.

The simplest weight pruning technique is sometimes referred to as Magnitude Pruning. It simply consists of removing the smallest present weights and thereafter retraining of the network. However, this method is not known to produce excellent results (Le Cun et al., 1990; Hassibi and Stork, 1993), since such weights, though small, may have a considerable influence on the



NN performance.

A better method is to quantify the sensitivity of the misfit function to removal of individual weights. The two best known algorithms proceeding in such a way are Optimal Brain Damage or OBD (Le Cun et al., 1990) and Optimal Brain Surgeon or OBS (Hassibi and Stork, 1993).

Both techniques approximate the variation  $\delta E$  of the least-mean-squares misfit  $E$  due to removal of a weight  $w_i$  by a second order Taylor expansion, i.e.

$$\delta E = \sum_i \frac{\partial E}{\partial w_i} \Delta w_i + \frac{1}{2} \sum_i \frac{\partial^2 E}{\partial w_i^2} (\Delta w_i)^2 + \frac{1}{2} \sum_{i \neq j} \frac{\partial^2 E}{\partial w_i \partial w_j} \Delta w_i \Delta w_j. \quad (\text{II.4.11})$$

Higher order terms are assumed to be negligible. Removal of weight  $w_i$  implies  $\Delta w_i = -w_i$ . As all pruning techniques are only applied after NN are trained and a local minimum has been attained, the first term on the right-hand side can be neglected. Moreover, the OBD algorithm assumes that the off-diagonal terms ( $i \neq j$ ) of the Hessian  $\partial^2 E / \partial w_i \partial w_j$  are zero. Hence, the sensitivity (also called saliency)  $s_i$  of the misfit function to removal of weight  $w_i$  is expressed by

$$s_i = \frac{1}{2} \frac{\partial^2 E}{\partial w_i^2} w_i^2. \quad (\text{II.4.12})$$

Weights with the smallest sensitivities are removed and the NN is retrained. Retraining must be done after the suppression of a single or several weights. The exact expression for the diagonal elements of the Hessian is given by le Cun et al. (1990).

The OBS technique is an extension of OBD in which the need of retraining does no longer exist. Instead of neglecting the off-diagonal elements, this technique uses the full Hessian matrix  $\mathbf{H}$  which is composed of both the second and third term in the right-hand side in equation (II.4.11). Again, suppression of weight  $w_i$  yields  $\Delta w_i = -w_i$  which is now formulated as  $\mathbf{e}_i^t \Delta \mathbf{w} + w_i = 0$ , where the vector  $\mathbf{e}_i$  represents the  $i^{\text{th}}$  column of the identity matrix. This leads to a variation  $\delta E_i$

$$\delta E_i = \frac{1}{2} \Delta \mathbf{w}^t \mathbf{H} \Delta \mathbf{w} + \lambda (\mathbf{e}_i^t \Delta \mathbf{w} + w_i) \quad (\text{II.4.13})$$

(with  $\lambda$  a Lagrange multiplier). Minimizing the expression (II.4.13) yields

$$\delta E_i = \frac{1}{2} \frac{w_i^2}{H_{ii}^{-1}} \quad (\text{II.4.14})$$

and

$$\Delta \mathbf{w} = -\frac{w_i}{H_{ii}^{-1}} \mathbf{H}^{-1} \mathbf{e}_i. \quad (\text{II.4.15})$$

The weight  $w_i$  resulting in the smallest variation in misfit  $\delta E_i$  in equation (II.4.14) is eliminated. Thereafter, equation (II.4.15) will tell how all the other weights have to be adapted to circumvent the need of retraining of the network. Yet, after the suppression of several weights, the NN is usually retrained to increase performance.

Although the method is well based on mathematical principles, it does have the disadvantage that not only the full Hessian but also its inverse have to be calculated. In particular for large networks, this may require intensive calculations and may even pose memory problems. However, the use of OBS becomes very interesting if the inverse of the Hessian or  $(\mathbf{A}^t \mathbf{A} + \beta \mathbf{I})^{-1}$  has already been approximated due to the application of a second order optimization algorithm for network training. The exact expression for the full Hessian matrix can be found in Hassibi and Stork (1993).

Finally, it should be noted that equation (II.4.11) is only valid for small perturbations  $\Delta w_i$ . Hence, OBD and OBS should not be used to remove very large weights. Moreover, Cottrell et al. (1995) showed that both OBD and OBS amount to removal of weights which are statistically null. Furthermore, the statistical approach of Cottrell et al. (1995) can be used to obtain a clear threshold to stop pruning with the OBD and OBS techniques as they proposed not to remove weights beyond a Student's  $t$  threshold which has a clear statistical significance (Hérault and Jutten, 1994).

Instead of pruning only links, whole neurons can be suppressed. Two techniques which proceed in such a way are those of Mozer and Smolensky (1989) and Sietsma and Dow (1991). The skeletonization technique of Mozer and Smolensky (1989) prunes networks in a similar way as OBD and OBS. However, now the removal of whole nodes on the misfit is quantified. Again those nodes showing small variations are deleted.

Sietsma and Dow (1991) propose a very simple procedure to prune nodes, yielding excellent results. They analyze the output of nodes in the same layer to detect Non-Contributing Nodes (NCU). Nodes which produce a near constant output for all training samples or which have a correlated output to other nodes are simply removed, since such nodes are not relevant for the performance of the network. A correlated output implies that these nodes have always identical or opposite output. Removal of nodes can be corrected for by a simple adjustment of biases and weights of all nodes they are connected to. Hence, in principle, no retraining is needed, although it is often applied to increase performance. Although Sietsma and Dow (1991) do not formulate their method in statistical terms, a statistical framework can easily be forged and removal can be done on inspection of averages and the covariance matrix. Moreover, this would allow for a Principal Component Analysis within each layer to suppress irrelevant nodes.

Note that both the skeletonization method and NCU do also allow for the pruning of input nodes. Hence, they can significantly reduce the number of internal variables describing the NN, which is of particular interest in the case of limited quantities of training samples.

All pruning techniques will increase the generalization capacity of the network, due to a decreased number of local minima. Other pruning techniques can be found in Karnin (1990), Pellilo and Fanelli (1993) and Cottrell et al. (1995). A short review of some pruning techniques including weight regularization methods is given in Reed (1993).

During pruning, a similar problem occurs as during training. When should pruning be stopped? In practice, very often pruning is simply stopped when the next NN cannot attain a predefined maximum misfit. However, this may not be an optimum choice. A better method is the use of any of the techniques described in the subsection on generalization. The non-linear theoretical criteria may be especially interesting as they include the trade-off of net complexity versus misfit.

#### II.4.4.6 Configuration optimization: Constructive methods

As a last possibility for creating optimal network structures, we will consider the so-called constructive methods. Such methods start from scratch, i.e. only input and output layers are defined, and will automatically increase the network size until convergence has been reached. The principal problem associated with this approach is to find a suitable stopping criterion. Otherwise, the training set will simply be memorized and generalization will be poor. Mostly theoretical measures evaluating the trade-off between net complexity and performance are used (see subsection on generalization).

Nowadays, constructive algorithms exist for both MLP and RBF networks and even combinations of these, i.e. NN using mixed activation functions. Probably the best known constructive



algorithm is the Cascade-Correlation method of Fahlman and Lebiere (1990). It starts with a fully connected and trained input and output layer. Next, a hidden node is added which is initially only connected to the input layer. To obtain a maximum decrease of the misfit, the output of the hidden node and the prediction error of the trained network is maximally correlated. Next, the node is linked to the output layer, weights from the input layer to the hidden node are frozen (i.e. no longer updated) and all links to the output layer are optimized. In the next iteration, a new hidden node is added, which is linked to the input layer and the output of all previously added nodes. Again, the absolute covariance of its output and the prediction error of the NN is maximized, after which its incoming links are again kept frozen and all links to the output nodes are retrained, etc. This procedure continues until convergence. It will be evident that each new node forms a new hidden layer. Hence, the algorithm constructs very deep networks in which each node is linked to all others. Moreover, the original algorithm does not use any stopping criterion, as input is assumed to be noiseless.

Two more recently proposed techniques which do not have these drawbacks are the incremental algorithms of Moody (1994) and Jutten and Chentouf (1995). These algorithms differ from Cascade-Correlation in such that only a single hidden layer is used and that all links are updated. The two methods differ in the number of neurons added per iteration: one (Jutten and Chentouf, 1995) or several (Moody, 1994), and their stopping criteria. Whereas Moody (1994) used the Generalized Prediction Error criterion of Moody (1992), Jutten and Chentouf (1995) analyze the misfit residuals. Further construction is ended if the characteristics of the measured misfit resemble the assumed noise characteristics.

A variant (Chentouf and Jutten, 1996b) of their algorithm also allows for the automatic creation of NN with several hidden layers. Its general way of proceeding is identical to the original algorithm. However, now it evaluates if a new neuron has to be placed in an existing hidden layer or if a new layer has to be created. Another variant (Chentouf and Jutten, 1996a) allows for the incorporation of both sigmoidal and Gaussian neurons. It simply evaluates which type of activation function yields the largest reduction in the misfit. See also Chentouf (1997).

The Dynamic Decay Adjustment (DDA) method of Berthold and Diamond (1995) is an incremental method for RBF networks which automatically estimates the number of neurons and the centers and variances of the Gaussian activation functions best providing an accurate classification of the training samples. It uses selected training samples as prototypes. These training samples define the centers of the Gaussian activation functions in the hidden layer neurons. The weight of each Gaussian represents its relative occurrence and the variance the region of influence. To determine these weights and variances, the method uses both a 'negative' and a 'positive' threshold. The negative threshold forms an upper limit for the output of wrong classes, whereas the positive threshold indicates a minimum value of confidence for correct classes. That is, after training, training samples will at least produce an output exceeding the positive threshold for the correct class and no output of the wrong classes will be larger than the negative threshold.

During training, the algorithm presents the training samples consecutively to the network. If a training sample cannot be correctly classified by the existing network then this training sample is used as a new prototype. Otherwise, the weight of the nearest prototype is increased to increase its relative occurrence. The variances of all Gaussians describing conflicting classes are reduced such that no conflicting class produces values larger than the negative threshold for this training sample. Output is not bounded, due to the linear activation functions which exist in the output nodes. Hence, this is a decision taking network, i.e. it only gives the most likely class for a given training sample, but not its exact likelihood.

The DDA algorithm of Berthold and Diamond (1995) has some resemblance with the Prob-

abilistic Neural Network (PNN) of Specht (1990). This network creates a Gaussian centered at each training sample. During training, only the optimum, common variance for all Gaussians has to be estimated. However, the fact that a hidden node is created for each training sample makes the network more or less a referential memory scheme and will render the use of large training sets very cumbersome. DDA, on the other hand, only creates new nodes when necessary.

Other incremental algorithms include Orthogonal Least Squares (OLS) of Chen et al. (1991), Resource Allocating Network (RAN) of Platt (1991) and Projection Pursuit Learning (PPL) of Hwang et al. (1994). A recent review of constructive algorithms can be found in Kwok and Yeung (1997).

## II.4.5 Practice

### II.4.5.1 A general strategy

How can the above described methods and techniques be used in a geophysical application? The following list contains some relevant points to be considered for any application. Particular attention should be paid to:

1) Choice of type of NN to be used, i.e. for static problems MLP or RBF. A preliminary data analysis or general considerations may already give an indication of the optimum choice. For instance, clusters in classification problems are often thought to be localized in input space. Hence, RBF nets may yield better results than MLP nets. However, notice that both types of NN are universal approximators, yielding that they are capable of producing identical results. Nevertheless, one type may be more suited for a particular application than the other type as these predictions are asymptotic results. If no indications exist, simply both have to be tried.

2) Choice of input parameters. In some problems, this may be a trivial question, though this may not generally be the case. In the extremest case, any parameter which can be thought off may be included, after which a Principal Component Analysis (PCA) or factor analysis may be used to reduce input space and thereby any redundancy and non-relevant parameters. Nevertheless, it should be realized that an adequate selection of parameters will significantly increase performance and quality of final results.

3) Suitable preprocessing techniques, i.e. any rescaling, filtering or other means allowing for an more effective use of the input parameters. Naturally, PCA or a factor analysis can be included here.

4) The training set and more particularly the number of available training samples. This number is of direct influence on the total number of internal variables allowed in the NN to keep training overconstrained. The total number of internal variables should never exceed the number of training samples. The largest fully connected NN can be calculated using equations (II.4.3) and (II.4.4). Naturally, such limitations will not exist if a very large training set is available.

5) Training algorithm and generalization measure. Naturally, a training algorithm has to be chosen. Conjugate gradient methods will yield superior performance than the standard backpropagation algorithm, since e.g. the first is proven to converge within a limited number of iterations, whereas the latter is not. Furthermore, a method has to be chosen to guarantee also a good performance in general. This can be any of the methods described in the subsection on generalization, i.e. cross-validation, a theoretical measure or residual analysis. However, these measures have to be calculated during training and not after convergence.

6) Configuration estimation, i.e. the choice between the use of a constructive or simplification method. An increasingly popular choice is using any constructive algorithm to obtain a suitable network configuration and thereafter the application of a pruning technique to realize a minimal optimum configuration. Sometimes, reinitialization and retraining of an NN may result in a



better than previous misfit and may allow for a continued pruning of the network. In such cases, the reinitialization has allowed for an escape of a local minimum.

#### II.4.5.2 An example

To illustrate how some of the above-described methods and techniques can be put in a general methodology, a single example will be considered which can relatively easily be solved using NN without the need for complicated processing schemes. Our example concerns the detection and extraction of reflections, ground roll and other types of noise in a deep seismic reflection experiment to enhance data quality. Use was made of the amplitude spectra, as it was assumed that a first distinction between signal types could be made on their frequency content. In addition, samples are now insensitive to phase perturbations.

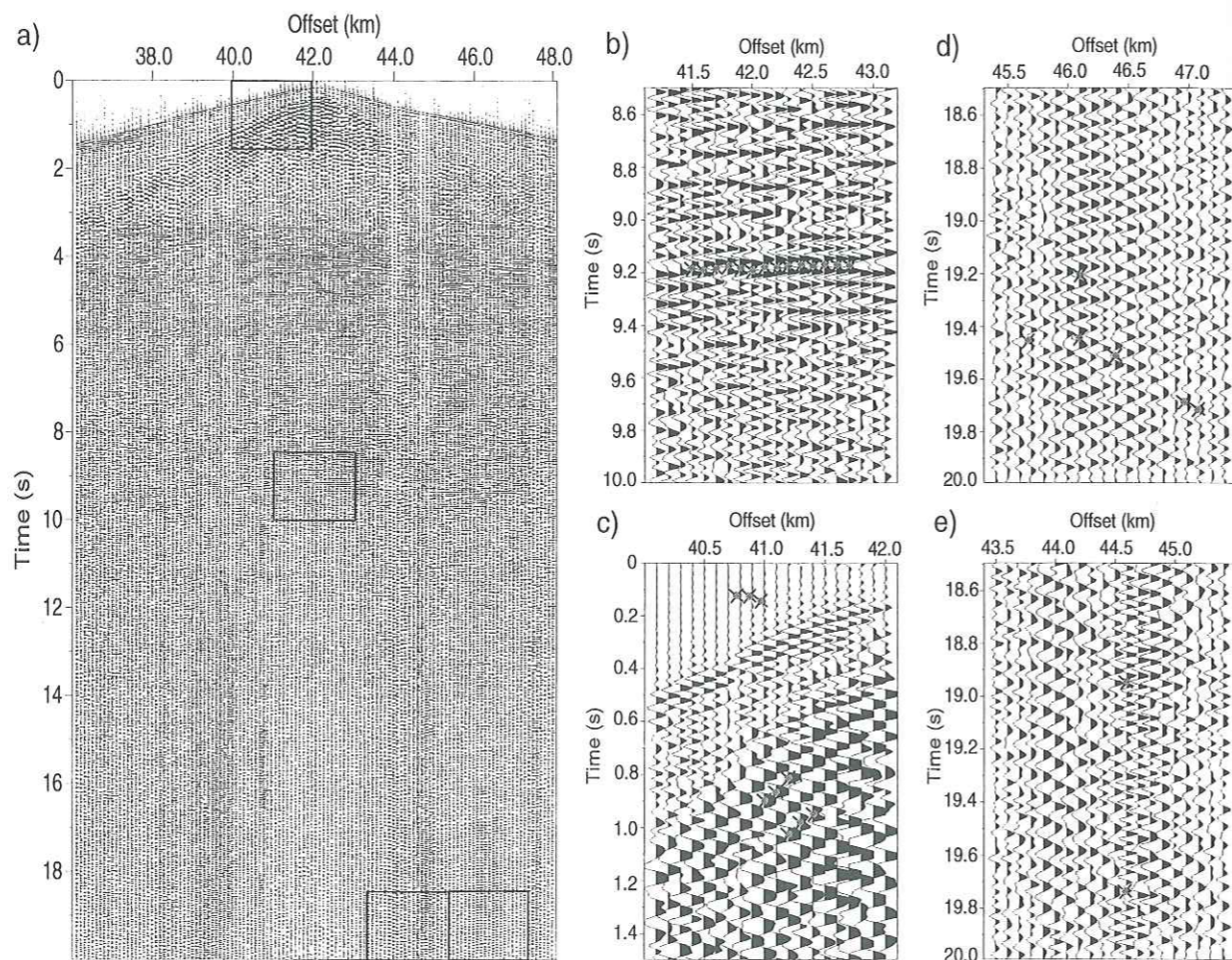


Figure II.4.5: Common shot gather + 31 pick positions. a) Original data, b) 14 reflection picks, c) 3 pre-arrival noise + 6 ground roll picks, d) 6 picks on background noise + bad traces and e) 2 more picks on bad traces.

Van der Baan and Paul (1999)<sup>7</sup> have shown that the application of Gaussian statistics on local amplitude spectra after the application of a PCA allows for an efficient estimation of the presence of reflections and therefore their extraction. They used a very simple procedure to

<sup>7</sup>Section II.3.

extract the desired reflections. In a common shot gather (Figure II.4.5.a) a particular reflection was picked 14 times on adjacent traces (Figure II.4.5.b). Local amplitude spectra were calculated using 128 ms (16 points) windows centered around the picks. These local amplitude spectra were put as column vectors in a matrix  $X$  and a PCA was applied using only a single eigenvector. That is, the first eigenvector of  $XX^t$  was calculated and henceforth all amplitude spectra were projected upon this vector to obtain a single scalar indicating resemblance with the 14 training samples. Once all 14 amplitude spectra have been transformed into scalars, their average and variance are calculated. The presence of reflection energy is now estimated by means of 1) a sliding window to calculate the local amplitude spectra, 2) a projection of this spectrum upon the first eigenvector and 3) the use of Gaussian statistics described by the scalar mean and variance to determine the likelihood of the presence of a reflection for this particular time and offset. Extraction results (obtained by means of a multiplication of the likelihood distribution with the original data) are shown in Figure II.4.6.a. More details can be found in Van der Baan and Paul (1999).

To obtain a good idea of the possible power of NN, their method will be extended to detect and extract two other categories of signal, namely ground roll and all remaining types of noise (including background noise, bad traces and pre-arrival noise). In this aim, more picks have been done on such non-reflections. Hence, ground roll (Figure II.4.5.c), background and pre-arrival noise (Figures II.4.5.c and II.4.5.d) and bad traces (Figures II.4.5.d and II.4.5.e) have been selected. This resulted in a total training set containing 14 reflections (identical to those used in Van der Baan and Paul (1999)) and 17 non-reflections.

Next, the two other categories of signal are extracted in a similar way as the reflections, i.e. Gaussian statistics are applied to local amplitude spectra after a PCA. Figure II.4.6.b and II.4.6.c show the results. A comparison of these figures with Figure II.4.5.a shows that good results are obtained for the extracted ground roll. However, in Figure II.4.6.c many laterally coherent reflection events are visible. Hence, the proposed extraction method is not well able to discern the third category of signals, i.e. all types of noise except ground roll.

Fortunately, the failure to extract all remaining types of noise is easy to explain. Whereas both reflections and ground roll are characterized by a specific frequency spectrum, the remaining types of noise display a large variety in frequency spectra containing both signals with principally only high or low frequencies. Therefore, the remaining types of noise have a multi-modal distribution which cannot be handled by a simple Gaussian distribution. To enhance extraction results, the remaining types of noise should be divided into several categories such that no multi-modal distributions will exist.

In the following we will show how different NN are able to produce similar and better results using both MLP and RBF networks. However, it should be noted that we did not want to test the influence of different generalization measures. Hence, results were manually selected – a procedure we do not recommend for general use.

As input parameters, the 9 frequencies in the local amplitude spectra are used. These 9 frequencies resulted from the use of 16 points (=128 ms) in the sliding window. All simulations have been performed using the SNNS (1996) package, which is very suited for those who do not wish to program their own applications and algorithms.

Equation (II.4.3) and (II.4.4) indicate that for a training set containing 31 samples and having 9 input parameters and 3 output nodes, already 3 hidden nodes results in an underconstrained training problem. Two hidden layers are out of the question. Even if expert systems are used (networks capable of recognizing only a single type of signal) then 3 hidden nodes also results in an underconstrained training problem. On the other hand, expert systems for extraction of reflections and ground roll may benefit from a PCA data preprocessing as it significantly reduces



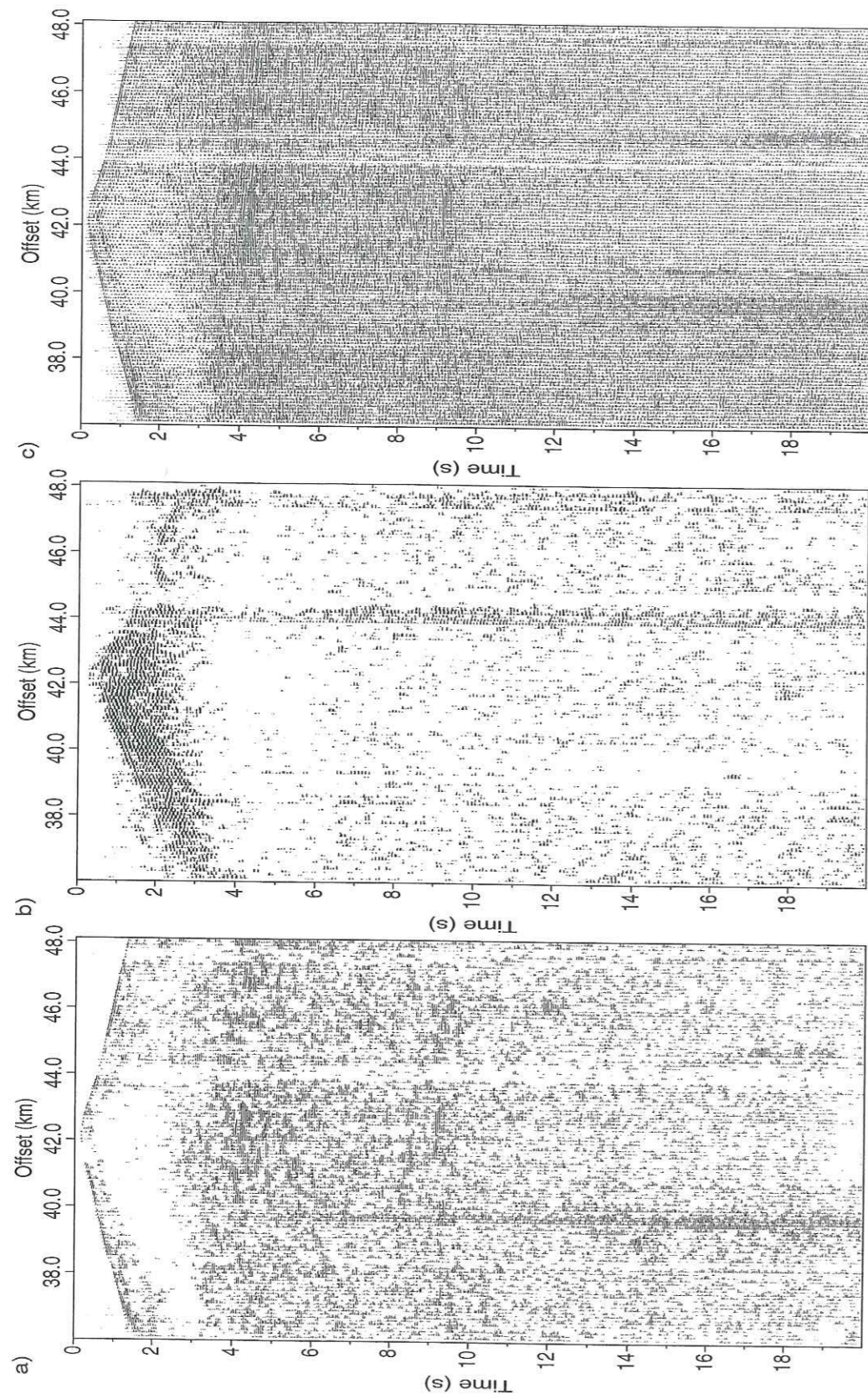


Figure II.4.6: Extraction results using Gaussian statistics and a PCA for a) reflections, b) ground roll and c) other types of noise. To be compared with Figure II.4.5.a.

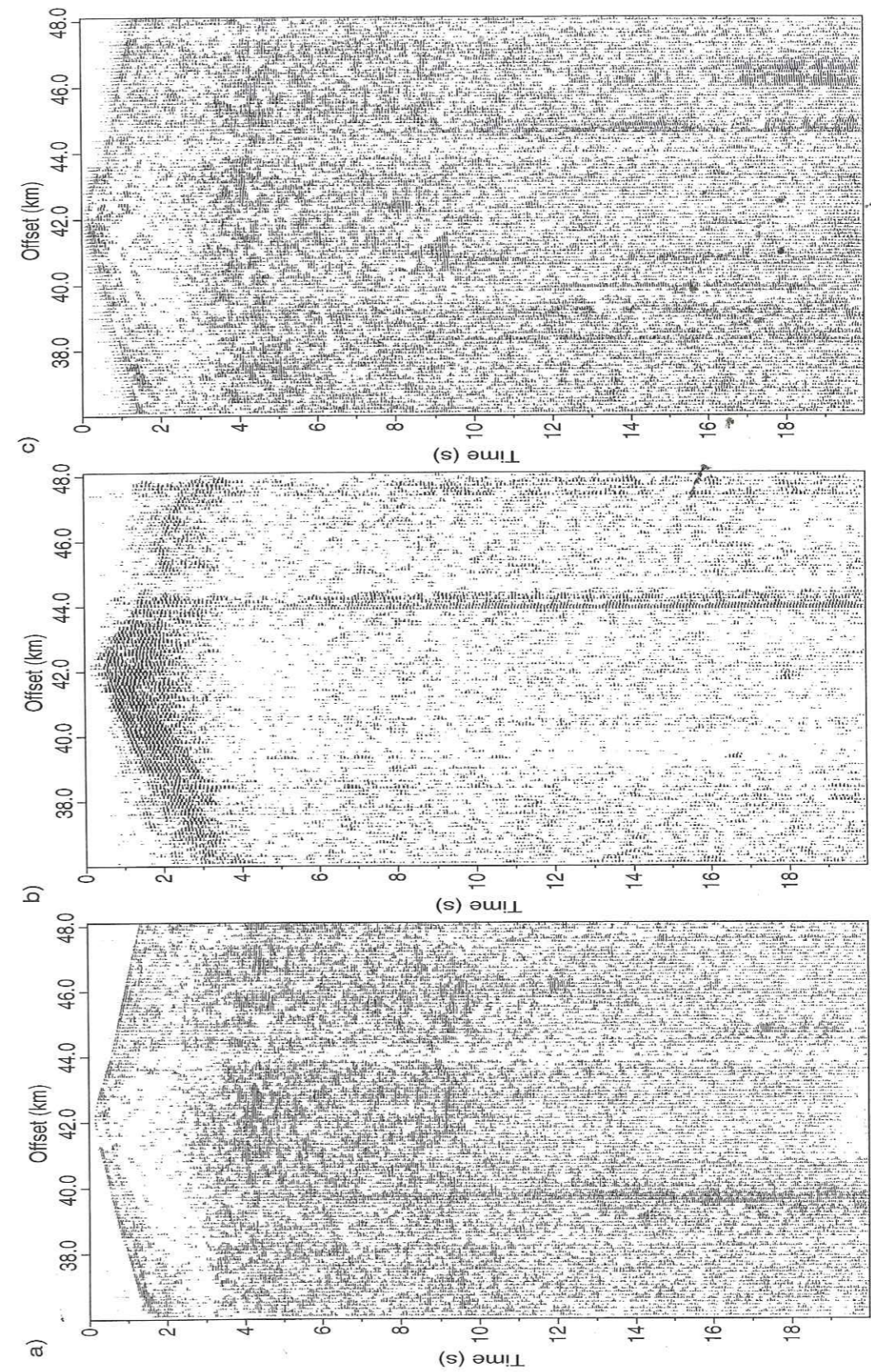


Figure II.4.7: Extraction results using an MLP network for again a) reflections, b) ground roll and c) other types of noise. Notice the improvement of extraction results for the remaining types of noise.



the number of input parameters and thereby allows for a larger number of hidden neurons.

The first network to be used was a so-called 9-5-3 MLP network, i.e. 9 input, 5 hidden and 3 output nodes. The network was trained until convergence. The fact that this may have resulted in an overfit is not of importance, since the network obtained was thereafter pruned using the Non-Contributing Units (NCU) method of Sietsma and Dow (1991). This particular method was chosen because it is capable of removing whole nodes at a time (including input nodes). Finally, this resulted in a 4-2-3 NN. The 4 remaining input nodes contained the second to the fifth frequency component of the amplitude spectra. The resulting signal extractions are displayed in Figure II.4.7. A comparison with the corresponding extraction results of the method of Van der Baan and Paul (1999) shows that more reflection energy has been extracted (Figure II.4.6.a versus II.4.7.a). Similar results are found for the ground roll (Figure II.4.6.b versus II.4.7.b). However, the extraction results for the last category containing all remaining types of noise has been greatly improved (compare Figure II.4.6.c and II.4.7.c). Nevertheless, some laterally coherent energy remains visible in Figure II.4.7.c, which may be attributable to non-detected reflective energy. Hence, results are amenable to some improvement, for instance by including lateral information.

The second network to be trained was a 9-5-3 RBF net. Again, the network was trained until convergence and thereafter pruned using NCU. The final network structure consisted of a 5-2-3 NN producing slightly worse results than Figure II.4.7.c for the remaining noise category as some ground roll was still visible after extraction. The 5 remaining input nodes were connected to the first 5 frequency components of the local amplitude spectra.

Hence, both MLP and RBF networks are able to solve this particular problem in a convenient and efficient way, whereas a more conventional approach encountered more problems. In this particular application, results did not benefit from a PCA data preprocessing as the distributions were too complicated (mixed and multi-modal). However, the use of a factor analysis might have been an option.

Although neither network was able to produce identical extraction results as those obtained by Van der Baan and Paul (1999) for the reflection energy, highly similar results could be obtained using different expert systems with either 4 input and a single output node or a single input and output node (after a PCA preprocessing of data). Thus, similar extraction results could be obtained using very simple expert systems without hidden layers.

#### II.4.6 Discussion and Conclusions

Neural networks are universal approximators. That is, they can obtain an arbitrary close approximation to any continuous function, be it associated to a direct or inverse problem. Therefore, they will constitute a powerful tool for the geophysical community.

Many different methods and techniques have been described in this paper to facilitate their use and to increase their performance. The last remaining principal issue is related to the training set. As in many other methods, the quality of the obtained results stands or falls with the quality of the training data.

Furthermore, one should first consider whether the use of neural networks for the intended application is worthwhile spending expensive research time. Generally, this question reduces to the practical issue of whether sufficient good training samples can be obtained to guarantee an overconstrained training procedure. This problem may hinder their successful application even after significant preprocessing of the data and reduction of the number of input parameters. If a negative answer has to be given to this pertinent question then a better alternative is the development of sound mathematical foundations for the particular application.

#### II.4.7 Acknowledgments

The first author wishes to thank Philippe Lesage for a first introduction to the domain of neural networks and for pointing to the existence of SNNS. In addition, discussions with Shahram Hosseini are acknowledged. Finally, we are grateful for the reviews of Bee Bednar, an anonymous reviewer and S.A. Levin to whom the note of caution about the use of "synthesized" data is due.



#### II.4.8 Les réseaux de neurones et la détection de signaux: Conclusions

La comparaison faite dans la section II.4.5.2 nous a montré que les RN offrent des possibilités intéressantes pour le problème de la détection de signaux. Ceci est dû au fait que ces outils mathématiques peuvent estimer des distributions statistiques sous-jacentes très compliquées de façon automatique pourvu bien entendu qu'il y a suffisamment d'exemples.

Cependant, cette capacité a des inconvénients, puisqu'elle fait que les RN sont plus difficiles à mettre en œuvre. Par conséquent, la méthode CVM est mieux appropriée pour des distributions sous-jacentes simples ou bien connues et les RN pour des distributions asymétriques ou multimodales.

## II.5 Transformée Continue en Ondelette

### Résumé

Une coupe sismique bruitée est difficile à interpréter. Cependant, nous avons vu dans les sections précédentes que les techniques de reconnaissance sont capables de détecter et d'extraire un signal spécifique. Par conséquent, ces techniques peuvent augmenter le rapport signal sur bruit et ainsi la qualité des données. Par exemple, les réseaux de neurones et la méthode de Classification par Vraisemblance Maximale peuvent accomplir cette tâche étant donné quelques points de référence.

Nous pouvons caractériser les réflexions à la fois en utilisant une Transformée de Fourier à Court Terme (TFCT) ou bien une Transformée Continue en Ondelette (TCO). La TCO fait une analyse d'échelle des données et ainsi extrait de façon automatique les échelles caractéristiques de l'énergie réfléchie (c'est-à-dire les périodes principales). D'autre part, la technique de la TFCT nécessite une spécification *a priori* de la période principale, puisque les résultats dépendent directement de la longueur de la fenêtre d'analyse.

S'il s'agit de l'extraction d'un seul type de signal, les 2 transformées produisent des résultats assez similaires. Par conséquent, nous recommandons d'utiliser une TFCT afin de caractériser les signaux, car cette technique est plus rapide et une analyse d'échelle n'est pas nécessaire. Par contre, si une extraction simultanée de plusieurs types de signaux est désirée, une analyse d'échelle est appropriée et ainsi une TCO donnera de meilleurs résultats.

### II.5.1 La Transformée Continue en Ondelette et la détection des signaux: Introduction

Les deux sections précédentes nous ont montrés qu'il est possible de détecter et d'extraire les réflexions dans les données d'une expérience de sismique réflexion profonde en utilisant une Transformée de Fourier à Court Terme (TFCT). Mais, est-ce que cette transformée est le mieux appropriée pour caractériser la signature des réflexions et d'autres signaux?

Une autre transformée qui est souvent utilisée pour analyser le contenu fréquentiel d'un signal non-stationnaire est la Transformée Continue en Ondelette (TCO). L'avantage de cette technique est qu'elle analyse les signaux sur des échelles différentes. Ceci a pour conséquence qu'il n'est plus nécessaire d'estimer *a priori* la longueur optimale de la fenêtre d'analyse. Malheureusement, ceci résulte aussi dans des temps de calcul plus importants.

Dans la suite, nous examinons dans quels cas il vaut mieux utiliser une TCO ou une TFCT.



## Recognition methods for S/N enhancement: A comparison<sup>8</sup>

Mirko van der Baan

*Laboratoire de Géophysique Interne et Tectonophysique,  
Université Joseph Fourier and CNRS, BP 53, 38041 Grenoble Cedex 9, France.*

### Abstract

Noisy seismic sections are hard to interpret. Recognition techniques, however, are able to detect and extract specific signals only and can, thereby, increase the signal-to-noise ratio and, thus, data quality. For instance, both neural networks and Maximum Likelihood Classification are able to perform this task given some reference picks.

Both the Short Term Fourier Transform (STFT) and the Wavelet Transform (WT) can be used to characterize reflections. The WT will perform a scale analysis of the data and can, thereby, extract automatically the characteristic scale lengths, i.e. principal periods, of reflection energy. The STFT-technique, on the other hand, requires the *a priori* specification of the main principal period, since results depend directly on the length of its tapering window.

Nevertheless, both transforms produce rather similar results if only a single type of signal has to be extracted. Hence, in that case, the use of an STFT is recommended, as it is faster than a WT and a scale analysis is not required. However, if several signals have to be extracted simultaneously then the use of a WT is appropriate as a scale analysis is warranted.

### II.5.2 Introduction

Seismic sections do nearly never attain the desired quality, which complicates their interpretation. Therefore, a method has to be devised which is capable of increasing the signal-to-noise ratio (S/N-ratio) and thereby data quality. A possible solution would be the extraction of desired signal only. Then, due to the extraction process, the presence of excessive noise will be reduced and, thereby, stack degeneration and migration artifacts. Nevertheless, in order to perform such an extraction, the desired signal has first to be detected, which requires a capable recognition method.

In a previous paper (Van der Baan and Paul, 1999; section II.3) a classical recognition technique in remote sensing, i.e. teledetection, has been presented, namely Maximum Likelihood Classification (MLC). The used MLC-technique enabled us to increase the S/N-ratio, since it calculates the likelihood of the presence of reflections in the data. Therefore, a multiplication of the obtained probability distribution of reflection energy with the original data will emphasize reflections, thereby resulting in an extracted, 'clean' section.

Basically, recognition techniques are composed of three stages, namely the selection of a training set, learning the typical characteristics and finally calculation of the desired probability distributions (Richards, 1993). See figure II.5.1.

Thus, in the first stage, a typical reflection is picked and, for each pick, the local amplitude spectrum is calculated. The amplitude spectra are used, since it is assumed that discrimination between different signal types can be done using the frequency content of signals. Moreover, use of the amplitude spectra allows for the detection of both reflections and diffractions, since these differ only by a phase shift (Troyer, 1970).

Next, in the learning stage, the variances and means are calculated for each frequency in the amplitude spectra of the picks. Finally, using a sliding window, local amplitude spectra are

8. Cette partie est extraite d'un article qui n'a jamais été soumis.



calculated and, by means of Bayesian statistics, the probability of the presence of the reflections can be estimated. The desired reflections can then be extracted by means of a simple multiplication of this likelihood distribution with the original data. Hence, a Short Term Fourier Transform (STFT) was used to obtain the desired likelihood distributions. For more details and the exact equations, the reader is referred to section II.3.

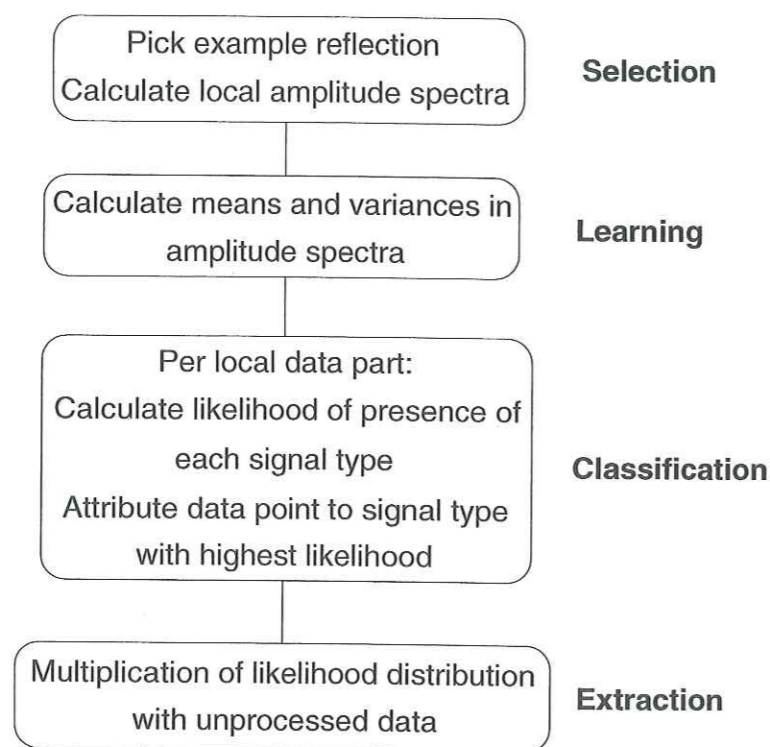


Figure II.5.1: Flow diagram of the MLC-technique using the STFT-technique for signal characterization. The first three steps are typical for any classification algorithm, whereas the fourth is required for signal extraction.

Although results were fine, two main inconveniences were encountered. First of all, the underlying statistical distribution of the spread in the characteristics of reflections had to be specified *a priori*. Second, an STFT uses a fixed window length, implying that the ideal length also had to be estimated *a priori*.

In this section, an alternative technique will be considered to characterize reflection energy, namely the Wavelet Transform (WT). This technique is able to circumvent the problem of fixed window lengths, as it performs a scale analysis of the data. Hence, it will determine automatically the characteristic scale length(s), i.e. principal periods. The problem of the *a priori* specification of the underlying statistical distribution has already been considered in the previous section (II.4) using neural networks.

The section begins with a more detailed comparison of WT versus STFT and continues with a comparison of all techniques concerning speed, ability to detect reflections and other particular problems.

### II.5.3 Maximum Likelihood Classification

MLC is the most commonly used supervised classification method for remote sensing image data analysis (Richards, 1993). It consists of three steps as displayed in Figure II.5.1. To classify data into different classes or categories, this technique makes use of the conditional probability density functions (pdf)  $p(\omega_i|\mathbf{x})$ , where  $\omega_i$  denote the  $I$  predefined spectral classes and the vector  $\mathbf{x}$  represents the attributes describing the signature of a local part of the data.

In this application, only two classes exist, namely reflections and non-reflections. Only the likelihood of the presence of reflection energy will be calculated, since the two are assumed to be complementary. The attribute vector  $\mathbf{x}$  will contain the amplitude spectra calculated from either the Fourier or wavelet coefficients, depending on which transform is used (STFT or WT).

Therefore, only knowledge of the conditional pdf  $p(\omega_{refl}|\mathbf{x})$  is required for the detection and extraction of reflection energy, since this function expresses the probability of the presence of a reflection given the attribute vector  $\mathbf{x}$  (i.e. some calculated Fourier or wavelet coefficients). The conditional pdf  $p(\omega_{refl}|\mathbf{x})$  can be estimated using Bayes' rule, which simplifies to the identity  $p(\omega_{refl}|\mathbf{x}) = p(\mathbf{x}|\omega_{refl})$  in the case of a single class (section II.3). The conditional pdf  $p(\mathbf{x}|\omega_{refl})$  expresses the likelihood that a reflection is present at the data point with attributes  $\mathbf{x}$  given the characteristics of this class. The learning phase of the MLC-technique mainly consists of estimating this distribution function in addition to the required characteristics.

In most applications, the conditional pdf  $p(\mathbf{x}|\omega_{refl})$  is simply assumed to be a normal or Gaussian distribution. Such a distribution was found to be too strict and a Principal Component was applied. Not only will use be made of a Principal Component Analysis, but also a slight adaptation of a Gaussian distribution will be employed. To damp the Gaussian distribution, it will be scaled, within the exponentials, with the size  $J$  of the attribute vector  $\mathbf{x}$  to correct for different window lengths:

$$p(\mathbf{x}|\omega_{refl}) = \prod_{j=1}^J \exp \left\{ \frac{-c(x_j - m_j)^2}{2J\sigma_j^2} \right\}. \quad (\text{II.5.1})$$

The constant amplitude factor preceding the product of exponentials has been neglected in this equation. The constant factor  $c$  can be used to adapt manually the effect of damping (reduced if increased), but is kept to 1.0 in this study. Thus, the characteristics of reflection energy are determined by the means  $m_j$  and the variances  $\sigma_j^2$  of the attributes of the reference picks, i.e. the local amplitude spectra for the STFT-technique. For a fair comparison, however, energy of each estimate of the attribute vector  $\mathbf{x}$  has to be normalized.

### II.5.4 Wavelets

#### II.5.4.1 STFT and WT in general

An STFT-analysis permits us to describe the evolution of the frequency content of a signal with time which is of special interest in case of non-stationary data. In fact, it is simply a repeated Fourier Transformation of the data using a sliding time window. Mathematically, this can be expressed as

$$\mathcal{F}[s](f, \tau) = \int s(t)g^*(t - \tau)e^{-2\pi ift} dt \quad (\text{II.5.2})$$

with  $s(t)$  the recorded time signal and  $g^*(t)$  the complex conjugate of the employed window  $g(t)$ . However, results of the analysis depend explicitly on the choice of window  $g(t)$ , concerning both its exact form and length, since these remain fixed (Rioul and Vetterli, 1991). Therefore,



an STFT-analysis will encounter problems if e.g. a short pulse is superposed on a long period signal, since the analyzing window will either be too short or too long (Lilly and Park, 1995).

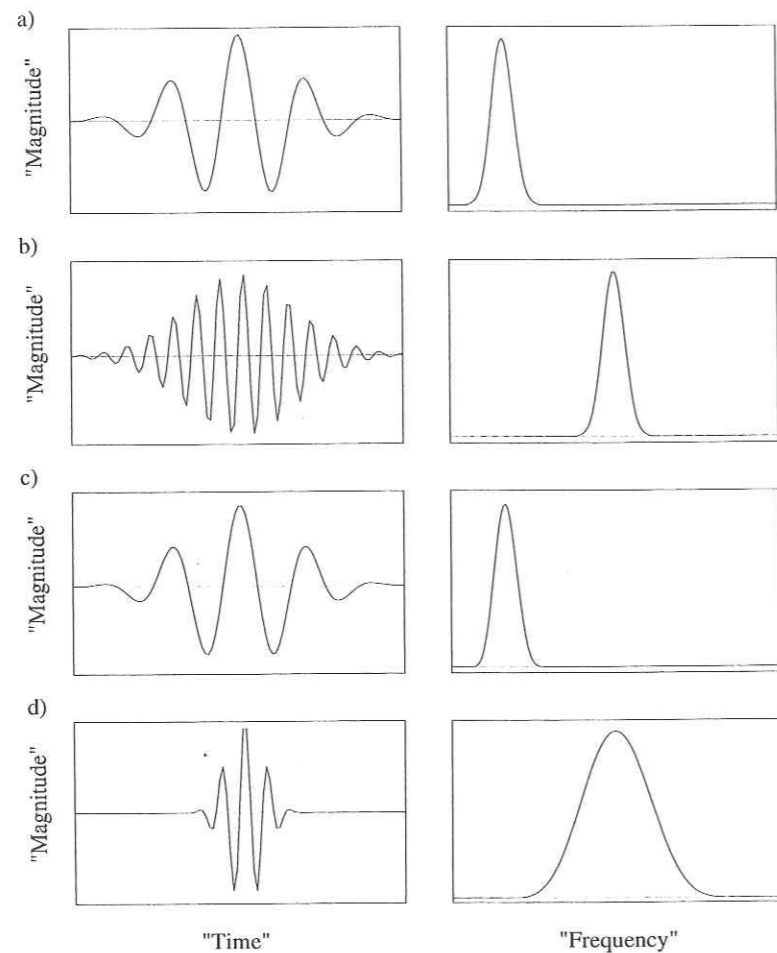


Figure II.5.2: Trade-off of time versus frequency resolution for both an STFT and a WT. The STFT-technique employs a constant window length. Thus, a) a long period and b) a short period signal will have an identical resolution (width of peak) in frequency domain, whereas time and frequency resolution are inversely proportional for c) large scale and d) small scale analyzing wavelets. Hence, small scale wavelets have an excellent time, but nearly no frequency resolution, whereas large scale wavelets show an opposite behavior. The STFT-technique uses a Gaussian tapering window for this example which makes a) nearly identical to c), the imaginary part of a particular Slepian mother wavelet. After Bear and Pavlis (1997).

The WT, on the other hand, forms a natural bridge between frequency and time analyses, since it decomposes the data using different window lengths or, more properly said, different scale lengths. Moreover, physically, the use of a WT is correcter for analyzing signals than an STFT due to the Uncertainty Principle of Heisenberg. This principle states that frequency and time resolutions are inversely proportional to each other. This trade-off between frequency and time resolution in an STFT is fixed by the choice of the window length, whereas for a WT these are not kept constant, but varied inversely proportional. In principle, all steps from infinite time resolution (i.e. the position of a Dirac) and no frequency resolution to no time but infinite frequency resolution (i.e. a mono-frequent wave) can be accounted for. See Figure II.5.2 for an illustration of this phenomenon of trade-off in resolutions.

Thus, to circumvent the problem of fixed window length, the WT should be considered. Mathematically, it can be written as

$$\mathcal{W}[s](a, \tau) = a^{-1/2} \int s(t) w^* \left( \frac{t - \tau}{a} \right) dt, \quad (\text{II.5.3})$$

where  $w(t)$  represents the shifted and scaled mother wavelet and  $a$  denotes scale (which is directly related to period). The normalization factor  $a^{-1/2}$  ensures that wavelets have identical energy for any scale. The mother wavelet represents that type of wavelet used for signal analysis.

The number of possible types of mother wavelets is very large. Required are only finite energy (often normalized), compact support in time (or sufficiently fast decay) to ensure localization in time and zero mean to create a wiggle, i.e. a wavelet (Kumar and Foufoula-Georgiou, 1997).

To obtain some more physical insight in the WT, we can consider the classical Morlet wavelet which consists of a real cosine and an imaginary sine multiplied with a Gaussian, i.e.

$$w(t) = \pi^{-1/4} e^{2\pi i f_0 t} e^{-t^2/2}. \quad (\text{II.5.4})$$

Substitution of the Morlet wavelet in equation (II.5.3) results in

$$\mathcal{W}[s](a, \tau) = a^{-1/2} \int s(t) \left\{ \pi^{-1/4} e^{-\frac{(t-\tau)^2}{2a^2}} e^{2\pi i f_0 \tau/a} \right\} e^{-2\pi i f_0 t/a} dt, \quad (\text{II.5.5})$$

which bears a strong resemblance with the STFT, equation (II.5.2), indicating already that, analogous to the STFT, results will depend on the precise choice of the mother wavelet.

Obviously, most mother wavelets will not span an orthonormal basis, yielding very often a non-optimal decomposition of the signal. Nevertheless, it is still possible to reconstruct the decomposed signal in the usual way, i.e. by means of an inner product of the obtained wavelet coefficients and the scaled and shifted wavelets, i.e.

$$s(t) = c_w \int \int_{a>0} \mathcal{W}[s](a, \tau) \left[ a^{-1/2} w \left( \frac{t - \tau}{a} \right) \right] \frac{dad\tau}{a^2}, \quad (\text{II.5.6})$$

where  $c_w$  is a constant depending only on the chosen mother wavelet  $w$ . However, a successful reconstruction does require a mother wavelet to be either real or complex analytic. The inverse transformation remains valid even if the wavelets are not orthogonal, but compose a redundant basis. The wavelets, however, do have to satisfy the before-mentioned criteria. Nevertheless, signals can only be reconstructed with zero mean (Rioul and Vetterli, 1991).

The possibility of signal reconstruction indicates that the WT preserves energy. The energy in wavelet domain is expressed by

$$E_s = \int |s(\tau)|^2 d\tau = c_w \int \int_{a>0} |\mathcal{W}[s](a, \tau)|^2 a^{-2} dad\tau. \quad (\text{II.5.7})$$

For more details on the WT, the interested reader is referred to the review articles of Rioul and Vetterli (1991) and Kumar and Foufoula-Georgiou (1997).

#### II.5.4.2 Slepian wavelets

As was already mentioned above, the exact shape of the mother wavelet greatly influences the obtained results. Thus, a suitable mother wavelet has to be chosen. To assess the influence of the mother wavelet on the results, a technique will be used which is capable of automatically creating a whole class of different wavelets, namely the so-called Slepian wavelets (Lilly and Park, 1995). These wavelets are not only easy to generate, but also complex analytic. Hence, reconstruction is possible.



This last property is also required to allow for the calculation of the so-called scalogram, the WT-equivalent of the spectrogram (i.e. squared amplitude spectrum). Hence, results produced by the use of these wavelets remain insensitive to any phase perturbations of the searched for reflections, just like the use of the local amplitude spectra of the STFT.

Another useful property of the Slepian wavelets is the possibility to create both time- and frequency-concentrated wavelets, i.e. emphasis can be placed on, respectively, a time or frequency analysis of the data. In this application, we are mainly interested in the position of reflections. Therefore, Slepian wavelets will be used which have a more time-concentrated character. Moreover, as the first Slepian wavelet is always very similar to a Morlet wavelet, the influence of this particular wavelet on the results is also examined.

The Slepian wavelets are a result of the classical problem of how to obtain accurate and stable estimates of the signal spectrum if the signal is due to a stochastic process or if it is deterministic but contaminated by noise. Naturally, an estimate of the signal spectrum is nearly always helpful in analyzing data. However, if the recorded signal is either due to a stochastic process or noise is superposed on the deterministic signal then the signal spectrum cannot be estimated accurately. The estimation will always fluctuate around some average value even if several independent realizations of the identical process exist (Lilly and Park, 1995). A related problem is the 'spectral leakage' of frequencies from outside the band of interest. In view of these problems, Lilly and Park (1995) desired to devise wavelets which 1) are localized both in time and frequency and 2) give an accurate estimation of the signal spectrum.

Although both properties are closely related, the first property demands for finite support in both time and frequency which is physically impossible (Slepian, 1983). However, an eigenvalue problem can be constructed which leads to functions which are bounded in time and have an optimized and nearly band-limited distribution in the frequency domain, in analogy with the discrete prolate spheroidal sequences of Thomson (1982) and Slepian (1983). Using these sequences, Park *et al.* (1987) were able to develop a multi-taper method which permits to obtain the desired accurate estimation of the spectrum.

Lilly and Park (1995) generalized the multi-taper method of Park *et al.* (1987). This generalization led to a class of wavelets which give statistically independent estimates of spectra with excellent resistance to spectral leakage. These wavelets were obtained by devising an eigenvalue problem which maximizes the fraction of energy  $\lambda$  between the frequency band  $|f \pm f_c| \leq f_w$ , i.e.

$$\lambda = \frac{\int_{-(f_c+f_w)}^{f_c+f_w} |W(f)|^2 df - \int_{-(f_c-f_w)}^{f_c-f_w} |W(f)|^2 df}{\int_{-1/2\Delta t}^{1/2\Delta t} |W(f)|^2 df}, \quad (\text{II.5.8})$$

where  $W(f)$  represents the discrete Fourier Transform of a discrete wavelet  $w_m$  with  $m = 1, \dots, M$ , i.e.

$$W(f) = \sum_{m=-M/2+1}^{M/2} w_m e^{2\pi i f m} \quad (\text{II.5.9})$$

and  $f$  is a continuous variable in the interval  $f \in [-1/2, 1/2]$ .

Optimization of the energy in this frequency band results in an eigenvalue problem  $\mathbf{A} \cdot \mathbf{w} = \lambda \mathbf{w}$ , where

$$A_{mn} = \frac{\sin[2\pi(f_c + f_w)\Delta t(m-n)]}{\pi(m-n)\Delta t} - \frac{\sin[2\pi(f_c - f_w)\Delta t(m-n)]}{\pi(m-n)\Delta t}. \quad (\text{II.5.10})$$

Solving the eigenvalue problem leads to  $M$  independent solutions  $w^{(k)}$  with eigenvalues  $1 > \lambda_1 > \lambda_2 > \dots > \lambda_M \geq 0$ . Use will only be made of the 'eigenwavelets' with eigenvalues close to 1, since only these have their spectral energy within the band of interest. The reader is referred to Lilly and Park (1995) for more details and for an application to Bear and Pavlis (1997).

Instead of using the deterministic variables  $f_w$  and  $f_c$  (which lose their sense for wavelets), it is more convenient to use the dimensionless values of, respectively, the time-bandwidth product,  $p = f_w M \Delta t$ , and the time-bandcenter product,  $p_c = f_c M \Delta t$ , with  $\Delta t$  the sample interval.

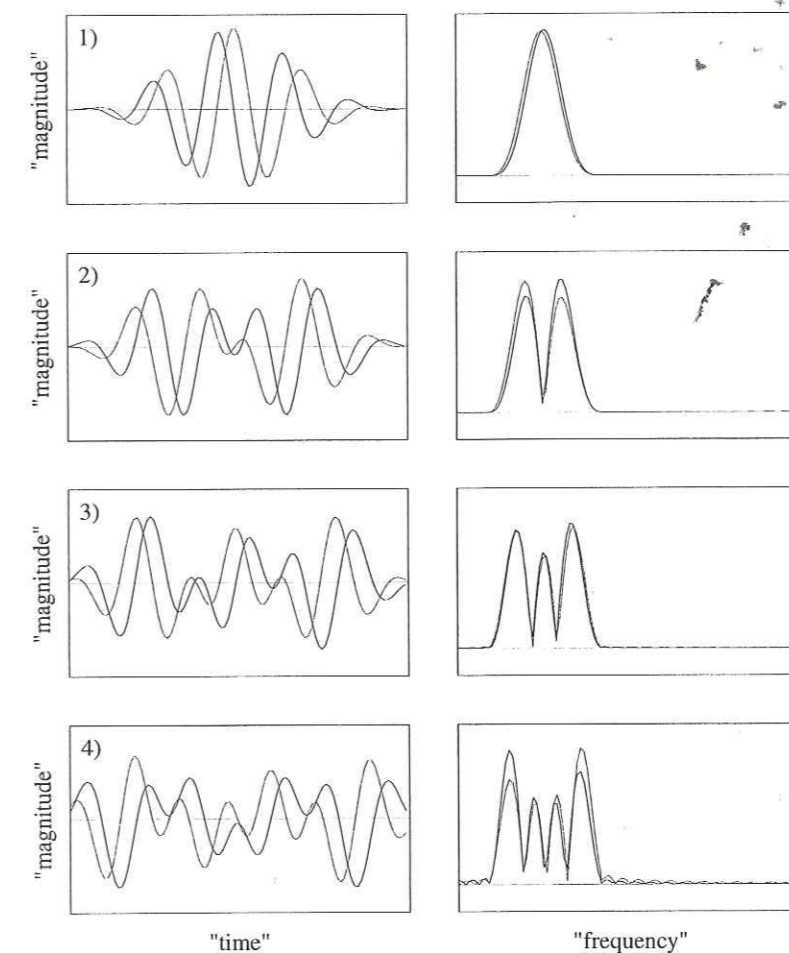


Figure II.5.3: First 4 used Slepian wavelets in both time and frequency domain with  $p = 3.0$  and  $p_c = 5.0$ , resulting in time-concentrated wavelets. After Lilly and Park (1995).

Besides their spectral properties, the so-called Slepian wavelets have some other interesting features, namely

1) The optimization is done for both positive and negative frequencies, leading to real wavelets. However, each wavelet comes with a counterpart having a  $90^\circ$  phase shift. Hence, complex analytic wavelets can be constructed.

2) The wavelets result from an eigenvalue problem. Thus, each individual wavelet is orthogonal. That is, wavelets are orthogonal for constant shift and scale. Notice that non-related classes of mother wavelets exist producing wavelets which are orthogonal independent of shift (position) and scale. However, such wavelets are not considered here.

Nevertheless, statistically independent estimates can be obtained using Slepian wavelets.



Such estimates can e.g. be used to obtain automatically uncertainty bounds (Bear and Pavlis, 1997).

3) Depending on the values of  $p$  and  $p_c$ , wavelets can be constructed with either a good time or a good frequency resolution. Namely, a large relative bandwidth  $p/p_c$  leads to time-concentrated wavelets, whereas  $p/p_c$  smaller will result in more frequency-concentrated wavelets. Therefore, a class of wavelets can be constructed with such properties as demanded for by a particular application.

In addition, within a particular class of wavelets, it can be said in general that the complexity of the spectral distribution increases with eigenvalue number (see Figure II.5.3). Moreover, the first wavelet has always a strong resemblance with the classical Morlet wavelet.

As a final remark, it should be noted that the Slepian wavelets are not continuous functions but discrete. Thus, often interpolation is necessary to perform a WT. The interpolation quality depends mainly on the choice of  $M$ . For values less than 100, the more complicated wavelets may not always be accurately constructed.

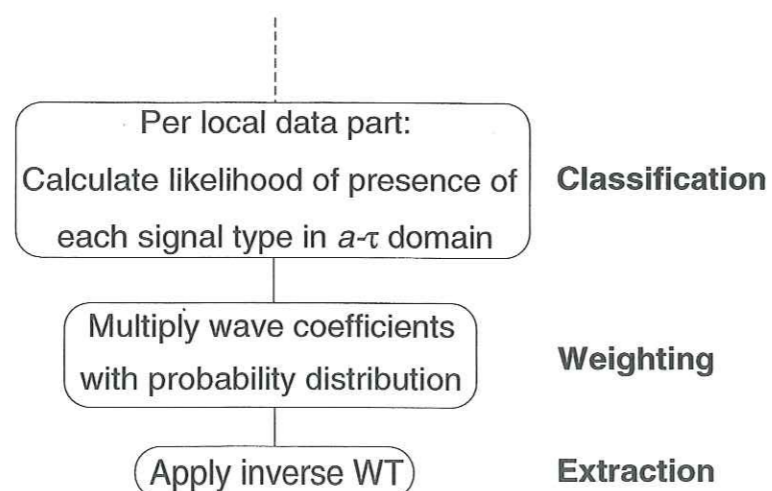


Figure II.5.4: Flow diagram for WT. To be compared with Figure II.5.1.

### II.5.4.3 Application to recognition

Application to recognition is straightforward. As is the case for recognition using an STFT, first a characteristic training set has to be selected, i.e. a reference reflection, and the wavelet coefficients are calculated for each pick. Next, averages and variances are estimated, which can then be used to obtain the likelihood distributions of the presence of reflection energy using equation (II.5.1). See also Figure II.5.1 for the analogous case of MLC using STFT.

The wavelet coefficients are calculated from discretized versions of the forward and inverse WT (respectively equations (II.5.3) and (II.5.6)) with logarithmic incremental steps of scale and linear shifts. Only 2 differences occur. Firstly, like in section II.3, energy has to be normalized for each local time  $\tau$  for an accurate estimation of probabilities. This normalization is slightly more difficult than in the previous case, since a correction for each individual scale has to be done using equation (II.5.7) for constant  $\tau$ . This correction includes a normalization factor  $a^{-2}$  for each single scale  $a$ . Notice that if a Principal Component Analysis is applied using equations (II.3.8) and (II.3.10), energy and thereafter vector length has to be normalized for a successful application.

Secondly, extraction cannot be done in time domain, because the obtained local probability is not directly related to a single time point, but to a whole range of adjacent time values. Thus, a multiplication of the likelihood distribution in time domain with the original data won't do. However, as the WT is invertible, extraction can be directly done in the  $a$ - $\tau$  domain by means of a multiplication of the wavelet coefficients with the derived probability distribution. Thereafter, application of the inverse WT, equation (II.5.6), will result in a 'cleaned' section up to a constant  $c_w$  which is simply neglected (see Figure II.5.4).

### II.5.5 A comparison

In this section a comparison will be made of the two different transforms for characterizing signals (STFT and WT) to determine the most suitable combination for detecting and extraction reflections and thereby enhancing the S/N-ratio. Not only the final extracted sections will be compared, but also computation time. The strategy is straightforward. Reflections are picked to estimate the characteristics of reflection energy. Next, the presence of reflection energy is estimated and 'cleaned-up' sections are produced.

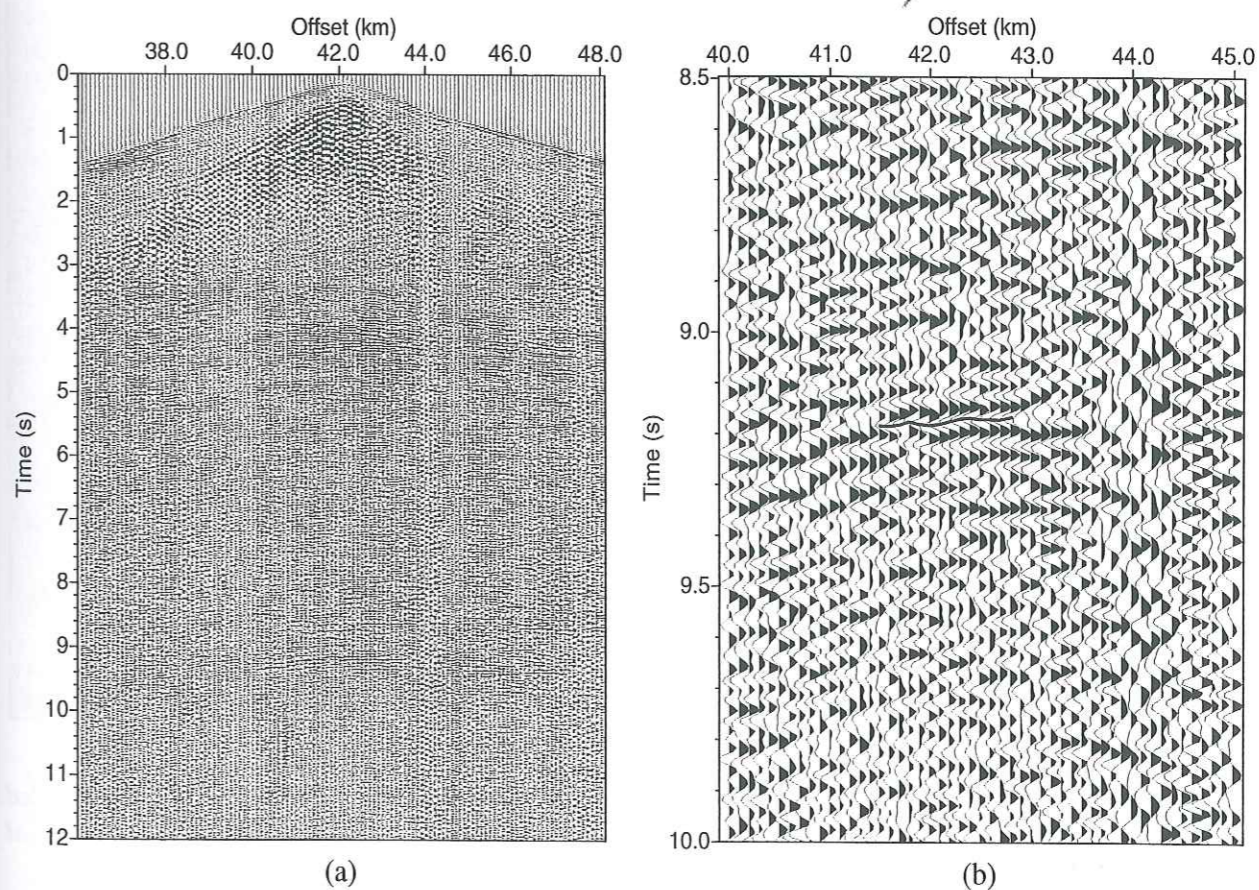


Figure II.5.5: a) The employed common shot gather for testing the different recognition methods and b) the 14 picks of the Moho reflection. Identical to Figure II.3.1 (reproduced for convenience).

To classify data, the MLC-technique used equation (II.5.1) for both transforms. The STFT employed always a 128 ms (16 point) sliding window with a 25 % cosine-taper, resulting in 9



frequencies in the amplitude spectra, whereas the WT used 10 logarithmic incremental scale steps of 1.342 points (i.e. powers of) with a minimum length of 9 points and shifts of 1 data point (8 ms). Especially for the longer scales, these shifts may have been too small and thus a reduction of computation time could have been obtained. Always the first four Slepian wavelets have been looked at with  $p = 3.0$  and  $p_c = 5.0$ , resulting in time-concentrated wavelets. Figure II.5.3 displays these wavelets in both time and frequency domain.

The used dataset is identical to the one used in sections II.3 and II.4. For more background, the interested reader is referred to section II.2. Methods are tested only on a single common shot gather (CSG). However, section II.3 shows that identical conclusions and results hold true for stacks. The used CSG is displayed in Figure II.5.5.a which shows clearly a limited S/N-ratio.

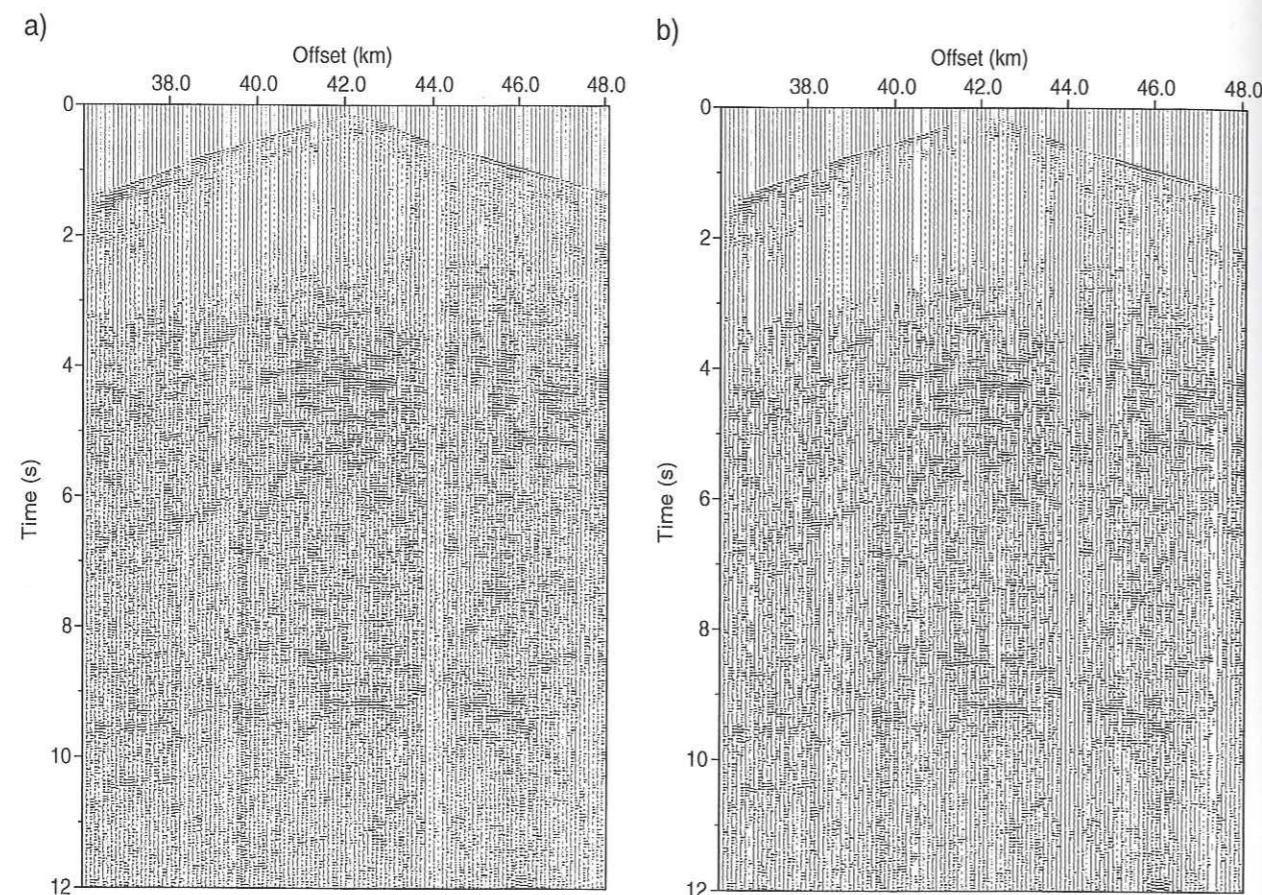


Figure II.5.6: Extraction results of the MLC-technique using an STFT analysis with a) the scaled likelihood distribution and b) a principal component analysis. To be compared with the original CSG displayed in figure II.5.5.a.

#### II.5.5.1 MLC using STFT

This is the method originally described in section II.3. Full details can therefore be found in that section.

The 14 picks used for a characterization of reflection energy are shown in Figure II.5.5.b. They are identical to those of section II.3. Extraction results are shown in Figure II.5.6.a.

Using a principal component analysis, equations (II.3.8) and (II.3.10) results in extracted section displaying slightly less background noise and therefore a slightly higher signal-to-noise ratio (Figure II.5.6.b).

The encountered inconveniences are the need of an *a priori* specification of window length and the underlying statistical distribution, although results remained quite stable if a longer window was used (256 ms instead of 128 ms). Nevertheless, a longer window resulted in extracted sections of a lesser quality (Figure II.5.10).

#### II.5.5.2 MLC using WT

The second tested combination consisted of MLC with the WT-technique for signal characterization. Since the MLC-technique does not necessitate the picks of non-reflections, the same picks were used as above (displayed in Figure II.5.5.b). Four different Slepian wavelets were tested in which the first was nearly identical to a Morlet wavelet. The wavelets were created such that an accurate time localization of the reflections was possible.

Figures II.5.7.a to II.5.7.d show the extracted sections using the first 4 Slepian wavelets (as displayed in Figure II.5.3). A large variety in extraction results is visible, although results remain comparable to those using an STFT (Figure II.5.6). Therefore, it can be concluded that the exact shape of the mother wavelet influences the results. Moreover, results depend also partly on the parameter settings of the discretization (i.e. step size and minimum length of wavelet). A better discretization using a smaller incremental step will result in general in more confident results. For instance, using 15 incremental step sizes of 1.184 points (instead of 10 steps of 1.342 points) produces more similar extraction results for the first 3 Slepian wavelets. From the fourth wavelet on, results seem to be less strict (more energy is extracted). The use of a principal component analysis results in most cases in a less strict (or vaguer) extraction. However, its exact influence remains unclear.

#### II.5.6 Discussion

A close comparison of the independently derived results displayed in Figures II.5.6 and II.5.7 shows that both transforms produce more or less identical extracted sections with S/N-ratios clearly higher than the original CSG in Figure II.5.5.a. The WT is able to produce slightly more and slightly less selective results using different mother wavelets.

A disadvantage of the WT is its computation time. As data is analyzed over several scales, more computation time is required than needed by an STFT. Nevertheless, applying undersampling of the data signals at the larger scales should save some time.

The use of the WT is very often advocated for analyzing non-stationary signals like seismic data, since it is capable of characterizing at the same time long and short period data. The STFT-technique is less suitable for this kind of data, due to its constant window length. Thus, the WT seemed to be very promising for signal recognition and, in particular, extraction of reflection energy. However, Figures II.5.6 and II.5.7 show that nearly identical results can be obtained using both techniques. Although Figure II.5.7 displays sometimes stricter extraction results than Figure II.5.6, behavior of the scaled likelihoods can be adapted manually by decreasing or increasing the constant  $c$  in equation (II.5.1). Thus, extraction results cannot only be modified using different mother wavelets, but also by a simple adjustment of the scaled likelihoods.

Moreover, the sole objective of the tested recognition techniques was the detection and extraction of a single type of signal. Thus, the optimum window length for the STFT-technique



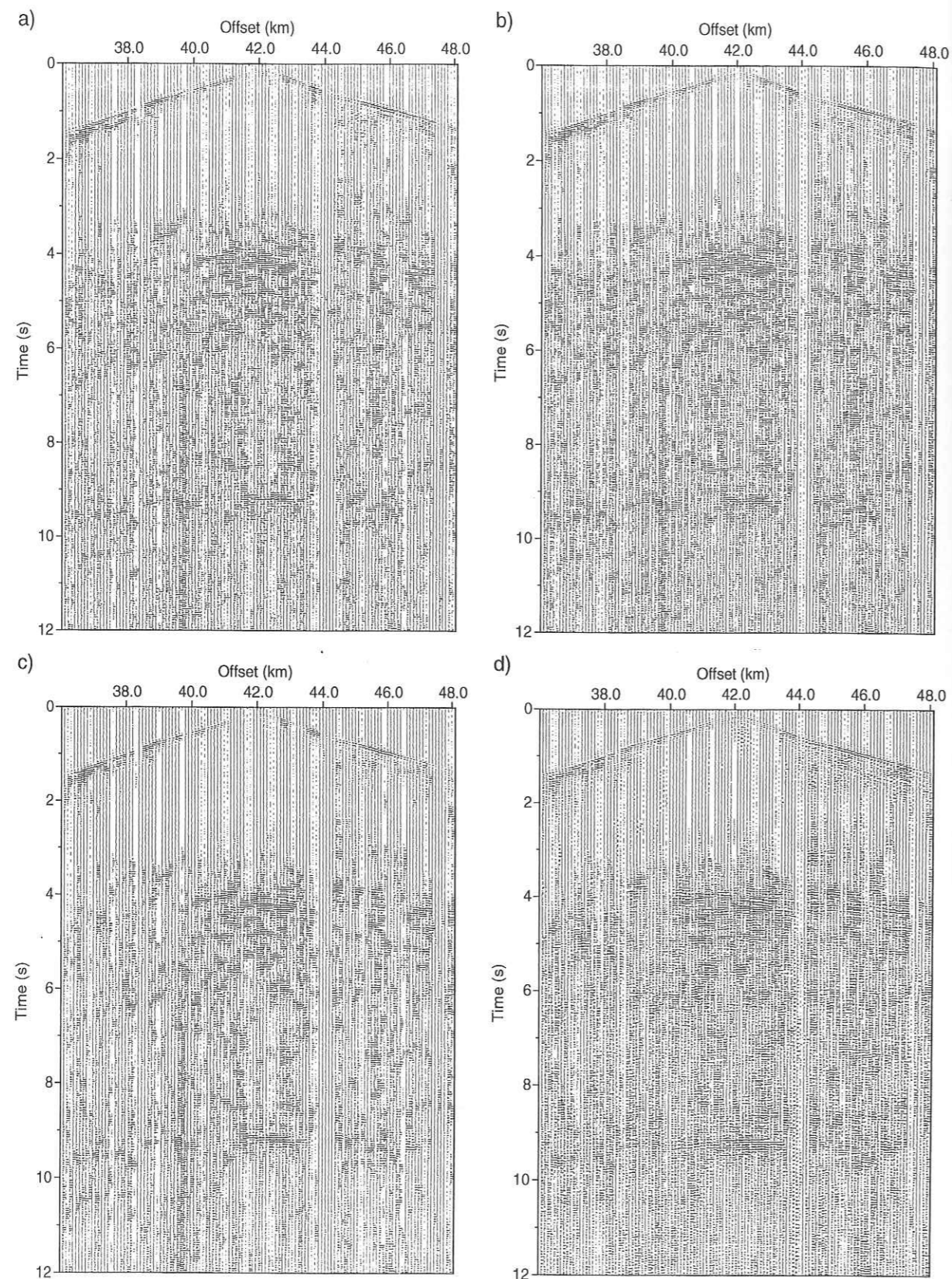


Figure II.5.7: As Figure II.5.6.a, but now using the first four Slepian wavelets (as shown in Figure II.5.3).

could more easily be estimated. However, if it had been desired to detect and extract simultaneously several signal types with different characteristic scale lengths (i.e. principal periods) then application of a WT would probably be more suitable.

In addition, in the case of very close to partly overlapping signals, the WT will offer most probably more possibilities, as the WT is better able to detect the characteristic scale lengths which allow for an optimum extraction. Although the superposition of different signal types in seismics is a common feature (e.g. reflections hidden by ground roll), any recognition or classification technique can only find the dominant signal. Therefore, using the WT-technique in combination with a recognition method will not enable us to extract hidden features. Such an application is a filtering problem and lies, therefore, beyond the scope of this paper.

As a last word concerning mother wavelets, it should be noted that the use of Slepian wavelets has made it evident that, to a large extent, results are independent of the exact form of the mother wavelet. However, to improve results, the used mother wavelet should reflect the complexity of the source wavelet, i.e. a simple source wavelet should not be dealt with by a mother wavelet having a complex frequency spectrum.

Fedi and Quarta (1998) use an interesting method to select the mother wavelet that produces the best results. They make use of the minimum entropy compactness criterion. This criterion is minimal for that particular wavelet needing the smallest number of non-negligible wavelet coefficients for an accurate description of the signal (in this case the reference examples). Therefore, this wavelet will yield an optimal decomposition of the reference examples, thereby allowing for an excellent distinction between this signal and other signals. Their criterion is strongly related to shortest description length criterions (Saito, 1994; Fedi and Quarta, 1998).

Unfortunately, this criterion cannot be applied in combination with a principal component analysis, as such an analysis reduces the number of coefficients to a single one. Hence, the number of non-negligible wavelet coefficients always equals one. On the other hand, the criterion might be applied on the eigenvectors  $\mathbf{u}^{(i)}$  of equation (II.3.10).

Finally, it should be noted that all tested recognition methods do not use information contained in the adjacent traces, but produce results for each trace separately. This may result in fragmentary images of the reflector positions as was shown in section II.3. However, a simple trace interpolation scheme (e.g. Kong et al., 1985), applied on the extracted sections, will emphasize laterally coherent energy and can, thereby, reconstruct the desired reflections even more clearly.

In addition, results are shown here only for CSGs. However, tests have shown that all tested recognition methods, once trained, can directly be applied on all CSGs contributing to a stacked section, without the need of constant repicking and retraining. Moreover, the method can also directly be applied after a stacked section has been obtained. In all cases, the S/N-ratio of the data will be enhanced significantly.

### II.5.7 Conclusions

Two different transforms for signal analysis have been tested for their ability of recognition of reflection energy. The use of the Wavelet Transform (WT) seems to be more justified for analyzing recorded non-stationary signals than the use of the Short Term Fourier Transform (STFT), due to Heisenberg's Uncertainty Principle. Moreover, the choice of an optimum window length of the STFT for characterizing simultaneously signal types having different principal periods or scale lengths is not without a problem. The WT-technique, on the other hand, performs a scale analysis of the data. Hence, it can automatically extract these principal scale lengths. This, however, has as a direct result that a WT is computationally more expensive



than an STFT. Moreover, if only a single type of signal has to be detected and extracted, this inconvenience of the STFT will not pose too frequently a problem. In addition, results of a WT-analysis still depend to some extent on the chosen mother wavelet.

Therefore, it may be concluded that, for the automatic detection and extraction of reflection energy, the most suitable combination still consists of the MLC-technique using STFT for signal characterization, at least in this particular application.

Possible applications of the recognition technique are manifold. Since it is capable of detecting and extracting reflection energy, the signal-to-noise ratio can be enhanced. Thus, interpretation is facilitated and stack degeneration and migration artifacts can readily be reduced. In addition, the MLC-technique can be used for interactive picking of reflections, possibly in combination with an automatic tracking algorithm. Finally, all recognition techniques can be used for data segmentation.

### II.5.8 Acknowledgments

The author wishes to thank Minh Nuygen and Jérôme Mars for introducing him to the wavelet domain.

### II.5.9 Extraction simultanée de plusieurs types de signaux

Dans les paragraphes précédents nous avons vu qu'une Transformée Continue en Ondelette (TCO) permet aussi d'extraire les réflexions. En effet, elle produit des résultats d'extraction qui sont de qualité comparable à ceux obtenus en utilisant une Transformée Fourier à Court Terme (TFCT).

En théorie, une TCO permet d'obtenir de meilleurs résultats qu'une TFCT à cause du fait qu'elle fait une analyse d'échelles, puisqu'elle n'utilise pas une longueur de fenêtre d'analyse fixée. Par conséquent, elle est mieux capable de distinguer entre des signaux ayant des fréquences caractéristiques différentes. Mais, si l'extraction d'un seul type de signal est désirée, cette analyse d'échelle n'aura pas beaucoup d'influence dans la plupart des cas. Ceci est probablement la raison pourquoi des résultats similaires ont été trouvés.

Afin d'exploiter au mieux les possibilités fournies par une TCO, nous refaisons le test fait dans la section II.4. Nous essayons donc d'extraire à la fois trois types de signaux en utilisant une TCO, à savoir les réflexions, les ondes de surface et tous les autres types de bruit.

Les pointés de référence sont montrés sur la figure II.4.5. La procédure d'extraction est similaire à celle utilisée dans les parties précédentes. Sur la figure II.5.8 nous montrons les résultats d'extraction en utilisant la méthode CVM (Classification par Vraisemblance Maximale) et sur la figure II.5.9 les résultats obtenus avec les RN (Réseaux de Neurones). La CVM a utilisé la première ondelette de Slepian (figure II.5.3.a). Donc, les résultats sur les figures II.5.7.a et II.5.8.a sont identiques. Les résultats d'extraction obtenus en utilisant une TFCT sont montrés sur les figures II.4.6 (CVM) et II.4.7 (RN).

Les résultats des figures II.5.9 et II.4.7 sont obtenus de la même manière. Donc, un RN de type 'MLP' et de forme 9-5-3 a été entraîné et ensuite élagué par la méthode 'NCU' de Sietsma et Dow (1991). Ceci a résulté dans un RN de type 3-2-3. Les trois entrées sont les échelles 5, 6 et 8. Ces échelles correspondent à des fréquences d'analyse de 19.2, 14.4 et 8 Hz.

Une comparaison des figures II.5.8 et II.5.9 nous montre que la méthode CVM et les RN ont obtenu des résultats similaires pour l'extraction des réflexions et des ondes de surface. Par contre, seuls les RN sont capables de bien extraire la troisième catégorie contenant tous les autres types de bruit. Les ondes de surface qui restent sur la figure II.5.9.b sont probablement dues à des effets de bords de l'analyse en ondelette. Ces effets de bords sont plus forts pour une TCO que pour une TFCT à cause des plus longues fenêtres d'analyse.

Les extractions des réflexions par la TCO sont plus strictes que par une TFCT (moins d'énergie a été extraite). Mais il est difficile de conclure quelle est la meilleure extraction.

Finalement, les extractions des ondes de surface par la TCO sont meilleures, car il reste moins de bruit de fond et de l'énergie isolée. Cette meilleure extraction est due à la fréquence principale des ondes de surface (entre 5 et 8 Hz). Cette période principale est plus longue que la longueur de la fenêtre d'analyse de la TFCT (128 ms). Par conséquent, cette longueur empêche une meilleure extraction. Les ondelettes, elles aussi, analysent les signaux à 8 Hz, mais elles utilisent une fenêtre de 562.5 ms.

Sur les figures II.5.10 et II.5.11 les résultats d'extraction sont montrés par une fenêtre d'analyse deux fois plus longue (256 ms au lieu de 128 ms) pour la TFCT. La conséquence est une nette dégradation des résultats d'extraction des réflexions, même si les résultats d'extraction des ondes de surface sont meilleurs. Cette dégradation est marquée par le fait qu'il y a beaucoup moins d'énergie extraite (figure II.5.10.a) ou beaucoup plus (figure II.5.10.b).



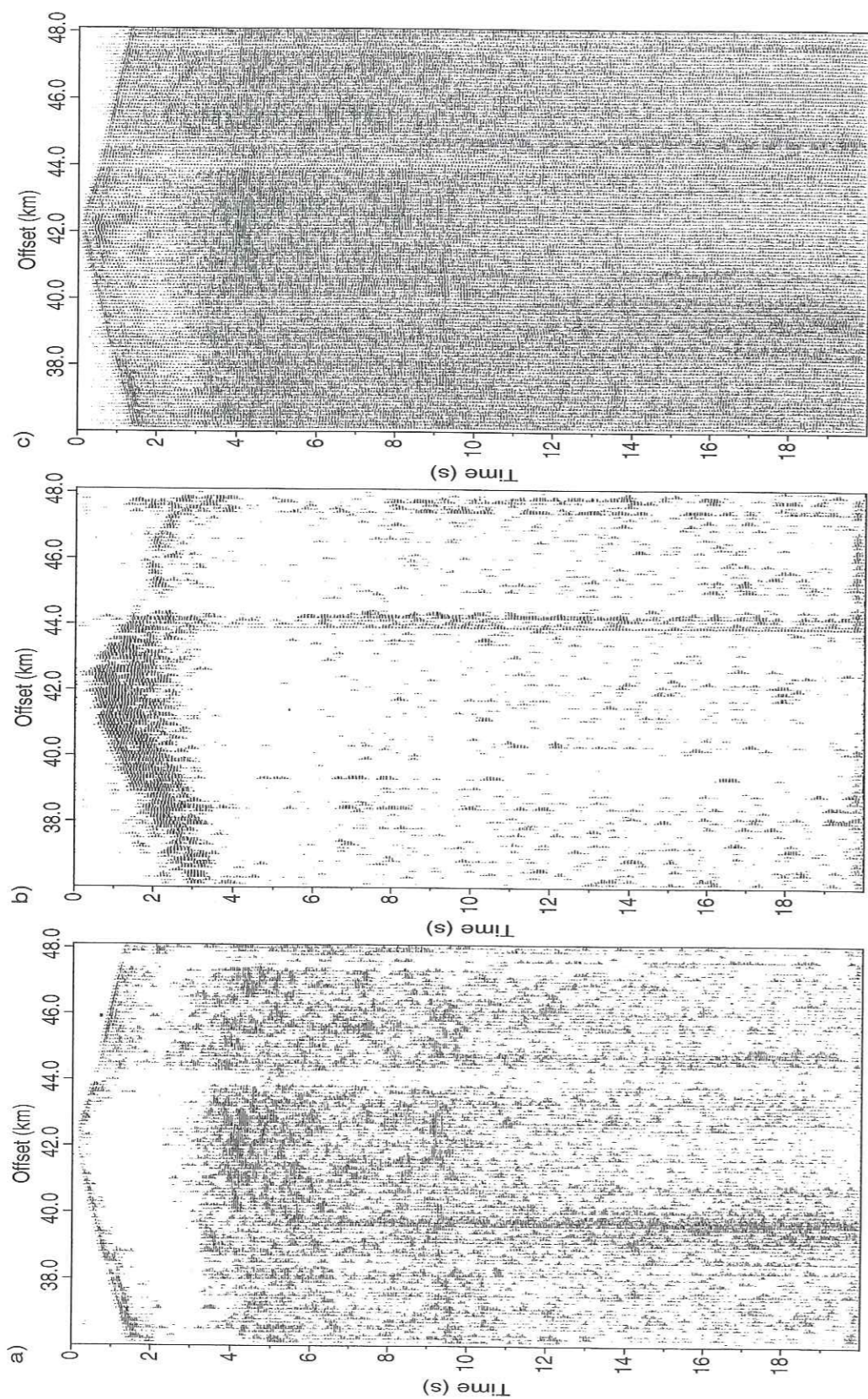


FIG. II.5.8 – Résultats d'extraction en utilisant la méthode CVM et une TCO. A comparer avec la figure II.4.6 où une TFCT a été utilisé. a) Réflexions, b) ondes de surface et c) autres types de bruit.

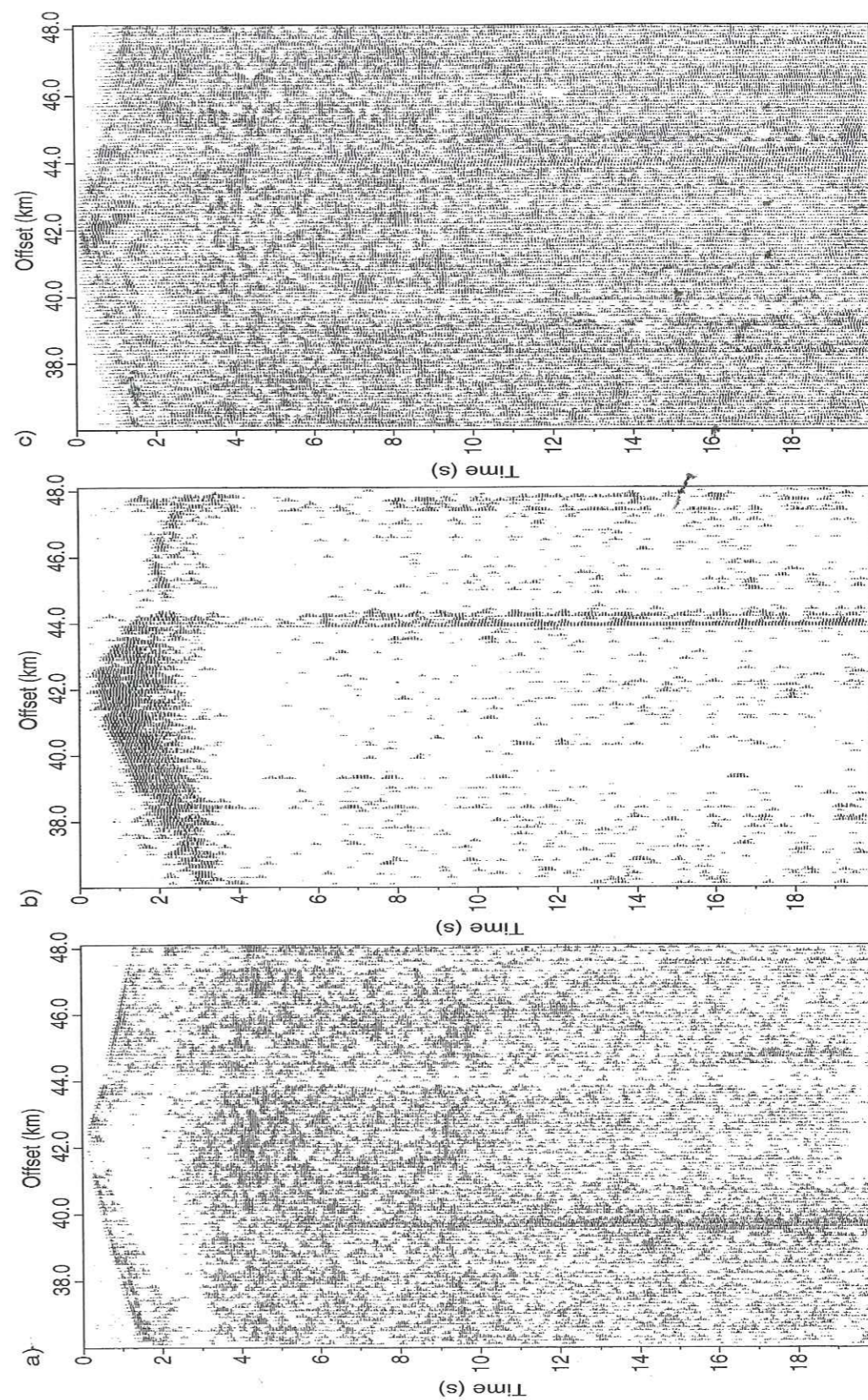


FIG. II.5.9 – Résultats d'extraction en utilisant un RN et une TCO. A comparer avec la figure II.4.7 où une TFCT a été utilisé. a) Réflexions, b) ondes de surface et c) autres types de bruit.



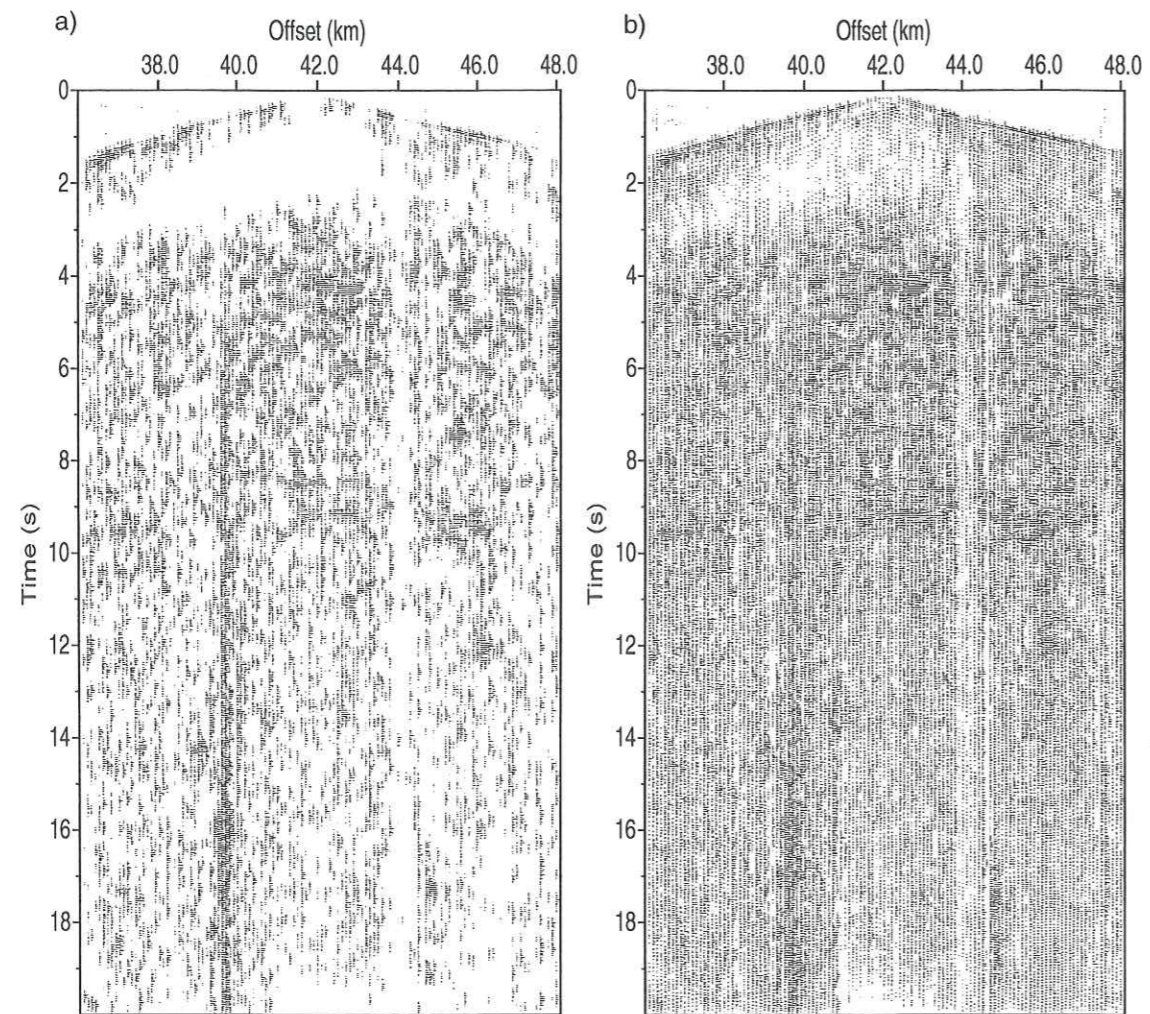


FIG. II.5.10 – Résultats d'extraction de la méthode CVM et une TFCT pour une fenêtre d'analyse de 256 ms. Extraction des réflexions en utilisant a) une distribution Gaussienne ou b) une analyse de composants principaux. A comparer avec figure II.4.6.a.

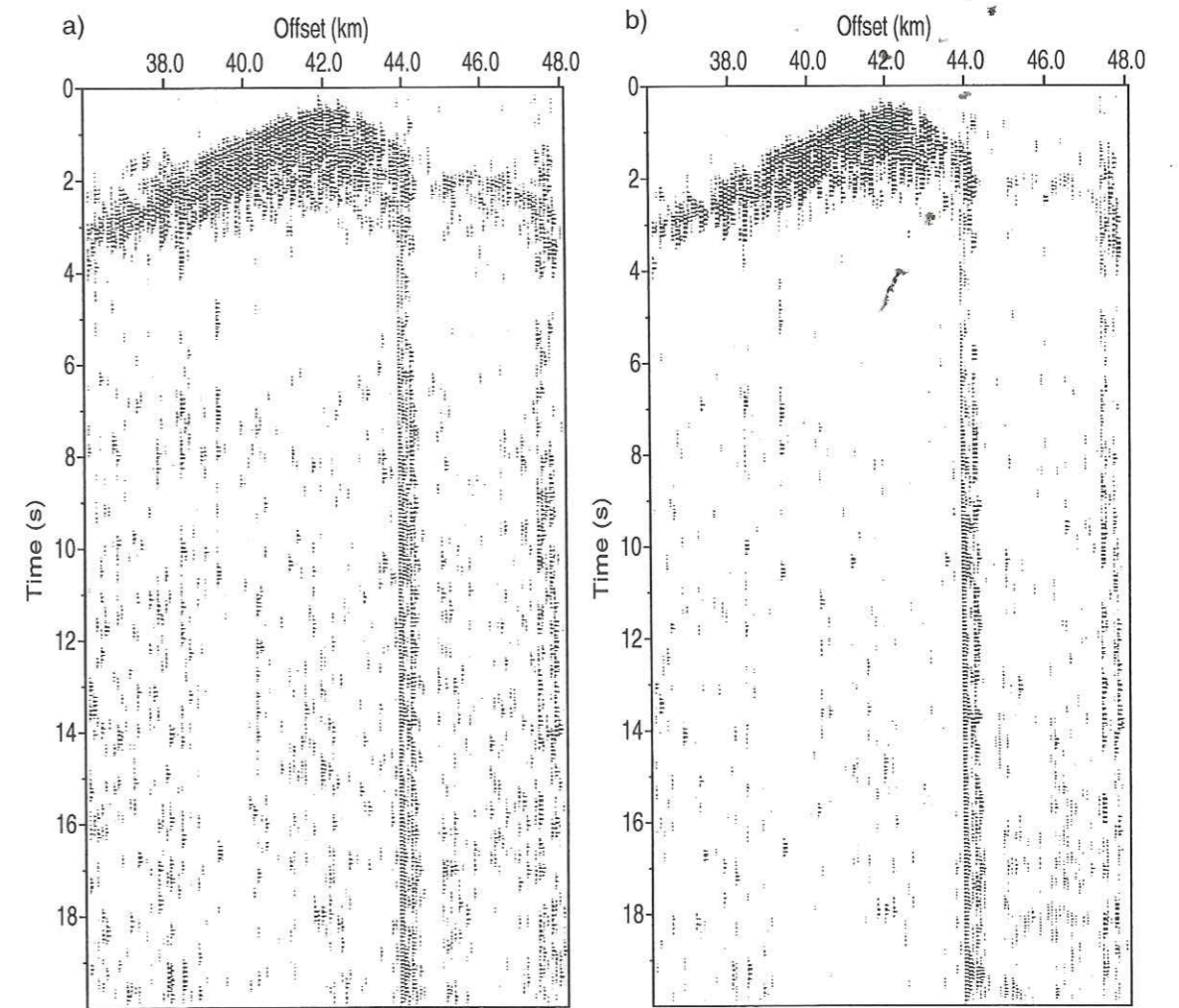


FIG. II.5.11 – Identique à la figure II.5.10. Extraction des ondes de surface par la méthode CVM et une TFCT pour une fenêtre d'analyse de 256 ms. a) Distribution Gaussienne et b) analyse de composants principaux. A comparer avec la figure II.4.6.b.



### II.5.10 La Transformée Continue en Ondelette et la détection des signaux: Conclusions

Les résultats obtenus nous ont montré qu'une Transformée Continue en Ondelette (TCO) et une Transformée de Fourier à Court Terme (TFCT) ont des avantages et des inconvénients opposés. D'une part, une TFCT est plus rapide, mais elle nécessite la spécification *a priori* de la longueur optimale de la fenêtre d'analyse. D'autre part, une TCO analyse les données sur des échelles différentes, mais elle est plus lente.

Par conséquent, pour l'extraction d'un seul type de signal, la TFCT est recommandée, puisqu'une analyse d'échelles n'est pas nécessaire et elle est plus rapide. Par contre, pour l'extraction simultanée de plusieurs types de signaux, il vaut mieux utiliser une TCO, puisqu'il est souvent difficile de définir la longueur optimale de la fenêtre d'analyse et par conséquent une analyse d'échelle est appropriée.

De nouveau, nous avons vu que les réseaux de neurones sont plus adaptés aux problèmes multi-modaux de classification qu'une méthode de Classification par Vraisemblance Maximale.

Finalement, une analyse de composants principaux appliquée aux attributs décrivant la signature locale du signal ne donne pas toujours de meilleurs résultats qu'une distribution Gaussienne pondérée et corrigée pour la longueur de la fenêtre d'analyse.

## II.6 Les données Ecors 2.5-D sommées et migrées

### Résumé

En comparant les résultats obtenus par Siliqi (1994) avec ceux obtenus par nos méthodes de détection et de reconstruction de signaux, nous montrons que 1) une application à 3 dimensions de ces techniques est indispensable, que 2) les méthodes actuelles de migration ne sont pas adéquates pour les données de sismique réflexion profonde et 3) que nos méthodes donnent des résultats pour la partie supérieure du profil avant migration qui sont similaires ou supérieures à ceux de Siliqi (1994) en ne traitant que 7 % des données. De plus, l'avantage de nos méthodes est d'être automatiques, alors que Siliqi (1994) a dû faire un pointé manuel de toutes les bonnes réflexions dans tout le volume des données. Finalement, le danger d'un pointé manuel est de négliger des structures qui n'apparaissent que clairement après migration, comme par exemple le synclinal d'Orriule.

### II.6.1 Introduction

Dans cette dernière étude, nous comparons les résultats obtenus en utilisant les méthodes décrites d'extraction et de reconstruction de signaux avec ceux obtenus par Siliqi (1994). Toutes les méthodes sont appliquées aux données Ecors 2.5-D (*cf.* section II.2).

Nos méthodes et celles utilisées par Siliqi (1994) diffèrent beaucoup. La seule analogie est que les deux méthodes utilisent une analyse  $\tau$ - $p$  afin d'extraire les réflexions et de reconstruire toute énergie 'manquante'. Cependant, nos méthodes sont complètement automatiques, tandis que la méthode de Siliqi nécessite une sélection manuelle des 'bonnes' traces afin d'arriver à un résultat correct. Ceci donne sans doute de meilleurs résultats, mais une telle sélection est extrêmement longue et fastidieuse. De plus, Siliqi a traité toutes les données, alors que nous n'en avons utilisé qu'une partie très limitée (moins que 7 %). Malgré tout, une comparaison directe des deux méthodes permet de bien évaluer la performance de nos méthodes d'extraction et de reconstruction.

Cette étude est composée de deux parties. D'abord nous comparons les coupes-somme et ensuite les coupes-somme migrées. Chaque partie commence avec une description plus étendue des méthodes utilisées. Finalement, nous proposons quelques améliorations des techniques utilisées. Il s'agit surtout des techniques de reconstruction et de migration.

### II.6.2 Une comparaison des résultats obtenus

#### II.6.2.1 La coupe-somme

Nous utilisons les lignes 130 et 132 de la coupe-somme afin de tester les méthodes proposées d'extraction de réflexions. Ces deux lignes de la coupe-somme se trouvent juste à côté de la ligne 4 de capteurs (figure II.2.4). Dans la section II.3, nous avons utilisé la ligne 133 (figures II.3.6 et II.3.7). Par conséquent, toutes les coupes-somme montrées se ressemblent. Afin de bien pouvoir comparer les résultats de cette partie avec ceux présentés dans la section II.3, le lecteur doit savoir que toutes les figures de la section II.3 sont montrées du nord au sud (de gauche à droite) alors que c'est le contraire ici. Pour les lecteurs intéressés par une comparaison des résultats obtenus par Siliqi (1994) et nous aux interprétations du profil Ecors Arzacq (Daignières et al., 1994), il faut savoir que le profil Arzacq passait par la ligne 132 de la coupe-somme du profil Ecors 2.5-D.



**Traitement** Afin d'augmenter le rapport signal sur bruit, nous utilisons la méthode décrite dans la section II.3, c'est-à-dire la méthode CVM avec une TFCT. Nous ne montrons que les résultats obtenus par cette combinaison, car toutes les autres combinaisons donnent des résultats similaires (section II.5). Nous avons choisi d'utiliser la TFCT, car elle est plus rapide qu'une TCO et parce que nous ne cherchons à extraire qu'un seul type de signal, les réflexions.

Les pointés de référence sont ceux faits sur le Moho et ils sont montrés sur la figure II.3.1. Afin d'augmenter la qualité de la coupe-somme, nous avons appliqué le schéma déjà utilisé dans la section II.3. Nous avons donc utilisé la procédure suivante:

- extraction des réflexions dans toutes les collections de traces à point de tir commun qui contribuent à la coupe-somme (en utilisant une fenêtre de 128 ms);
- application de la technique de reconstruction (Kong et al., 1985) après extraction sur les mêmes collections de traces (en utilisant des ensembles de 10 traces et des déplacements de 5 traces);
- sommation des traces traitées (avec une taille des pavés de 50 x 50 m et des distances maximales source - capteur limitées à 6 km);
- extraction des réflexions dans la coupe-somme (en utilisant de nouveau une fenêtre de 128 ms et les mêmes pointés de référence sur le Moho);
- application de la technique de reconstruction sur la coupe-somme après extraction (en utilisant des ensembles de 40 traces et déplacements de 20 traces).

La deuxième application de la technique d'extraction est nécessaire afin d'enlever toute énergie non-cohérente introduite par la sommation des traces. La deuxième reconstruction utilisait 40 traces par ensemble afin d'extraire seulement les grandes structures géologiques (40 traces correspondent à une distance de 2 km). La coupe-somme résultante est montrée dans la figure II.6.1.a.

Siliqi (1994) a choisi une autre approche afin de tenter de résoudre le problème de la faible qualité des données Ecors 2.5-D. Il n'a effectué aucun traitement des traces avant sommation, mais il a fait la sommation en utilisant les données brutes.

Comme nous il se sert de pavés de taille 50 x 50 m. Cette taille a l'inconvénient que 20 % des pavés ont une couverture nulle, car la taille normale des pavés serait de 50 x 200 m, puisqu'il y avait à l'acquisition 100 m entre capteurs et 400 m entre sources. Il a choisi cette taille pour qu'elle augmente légèrement la qualité après sommation. Par conséquent, cette taille lui a permis de mieux suivre les réflexions, ce qui a facilité le pointé.

Nous avons tous les deux utilisé le modèle de vitesse estimé par la CGG, qui a effectué le traitement préalable des données Ecors 2.5-D. Ce modèle est montré dans la figure 6 du chapitre 2 de la thèse de Siliqi.

Ensuite, afin d'extraire les réflexions d'une façon automatique, il a appliqué une analyse  $\tau$ - $p$ - $az$ , c'est-à-dire une analyse  $\tau$ - $p$  à 3 dimensions où la troisième dimension est représentée par l'azimut. En effet, il a effectué des sommations sur un disque plan avec un rayon de 200 m (incluant au total 49 traces). Au lieu d'utiliser la semblance pour faire 'ressortir' toute énergie cohérente, il a appliqué l'algorithme de Bano (1989) qui s'est inspiré de la méthode de Harlan et al. (1984). Cet algorithme applique une analyse  $\tau$ - $p$  sur une partie de la coupe-somme ne contenant que du bruit et calcule l'écart-type du bruit après les sommations obliques. Ensuite, la méthode utilise deux fois cet écart-type comme seuil d'amplitude sur une analyse  $\tau$ - $p$  des données avant d'appliquer la transformée  $\tau$ - $p$  inverse.

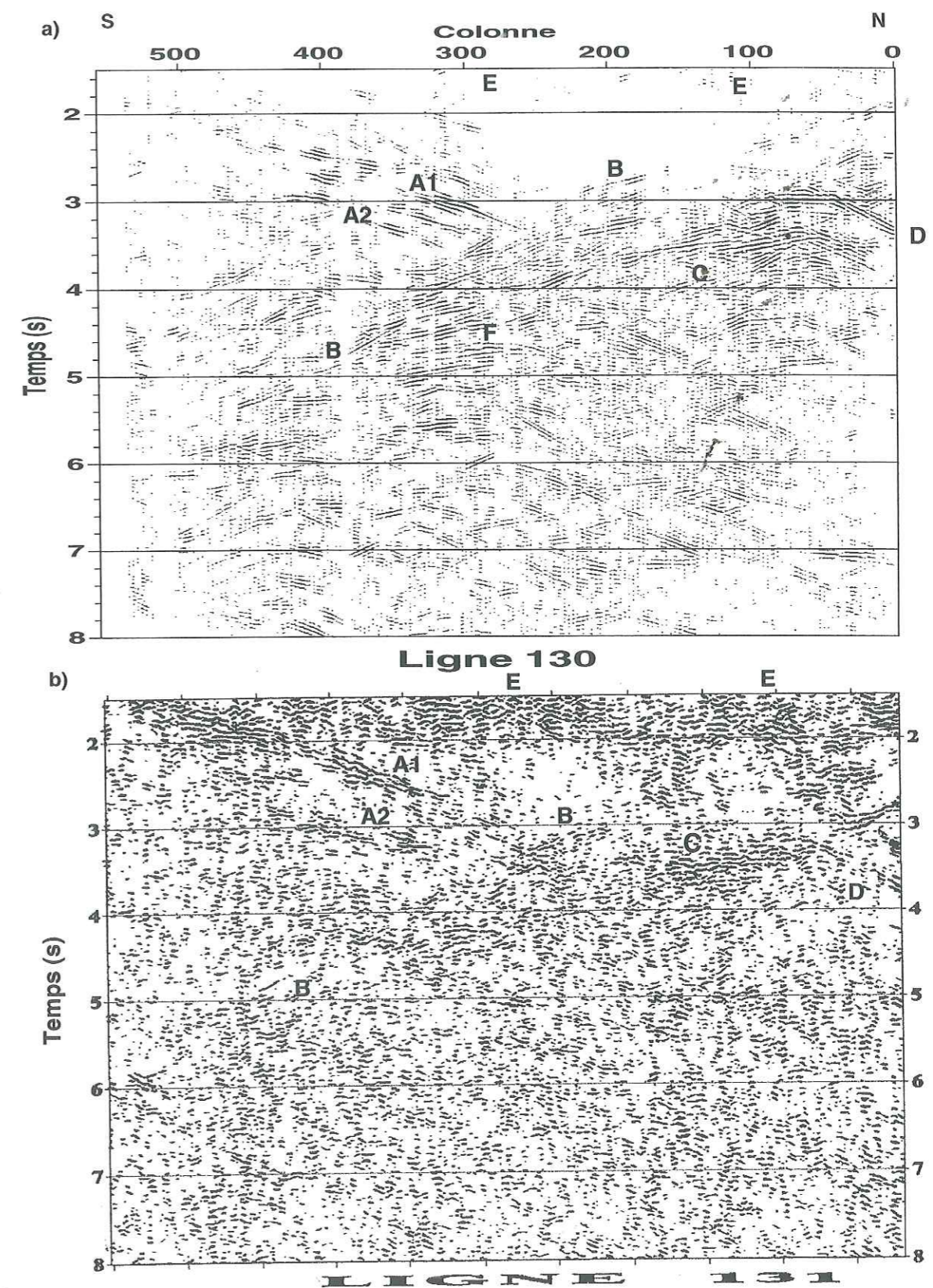


FIG. II.6.1 - Partie supérieure de 2 lignes voisines de la coupe-somme des données Ecors 2.5-D. Résultats obtenus a) par nous et b) par Siliqi (1994).



Siliqi (1994) a généralisé cette méthode à 3 dimensions et l'a appliquée aux données Ecors 2.5-D après sommation des traces. La figure II.6.1.b nous montre les résultats pour la ligne 131. Cette méthode n'a non seulement fait 'ressortir' les réflexions, mais elle a aussi permis de reconstruire toute énergie 'manquante' comme par exemple dans une grande partie des pavés à couverture nulle.

**La partie supérieure** Si nous examinons la coupe-somme de Siliqi (1994) de la partie supérieure des données (jusqu'à 8 s; figure II.6.1.b), nous voyons quelques structures majeures marquantes.

- D'abord, il y a 2 réflecteurs parallèles inclinés vers le nord. Le premier ( $A_1$ ) commence autour de la colonne 500 à 1.8 s et se poursuit jusqu'à la colonne 280 à 3.2 s. Le second ( $A_2$ ) se trouve plus bas et est plus court (de la col. 450 à 2.8 s jusqu'à la col. 350 à 3.2 s).
- Ensuite, il y a un réflecteur ( $B$ ) à pendage sud, qui commence à la col. 225 vers 3 s et qui continue jusqu'à la col. 450 à 5 s.
- Puis, nous pouvons voir quelques réflecteurs horizontaux ou légèrement inclinés vers le sud entre les col. 50 et 200 vers 3.5 s ( $C$ ).
- Ces réflecteurs sont bornés au nord par un autre réflecteur à pendage nord entre les col. 100 et 0 vers 3.2 à 3.8 s ( $D$ ).
- Finalement, la coupe-somme montre beaucoup d'énergie entre 1.5 s et 2 s ( $E$ ). Cette énergie n'est pas très cohérente, mais semble être liée à des réflecteurs horizontaux.

Nous retrouvons les mêmes structures géologiques sur la ligne 130 de notre coupe-somme. Nous n'avons pas utilisé la ligne 131, puisqu'elle est partiellement vide (à cause de la taille choisie pour les pavés). Cela n'a posé aucun problème à Siliqi, car, contrairement à nous, il a appliqué une analyse  $\tau$ - $p$  à 3 dimensions. Par conséquent, il a extrait de l'énergie des lignes voisines de la coupe-somme. Néanmoins, les mêmes structures géologiques sont présentes dans nos résultats.

- Comme lui, nous avons extrait des réflexions inclinées vers le nord de la col. 450 à 2.4 s jusqu'à la col. 275 à 3.2 s ( $A_1$ ). Un autre réflecteur parallèle ( $A_2$ ) se trouve légèrement plus bas.
- Nous pouvons voir le réflecteur ( $B$ ) incliné vers le sud qui commence vers la col. 175 à 2.7 s et qui se prolonge jusqu'à la col. 450 à 5.5 s. Ce réflecteur semble recouper les 2 réflecteurs précédents. Cela pourrait indiquer la présence d'un chevauchement dans le socle (comme nous l'avons déjà suggéré dans les conclusions de la section II.3). Par contre, le réflecteur à pendage sud recoupe aussi beaucoup d'autres réflecteurs ( $F$ ) inclinés vers le sud qui sont plus horizontaux (par exemple entre les col. 300 et 400 entre 4 et 5 s).
- Entre les col. 50 et 200, nous voyons beaucoup de réflecteurs semi-horizontaux vers 3.5 s ( $C$ ). Ces réflecteurs semblent dus à la présence d'un anticlinal.
- Le réflecteur ( $D$ ) incliné vers le nord entre les col. 80 et 0 de 2.8 à 3.4 s renforce l'idée de la présence d'un anticlinal.
- Finalement, nous n'avons extrait aucune énergie entre 1.5 et 2 s ( $E$ ). Ceci nous donne à croire que toute énergie à cette profondeur sur la figure II.6.1.b est due aux ondes de surface non enlevées par Siliqi. Par conséquent, cette énergie ne représente aucun réflecteur.

Même si les deux méthodes ne donnent pas des images identiques des réflecteurs présents dans les données, elles ont extrait les mêmes structures majeures. Donc, nous pouvons conclure que les deux méthodes sont capables d'augmenter la qualité des données.

**La partie profonde** Afin d'extraire la réflexion sur le Moho, nous n'avons changé aucune étape dans notre procédure de traitement des données. Par contre, Siliqi (1994) a effectué un pointé manuel de toutes les 'bonnes' traces où le Moho est visible. Il n'a sommé que les traces montrant clairement la réflexion sur le Moho et il a ensuite appliqué sa méthode  $\tau$ - $p$ - $az$  pour augmenter la qualité de la coupe-somme de la partie inférieure des données (entre 7 et 12 s). Les résultats des deux coupes-sommes sont montrés sur la figure II.6.2 en utilisant la ligne 132.

Siliqi a trouvé le Moho sur 12 % des traces dispersées sur un quart des tirs. Il a montré que l'utilisation des 'bonnes' traces améliore considérablement le rapport signal sur bruit dans la partie profonde des données, malgré une grande augmentation du nombre de pavés à couverture nulle (de 20 % à 72 %). Même après l'analyse  $\tau$ - $p$ - $az$ , 46 % des pavés restent vides. Par contre, les figures 2 et 3 de son chapitre 6 montrent bien que le Moho n'est que difficilement visible en utilisant une analyse  $\tau$ - $p$ - $az$  après sommation de toutes les traces, tandis qu'il est facile à identifier en ne sommant que les 'bonnes' traces.

Dans la majorité des cas, on ne peut distinguer une réflexion nette sur le Moho dans les profils de sismique réflexion profonde. Par conséquent, le Moho est souvent défini comme la limite inférieure des réflexions continues. Avec cette définition, sur la figure II.6.2.b, le Moho se trouve à 9.4 s entre les colonnes 225 et 325 ( $M_1$ ), à 9.8 s au-delà la col. 150 ( $M_2$ ) et probablement à plus que 10 s entre les col. 0 et 100 ( $M_3$ ). Ceci indique un Moho à pendage nord. Autrement, dans la coupe-somme de Siliqi, nous pouvons voir deux parties montrant beaucoup d'énergie cohérente (entre les col. 280 et 325 et entre les col. 0 et 120) et une partie intermédiaire ne montrant presque pas d'énergie extraite.

Il est difficile de trouver le Moho dans notre coupe-somme (figure II.6.2.a) en utilisant la définition courante indiquant la position du Moho. Cependant, nous voyons clairement une réflexion ( $M_1$ ) entre les col. 225 et 340 à 9.6 s, qui s'incline légèrement vers le sud. Par contre, la partie nord du profil est plus obscure. Une réflexion ressort autour de la col. 70 à 9.7 s ( $M_3$ ) et il y a quelques structures à côté à pendage nord ( $M_2$ ), mais aucune structure ne peut être identifiée au-delà 10 s. Par conséquent, nos résultats concernant le Moho sont difficiles à concilier avec ceux obtenus par Siliqi. Même l'alternance des bandes verticales noires et claires ne se trouve pas dans nos résultats. Dans notre cas, la quantité d'énergie extraite diminue simplement vers le nord.

Curieusement, dans la collection de traces à point à tir commun 8184 (figure II.3.1), nous avons pointé le Moho à 9.2 s. Cette collection de traces se trouve juste à côté de la ligne montrée (figure II.6.2.a). Par conséquent, on ne voit pas clairement à quoi correspond exactement l'énergie réfléchie autour de la col. 300 à 9.6 s (nous, figure II.6.2.a) ou à 9.4 s (Siliqi, figure II.6.2.b). Cependant, les 2 coupes-sommes montrent des réflexions à 9.2 s autour de cette colonne. Par conséquent, il est évident que la définition utilisée afin de localiser le Moho ne donne qu'une surestimation de la profondeur du Moho à cause des réflexions multiples tardives.

Pour récapituler, les deux méthodes d'amélioration de la qualité des données donnent des images différentes des localisations exactes des réflecteurs profondes. Néanmoins, les mêmes structures principales sont visibles dans les 2 extractions pour la partie supérieure du profil. Par contre, la partie profonde montre des grandes différences qui sont probablement dues à la sélection manuelle des 'bonnes' traces faite par Siliqi.



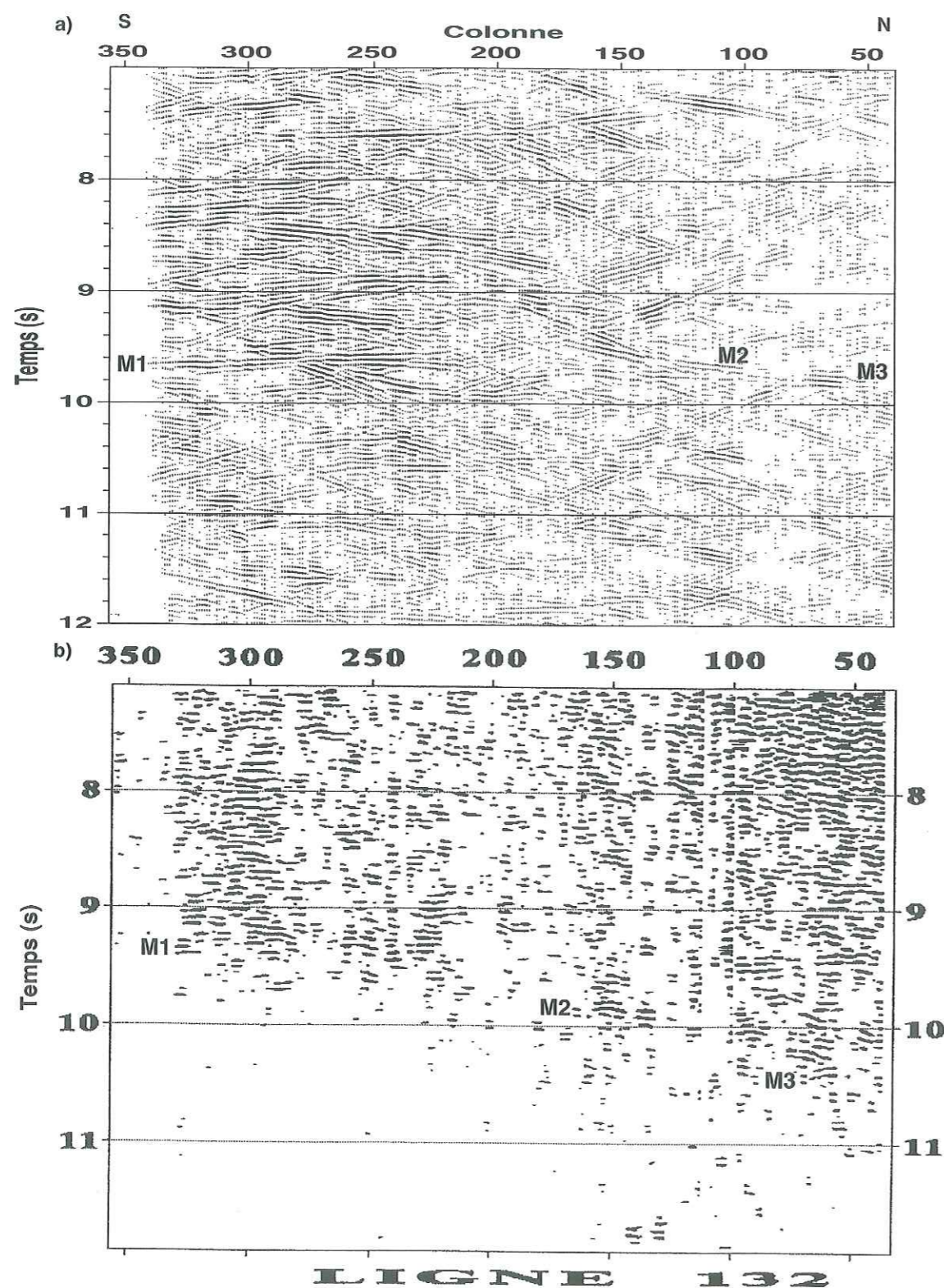


FIG. II.6.2 – Partie profonde de la ligne 132 de la coupe-somme obtenue a) par nous et b) par Siliqi (1994).

### II.6.2.2 Les données migrées

**Procédure de migration de Siliqi** De nouveau, Siliqi (1994) et nous n'avons pas migré les données de la même manière. Siliqi a appliqué une méthode géométrique simple ne migrant que des segments de droites, alors que nous avons utilisé une méthode classique de migration  $f-k$ , à savoir la méthode de déphasage de Gazdag (1978). Une autre différence entre les méthodes de migration est que Siliqi a fait une migration à 3 dimensions de toutes les données, alors que nous n'avons fait qu'une migration à 2 dimensions de quelques lignes de la coupe-somme.

Par conséquent, nous avons effectué des pré-traitements différents des données. Siliqi était obligé de pointer des segments de droites représentant les réflexions, tandis que nous avons simplement utilisé les données après sommation. Les avantages de la méthode de migration choisie par Siliqi est qu'elle permet d'avoir une influence directe sur les résultats (par exemple en ne pointant que les 'bonnes' réflexions) et qu'elle est beaucoup moins sensible à l'aliasage spatial et d'autres artefacts de migration. L'inconvénient est qu'elle nécessite de nouveau un pointé de toutes les données.

Afin d'effectuer une migration à 3 dimensions des données, Siliqi a choisi la méthode intéressante de Jakubowicz et Levin (1983). Cette méthode nous dit que nous pouvons faire des migrations à 3 dimensions dans un milieu à vitesse constante en faisant 2 migrations successives dans 2 directions perpendiculaires. Dans un milieu à vitesse constante, il suffit de migrer d'abord les lignes et ensuite les colonnes. Siliqi a migré les données Ecors 2.5-D en utilisant cette technique avec une vitesse de 5 km/s pour la partie supérieure et 6 km/s pour la partie profonde.

Ainsi, la procédure choisie par Siliqi est composée des étapes suivantes:

- D'abord, il a pointé des segments de droites le long des 160 lignes de la coupe-somme. Ceci a résulté en 20000 segments.
- Ensuite, il a migré les segments pointés le long des lignes.
- Puis, il a refait le pointé le long des 800 colonnes. Ceci a réduit le nombre de segments à 9000.
- Après avoir refait le pointé, il a effectué la deuxième migration le long des colonnes.
- Finalement, il a effectué un troisième pointé le long des lignes à nouveau. Ce troisième pointé a donné une image finale composée de 8600 segments.

La figure II.6.3 nous montre les résultats après la première, deuxième et dernière étape.

**Notre procédure de migration** Nous n'avons ni pointé ni utilisé une méthode géométrique de migration. En effet, nous n'avons appliqué que des méthodes automatiques de traitement de traces. Cependant, nous avons légèrement modifié notre procédure de traitement. Au lieu d'utiliser une seule ligne de la coupe-somme des données Ecors 2.5-D et d'appliquer ensuite la technique de reconstruction de Kong et al. (1985), nous avons construit 11 lignes de cette coupe-somme (les lignes 128 à 138) par la procédure décrite ci-dessus (sans une deuxième application de la méthode de reconstruction). Ensuite, ces lignes ont été sommées (superposées) et la technique de reconstruction a été appliquée sur la coupe-somme résultante.

Nous avons choisi cette modification afin d'obtenir une coupe-somme ressemblant plus au résultat d'une analyse  $\tau-p-az$ . Même si cette modification n'est qu'une analyse  $\tau-p-az$  simplifiée, elle a bien tendance à homogénéiser les résultats. Le choix d'utiliser 11 lignes est basé sur le rayon de 200 m utilisé par Siliqi qui représente 9 lignes. Nous avons utilisé 2 lignes de plus, mais les résultats ne changent que légèrement si nous utilisons 2 lignes de moins. La coupe-somme



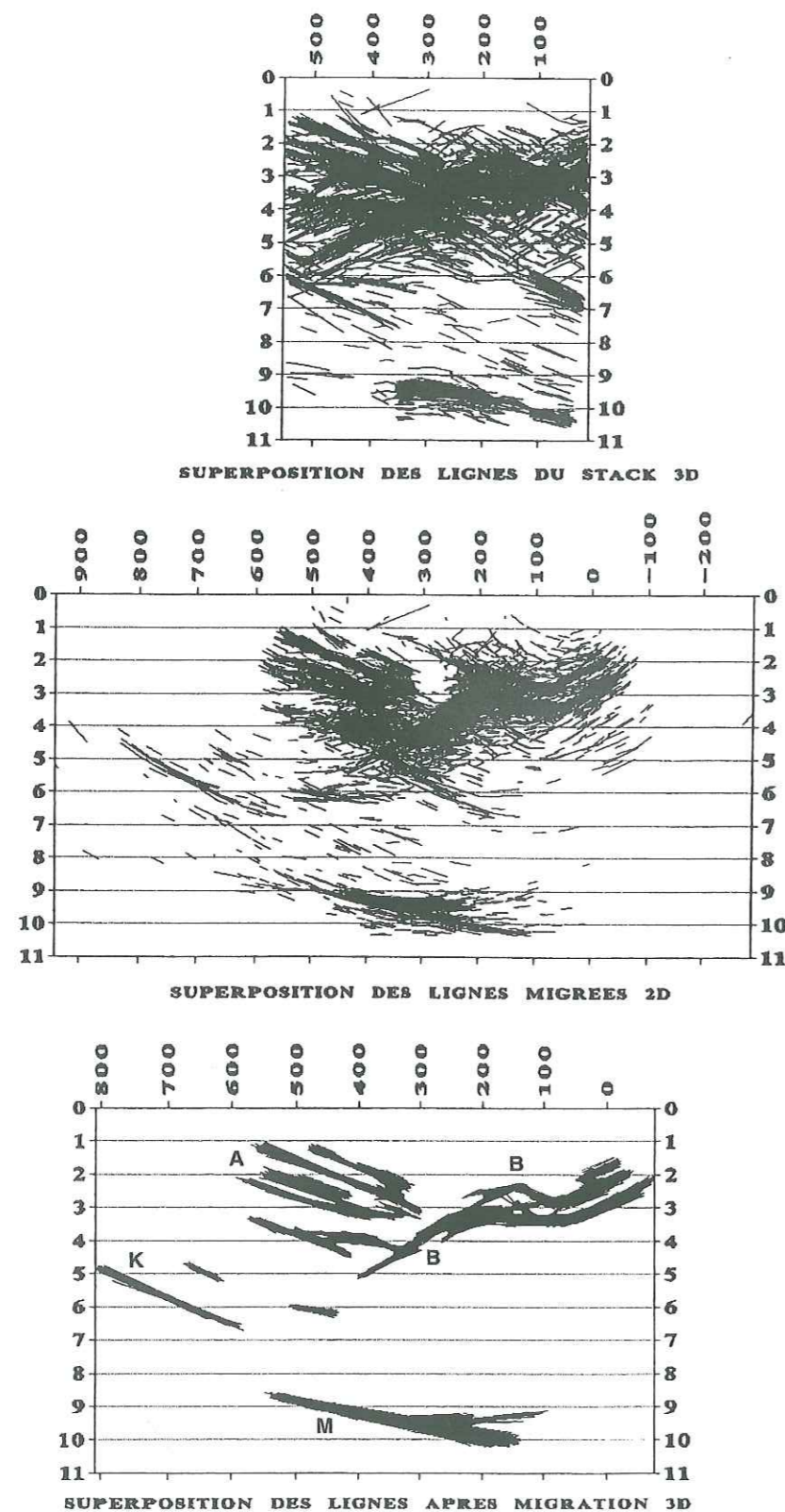


FIG. II.6.3 – Migration à 3 dimensions des segments pointés d'après Siliqi (1994). a) Superposition de tous les segments pointés le long des lignes. b) Superposition de ces segments après migration le long des lignes. c) Superposition des segments restants après migration le long des colonnes et le troisième pointé.

résultante est montrée sur la figure II.6.4.a. Nous retrouvons à peu près les mêmes structures géologiques dans cette coupe-somme que sur les figures II.6.1 et II.6.2. Seulement, elles sont plus continues et donc plus faciles à détecter.

Ensuite, après avoir fait une nouvelle coupe-somme, nous l'avons migrée à 2 dimensions avec la méthode de déphasage de Gazdag (1978). Cette méthode donne moins d'artefacts de migration que plusieurs autres méthodes, mais elle utilise beaucoup plus de temps de calcul (plusieurs heures). Cependant, cette technique produit beaucoup plus d'artefacts que la migration géométrique des segments de droites utilisée par Siliqi. La coupe-somme migrée est montrée sur la figure II.6.4.b.

Les données ont été migrées avec un autre modèle de vitesses que celui utilisé par Siliqi. Notre modèle est basé sur la moyenne des modèles de vitesses utilisées pour la sommation des traces. Le modèle commence avec une vitesse de 3.5 km/s à la surface. Puis, les vitesses augmentent linéairement jusqu'à 7 km/s à une profondeur de 5 km et ensuite jusqu'à 8 km/s à 8 km. Au-delà de cette profondeur, la vitesse reste constante.

### II.6.2.3 Comparaison des résultats après migration

Maintenant, nous tentons de comparer les résultats de la coupe-somme et de la migration (figures II.6.3 et II.6.4). Il est évident que Siliqi (1994) a obtenu de meilleurs résultats. On peut donner plusieurs explications pour cela. Premièrement, la sélection manuelle des segments de réflexions lui a permis d'avoir une influence directe sur les résultats obtenus. Deuxièmement, Siliqi a utilisé le jeu complet des données, tandis que nous n'avons utilisé que 7 % du volume total. Troisièmement, nous n'avons effectué qu'une migration à 2 dimensions et non à 3 dimensions. Finalement, la méthode géométrique de migration de segments n'introduit presque pas d'artefacts (comme par exemple des effets de bords).

Une comparaison directe des 2 coupes-sommées migrées est très difficile. Par conséquent, la façon la plus simple pour comparer les résultats obtenus est de comparer notre coupe-somme (figure II.6.4.a) avec la coupe-somme et la migration de Siliqi (figure II.6.3) et ensuite de regarder où les structures détectées se trouvent dans notre coupe-somme migrée (figure II.6.4.b).

**La coupe-somme** Les réflecteurs (A) inclinés vers le nord entre 1.5 et 3 s autour de la col. 400 sont faciles à détecter sur la figure II.6.4.a. Ils sont encore visible jusqu'à 3.6 s autour de la col. 225 et se prolongent peut-être même plus loin. Ces réflecteurs coupent 2 autres structures. La première est une structure de type anticlinal (B) à 2.5 s vers la col. 150. L'extrémité du flanc sud de cet anticlinal se trouve vers la col. 430 à 5.5 s (comme nous l'avons déjà suggéré sur la figure II.6.1.a). La deuxième (G) est un autre réflecteur à pendage sud situé autour de la col. 400 entre 3.2 et 4 s. Ce dernier réflecteur appartient soit à un autre anticlinal soit indique la présence d'une discontinuité structurale. Ensuite, nous voyons des réflecteurs entrecroisés (H) autour de la col. 40 à plusieurs profondeurs (1.2, 2.5, 3, 3.6 s).

Quant aux structures profondes, le Moho est clairement visible à 9.3 s entre les col. 350 et 225. Juste au-dessus, on trouve un autre réflecteur profond (N, vers 8.5 s). De plus, nous voyons un dernier réflecteur profond autour de la col. 80 à 10 s ( $M_3$ ). Ce réflecteur-là a aussi été détecté par Siliqi (figure II.6.3.b). Cependant, sur notre figure, ce réflecteur ne semble pas être lié au Moho, mais à son précurseur (N).

Par conséquent, nous avons identifié des structures identiques dans notre coupe-somme (figure II.6.4.a) et celle de Siliqi (figure II.6.3.a) à l'exception du précurseur du Moho, des réflecteurs entrecroisés autour de la col. 40 (H) et du réflecteur à pendage sud autour de la col. 400 à 3.5 s (G).



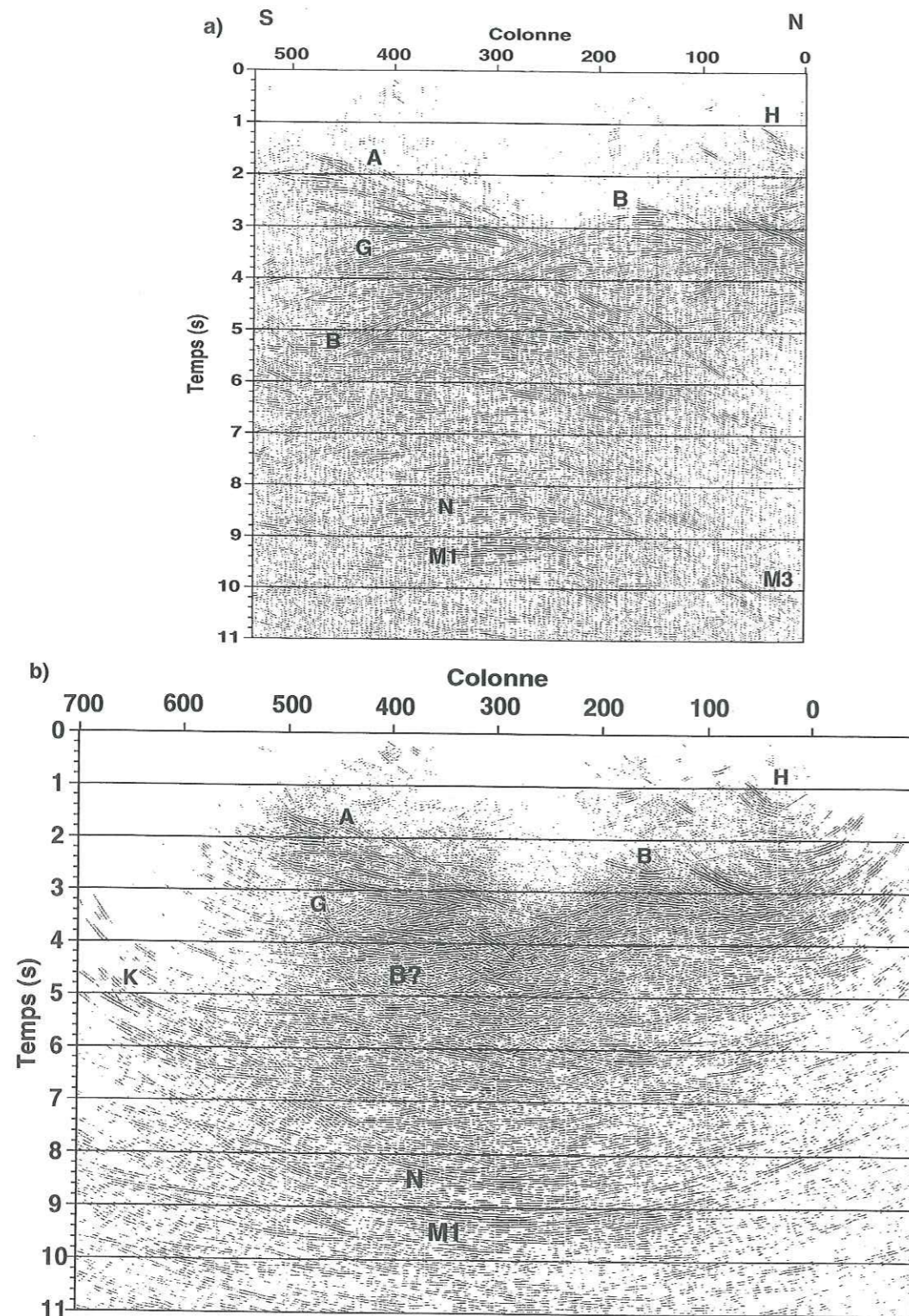


FIG. II.6.4 – Notre coupe-somme migrée. a) Coupe-somme finale résultant de la superposition des lignes 128 à 138 de la coupe-somme des données Ecors 2.5-D. b) Migration de la coupe-somme finale par la méthode de Gazdag (1978).

**La coupe-somme migrée** Après migration (figure II.6.4.b), les réflecteurs (A) inclinés vers le nord (maintenant entre les col. 500 et 350 et entre 1.5 et 3.2 s) restent bien visibles. De plus, les réflecteurs entrecroisés (H) autour de la col. 40 ont été transformés en un synclinal. Il s'agit du synclinal d'Orriule (figure II.2.3) qui est situé dans le Bassin de Mauléon entre les Chevauchements de Saint-Palais et Sainte-Suzanne. Les données Ecors 2.5-D confirment ainsi l'existence de ce synclinal, même si il n'a pas été vu par Siliqi (1994). Ce synclinal a aussi été mis en évidence par le profil Arzacq (Daignières et al., 1994).

La structure anticlinale (B, col. 150 à 2.5 s) est devenue peu visible après migration, mais peut-être peut-elle être interprétée comme la rampe (figure II.6.3.c) du Chevauchement de Sainte-Suzanne (un segment du Chevauchement Frontal Nord Pyrénéen). Cette structure a été interprétée ainsi par Siliqi (voir figure II.2.2.c).

Même après migration, il n'est pas clair si le réflecteur (G) à pendage sud (col. 400 à 3.5 s) appartient à un anticlinal ou une discontinuité structurale.

Quant aux structures profondes, la migration a complètement étalé les réflexions sur le Moho ( $M_1$ ) et son précurseur N (des 'sourires' sont apparus). Par conséquent, la migration n'a apporté rien de nouveau à nos connaissances concernant l'orientation et une bifurcation éventuelle du Moho. Siliqi (1994) a trouvé une telle bifurcation et a proposé de l'interpréter comme le contact entre les plaques ibérique et européenne (figure II.2.2.c).

Même si notre migration est caractérisée par de forts effets de bords, nous voyons apparaître un réflecteur intéressant à pendage nord entre les col. 700 et 600 entre 4.5 et 6 s (K) dans la coupe migrée (figure II.6.4.b). Un tel réflecteur a été interprété par Siliqi (1994) comme trace de la Faille Nord Pyrénéenne (figure II.2.2.c). Seulement, son réflecteur se trouve légèrement plus bas.

### II.6.3 Discussion

**La coupe-somme** La figure II.6.1 nous montre que nos méthodes de détection et de reconstruction donnent des résultats similaires à ceux obtenus par l'analyse  $\tau$ - $p$ - $az$  appliquée par Siliqi (1994). A première vue, ceci indique que l'extraction d'un signal spécifique, que nous avons fait, n'est pas nécessaire pour améliorer la qualité de la coupe-somme, puisqu'une telle analyse est aussi une technique de reconstruction. Il existe 2 raisons expliquant ce phénomène. Premièrement, la méthode de sommation de traces fait déjà appel à une procédure de sélection de signaux, puisqu'elle fait 'ressortir' tous les signaux ayant une courbure apparente spécifique. Deuxièmement, Siliqi n'a pas appliqué une technique de reconstruction à 2 dimensions, mais à 3 dimensions. Il faut toutefois remarquer que les résultats que nous obtenons avec une reconstruction à 3 dimensions très simple (figure II.6.4.b, reconstruction le long des lignes et sommation le long des colonnes) sont largement meilleurs que ceux de Siliqi (figure II.6.1.b). La coupe-somme résultante est beaucoup plus facile à interpréter. De plus, nous retrouvons beaucoup d'éléments du pointés de la figure II.6.3.a dans notre coupe-somme II.6.4.a, alors que nous n'avons utilisé que 7 % des données disponibles.

Nous pouvons remarquer le même phénomène sur les figures II.6.2, II.6.3.a et II.6.4.a concernant les structures profondes. Même si, initialement, il y a une nette différence de qualité entre les figures II.6.2.a et II.6.2.b, nous voyons que cette différence n'est plus si grande entre les figures II.6.3.a et II.6.4.a. Les plus grandes différences sont que nous ne trouvons pas une réflexion sur le Moho continue à pendage nord entre les colonnes 400 et 50, que nous voyons clairement un précurseur du Moho et que la réflexion profonde autour de la colonne 80 à 10 s semble, sur notre coupe-somme, être liée au précurseur et non au Moho.

A ce stade, il faut noter qu'il serait très simple d'adapter nos méthodes à 3 dimensions. En



effet, les techniques d'extraction traitent les données trace par trace. Par conséquent, aucune modification est nécessaire. De plus, la technique de reconstruction de Kong et al. (1985) fait appel à une analyse  $\tau$ - $p$ . L'adaptation à 3 dimensions se fait naturellement en appliquant une analyse  $\tau$ - $p$ - $az$ . Finalement, la semblance se calcule aussi facilement en 3 qu'en 2 dimensions.

Néanmoins, des tests faits par Siliqi nous montrent clairement que la semblance est une mesure très sensible à la présence de perturbations statiques des réflexions, puisque celles-ci ne se trouvent plus sur une ligne droite. En 3 dimensions, ce phénomène est encore plus sévère. Ceci a pour conséquence que la semblance n'est probablement plus une bonne mesure de cohérence en 3 dimensions. Nous envisageons 2 alternatives. Naturellement, nous pouvons appliquer la méthode de Bano (1989). Une autre alternative est d'appliquer la technique d'extraction directement dans le domaine  $\tau$ - $p$ - $az$  comme nous l'avons déjà suggéré dans la section II.3. Au cas où on choisirait la méthode de Bano (1989), nous recommandons de ne pas utiliser un seuil strict de 2 fois l'écart-type du bruit, mais un seuil plus graduel afin d'éviter des effets de bords.

**La migration** Il est évident que la méthode de migration géométrique appliquée par Siliqi (1994) a résulté en une meilleure reconstruction de l'image des structures géologiques que la technique de Gazdag (1978) que nous avons choisie. Nous pouvons citer plusieurs raisons expliquant ce fait. Premièrement, la méthode géométrique introduit moins d'effets de bords et d'autres artefacts de migration dans des données d'une faible qualité. Deuxièmement, la sélection manuelle des segments à migrer a permis à Siliqi d'avoir une influence directe sur les résultats obtenus. Troisièmement, il a effectué une migration des toutes les données disponibles à 3 dimensions, alors que nous n'avons migré que 7 % des données à 2 dimensions.

Néanmoins, nous retrouvons beaucoup de structures géologiques mises en évidence par Siliqi dans notre coupe-somme migrée. De plus, nous avons clairement montré la présence du synclinal d'Orriule et peut-être de la Faille Nord Pyrénéenne. Ceci nous montre le danger d'une sélection manuelle des pointés à migrer. En effet, ces 2 structures sont peu visibles dans notre coupe-somme. Par conséquent, elles sont difficiles à pointer ce qui augmente le risque de les négliger.

Cependant, une méthode de migration géométrique est probablement la plus appropriée pour des données d'une faible qualité (Warner, 1987). Par conséquent, il faut chercher une méthode automatique de pointé afin d'éviter tout travail fastidieux.

Parmi les méthodes de reconstruction introduites dans la section II.3, il y a 2 qui peuvent accomplir une telle tâche. La méthode de squelettisation (Le et Nyland, 1990; Lu et Cheng, 1990) essaie de corréler les formes d'ondes des traces voisines pour produire une esquisse structurale de la coupe considérée. Li et al. (1997) nous montrent comment cette technique peut être appliquée à des profils de sismique réflexion profonde. Malheureusement, nous avons déjà remarqué que cette technique exige une bonne qualité de données.

Ensuite, il y a la méthode statistique de Hansen et al. (1988). Elle analyse localement les données sous des angles différents en cherchant tous les segments ayant plus d'amplitudes positives qu'un seuil fixé. Puis, elle extrait toutes les parties des données où des segments ont été détectés (section II.3.1). En n'utilisant que les segments détectés, cette méthode peut aussi extraire des segments de pointés et, ainsi, être utilisée au lieu d'une sélection manuelle.

Alsdorf (1997) recommande aussi d'utiliser une méthode de migration géométrique. Il se sert d'une méthode ressemblant à la squelettisation afin de pointer les segments à migrer.

**Chevauchement de Saint-Palais** Un phénomène curieux est le fait que ni les données Ecors 2.5-D ni le profil Arzacq (Daignières et al., 1994) n'ont mis en évidence le Chevauchement de Saint-Palais. Pourtant, il s'agit d'une structure majeure dans le Bassin de Mauléon. Souvent, la présence d'une faille n'est indiquée dans des données sismiques que par de réflecteurs

tronqués; la faible qualité des données Ecors 2.5-D ne permet probablement pas de détecter la faille de cette façon. Par contre, la rampe du Chevauchement de Sainte-Suzanne est bien visible dans ces données (figures II.6.3.c et II.2.2.c). Il est donc difficile de comprendre pourquoi aucun des deux profils n'a trouvé d'indications de l'existence du Chevauchement de Saint-Palais.

#### II.6.4 Conclusions

Nos méthodes de détection et de reconstruction nous ont permis d'améliorer sensiblement la qualité de données Ecors 2.5-D et ainsi de faciliter l'interprétation du profil. Une comparaison de nos résultats avec ceux obtenus par Siliqi (1994) nous a montré que

- Un traitement de reconstruction à 3 dimensions des données est indispensable afin d'améliorer la qualité des données.
- L'interprétation des données de sismique réflexion profonde migrées est la plus facile si les données sont migrées par des méthodes simples de migration géométrique des segments pointés. Par conséquent, il faut chercher des méthodes automatiques pour pointer les segments à migrer.
- Nos méthodes automatiques de détection et de reconstruction ont extrait et mis en évidence plusieurs structures géologiques trouvées par Siliqi en pointant manuellement des réflecteurs. De plus, nous avons détecté le synclinal d'Orriule. Pourtant, nous n'avons traité que 7 % du volume total des données Ecors 2.5-D et ceci sans aucune intervention manuelle hors le pointé de la réflexion sur le Moho dans un seul point de tir commun.



## II.7 Conclusions

Nous avons montré qu'il est possible d'améliorer la qualité des données de sismique réflexion profonde à l'aide de techniques automatiques de détection et de reconstruction de signaux. Cette qualité améliorée facilitera ensuite l'interprétation des données.

**Détection** Nous avons testé deux méthodes de détection, la méthode de Classification par Vraisemblance Maximale et les réseaux de neurones. Ces deux techniques ont des avantages et inconvénients opposés.

- La Classification par Vraisemblance Maximale est facile à mettre en œuvre. Par contre, elle donne de mauvais résultats si l'on ne connaît pas la distribution statistique sous-jacente des différentes catégories de signaux. Ceci pose surtout des problèmes dans le cas où la vraie distribution sous-jacente est multi-modale ou asymétrique, puisqu'elle doit être estimée par des approximations successives.
- Les réseaux de neurones sont capables d'estimer cette distribution sous-jacente de façon automatique, mais souvent il est plus difficile de les mettre en œuvre.

**Attributs** Une bonne détection de signaux exige de connaître les attributs permettant au mieux de caractériser les signaux et de distinguer entre les différentes catégories. Deux transformées souvent utilisées pour analyser des signaux sismiques sont la Transformée de Fourier à Court Terme et la Transformée Continue en Ondelette. De nouveau, ces deux transformées ont des avantages et inconvénients opposés.

- La Transformée de Fourier à Court Terme est très rapide, mais elle nécessite une spécification *a priori* de la longueur de la fenêtre d'analyse, puisque cette longueur reste fixée.
- La Transformée Continue en Ondelette est moins rapide, mais elle analyse les données sur des échelles différentes. Par conséquent, elle est capable de déterminer la ou les longueurs caractéristiques des données de façon automatique.

Des tests nous ont montré que la Transformée de Fourier à Court Terme est optimale si nous ne voulons extraire qu'un seul type de signal. Autrement, une Transformée Continue en Ondelette est plus appropriée.

**Reconstruction** L'inconvénient principal des techniques décrites de détection est qu'elles ne peuvent fournir aucune estimation de toute énergie 'manquante'. Ceci requiert une technique de reconstruction. Dans cette étude, nous avons appliqué la technique de Kong et al. (1985) qui fait appel à des sommations obliques pondérées par la semblance. Cette méthode donne de bons résultats après l'extraction d'un signal particulier. Cependant, il serait très utile d'appliquer une extension de cette technique à 3 dimensions.

**Ecors 2.5-D** Afin d'illustrer au mieux les capacités des techniques de détection et de reconstruction, nous les avons appliquées sur les données de l'expérience de sismique réflexion profonde Ecors 2.5-D dans les Pyrénées. Cette chaîne de montagnes est une région à tectonique très complexe. De plus, les données Ecors 2.5-D ont une faible qualité.

Après migration, nos méthodes donnent des résultats similaires à ceux obtenus par Siliqi (1994), mais d'une qualité plus faible. Ceci est dû à l'algorithme de migration choisi et au fait que nous n'avons appliqué rien que des méthodes automatiques sur 7 % des données, tandis que



Siliqi (1994) a pointé manuellement toutes les bonnes réflexions dans le volume total de données. Par contre, avant migration, nos méthodes obtiennent des coupes-sommes sismiques d'une qualité supérieure pour la croûte supérieure et légèrement inférieure pour la croûte inférieure, puisque Siliqi (1994) a pointé manuellement toutes les traces où le Moho est visible. Nous avons montré aussi que le danger d'un pointé manuel sur des coupes-sommes est de négliger des structures qui n'apparaissent que clairement après migration.

**Migration** La méthode simple de migration de segments pointés de Siliqi (1994) s'est montrée meilleure que les méthodes courantes migrant les amplitudes des traces. L'inconvénient principal est qu'elle nécessite un pointé manuel de tous les segments à migrer. Par conséquent, il faut chercher des méthodes automatiques afin de pointer toutes les réflexions dans des données d'une faible qualité.

Localisation d'onde



## Chapitre III

# Localisation d'onde

### Résumé

Le phénomène de la diffusion multiple dans des milieux très hétérogènes est très compliqué. Même si une approche statistique en simplifie la rigueur mathématique de façon considérable, elle ne garantit pas que les quantités prédites et observées coïncident. La théorie de la localisation d'onde est une théorie sur la diffusion multiple qui utilise des quantités auto-moyennantes résolvant ainsi ce problème. Donc, cette théorie nous permet d'étudier l'influence de la présence d'hétérogénéités sur l'atténuation apparente et la dispersion d'une impulsion traversant un milieu aléatoire et de comparer des prédictions théoriques avec les résultats de simulations numériques.

Cette théorie s'applique à des milieux acoustiques et stratifiés où la densité est constante sur presque toute la gamme de fréquences, lorsque l'angle d'incidence reste inférieur à  $30^\circ$  et l'écart-type des fluctuations de l'incompressibilité ne dépasse pas 30 %. De plus, elle montre que, à l'exception des milieux fractals, l'atténuation apparente est maximale dans le domaine de la diffusion de type 'Mie'. Pour des milieux fractals, elle ne dépend presque pas de la fréquence. Finalement, les vitesses de groupe et de phase sont en concordance avec la théorie d'homogénéisation et les relations de Kramers-Krönig. Leur comportement exacte ne dépend que de la rugosité du milieu pour les fréquences intermédiaires.

Ensuite, nous appliquons cette théorie à l'inférence statistique. Nous résolvons le problème inverse en mesurant la dépendance fréquentielle des rapports spectraux de plusieurs réflexions majeures. Des tests synthétiques montrent que les rapports spectraux mesurés sont contaminés par des réflexions multiples tardives, qui ne sont pas prévues par la théorie de la localisation d'onde, puisqu'il s'agit d'une théorie en transmission. Elles tendent à augmenter l'énergie totale et ainsi à sous-estimer la magnitude typique des hétérogénéités et l'atténuation intrinsèque. Malheureusement, leur influence exacte est difficile à prévoir.

Finalement, nous avons utilisé la théorie de la localisation d'onde pour estimer la magnitude  $\sigma_c$  et la dimension  $a$  typiques des fluctuations de vitesses et le facteur de qualité de l'atténuation intrinsèque  $Q_{in}$  dans la croûte sous le Bassin de Mauléon dans les Pyrénées Occidentales. La croûte supérieure y est caractérisée par  $a \sim 15$  m,  $\sigma_c \leq 7.5$  % et  $Q_{in} \geq 350$  et la croûte inférieure par  $a \sim 40$  à  $50$  m,  $\sigma_c \leq 5.5$  % et  $Q_{in} \geq 1600$ . Ceci nous montre que la croûte supérieure est plus hétérogène et plus atténuante que la croûte inférieure et que les dimensions typiques des hétérogénéités y sont plus petites.



### III.1 Introduction

Le deuxième exemple que nous considérons afin d'illustrer l'utilité d'une approche statistique pour des problèmes géophysiques est la propagation d'ondes dans des milieux aléatoires et en particulier l'influence de la présence des hétérogénéités présentes sur l'atténuation apparente et la dispersion d'une onde plane traversant un tel milieu. Ce problème est caractérisé par une très grande complexité mathématique, surtout si nous utilisons une approche déterministe. Car dans ce cas, il faut décrire exactement chaque hétérogénéité (localisation, forme, magnitude, etc.), ce qui aboutit à un nombre de paramètres énorme. Si, par contre, nous faisons appel à une approche statistique, les milieux seront complètement décrits par des fonctions d'autocorrélation spatiale. Ces fonctions ne dépendent que de trois paramètres: la longueur et la magnitude caractéristiques des hétérogénéités et le type exact de fonction. Ce dernier paramètre décrit la rugosité des hétérogénéités.

Cette complexité mathématique nous empêche d'améliorer nos connaissances sur le phénomène de la propagation d'ondes dans un milieu complexe et surtout sur le processus de la diffusion forte. Donc, en réduisant le nombre de paramètres, nous simplifions les mathématiques considérablement. Par conséquent, il nous sera beaucoup plus facile de comprendre l'influence de la présence d'hétérogénéités sur l'atténuation apparente et la dispersion d'une impulsion.

Dans la première partie de ce chapitre (section III.2), nous décrivons la théorie de la localisation d'onde qui a été utilisée afin de mieux comprendre certains phénomènes physiques liés à la diffusion multiple causée par les hétérogénéités présentes. Nous examinons surtout l'atténuation apparente et la dispersion d'une onde plane acoustique résultant des diffractions multiples pendant qu'elle traverse un milieu aléatoire stratifié. De plus, nous décrivons en détails les difficultés posées par les diffractions multiples et plusieurs autres méthodes proposées dans la littérature.

Dans la deuxième partie (section III.3), nous examinons l'aptitude de la théorie de la localisation d'onde à l'inférence statistique, c'est-à-dire à la détermination des propriétés statistiques du milieu. Nous essayons donc de calculer par inversion les paramètres statistiques décrivant un milieu hétérogène en mesurant l'atténuation d'une réflexion profonde en fonction de la fréquence. Nous nous concentrons surtout sur les divers problèmes rendant une telle inversion très compliquée. Nous voulons utiliser des données en réflexion et non en transmission, puisqu'elles sont plus faciles à obtenir et puisqu'elles nous permettent d'étudier la Terre jusqu'à une plus grande profondeur. De nouveau, nous appliquons la méthode développée sur des données de l'expérience de sismique réflexion profonde Ecors 2.5-D.



## III.2 Propagation d'ondes en milieu aléatoire

### Résumé

Le phénomène de la diffusion multiple dans des milieux très hétérogènes est très compliqué. Même si une approche statistique en simplifie la rigueur mathématique de façon considérable, elle ne garantit pas que les quantités prédites et observées coïncident. Afin de résoudre ce problème, il faut faire appel à des quantités auto-moyennantes.

La théorie de la localisation d'onde est une théorie sur la diffusion multiple qui utilise de telles quantités auto-moyennantes. Cette théorie nous permet donc d'étudier l'influence de la présence d'hétérogénéités sur l'atténuation apparente et la dispersion d'une impulsion traversant un milieu aléatoire et de comparer des prédictions théoriques avec les résultats de simulations numériques. Nous calculons la longueur de localisation (profondeur de pénétration), la réciproque du facteur de qualité et les vitesses de groupe et de phase pour plusieurs milieux aléatoires décrits par des fonctions d'autocorrélation différentes. Les calculs sont limités à des milieux acoustiques et stratifiés où la densité est constante. Cependant, la rugosité des milieux étudiés varie de très lisse à fractale. De plus, incidence non-verticale est permise. Ensuite, nous comparons les prédictions théoriques à des résultats de simulations numériques.

Cette étude nous a permis de tirer les conclusions suivantes. 1) Les prédictions théoriques et les simulations numériques coïncident pour presque toutes les fréquences si l'angle d'incidence reste inférieur à  $30^\circ$  et si l'écart-type des fluctuations de l'incompressibilité ne dépasse pas 30 %. 2) A l'exception des milieux fractals, l'atténuation apparente est maximale dans le domaine de la diffusion de type 'Mie' où l'échelle typique des hétérogénéités et la longueur d'onde sont du même ordre de grandeur. Par contre, dans des milieux fractals, aucune échelle typique des hétérogénéités n'existe et la réciproque du facteur de qualité devient constante sur une large gamme de fréquences. 3) Les vitesses de groupe et de phase sont en concordance avec la théorie d'homogénéisation et les relations de Kramers-Krönig. Donc, elles convergent vers les vitesses effective et géométrique dans les limites des basses et hautes fréquences respectivement. Cependant, pour des fréquences intermédiaires, leur comportement dépend fortement du type de milieu et il peut exister plusieurs extrema et des phases d'Airy.



### Acoustic wave propagation in 1-D random media: The wave localization approach<sup>1</sup>

Mirko van der Baan

*Laboratoire de Géophysique Interne et Tectonophysique,  
Université Joseph Fourier and CNRS, BP 53, 38041 Grenoble Cedex 9, France.*

#### Abstract

Multiple wave scattering in strongly heterogeneous media is a very complicated phenomenon. Although a statistical approach may yield a considerable simplification of mathematics, no guarantee exists that the theoretically predicted and the observed quantities will coincide. The solution of this problem is to use self-averaging quantities only.

A multiple scattering theory which makes use of such self-averaging quantities is the so-called wave localisation theory. Therefore, this theory allows to study both numerically and theoretically the influence of the presence of heterogeneities on the frequency dependent dispersion and apparent attenuation of a pulse traversing a random medium. I will calculate the localisation length (penetration depth), the inverse quality factor and both the group and phase velocities for several chaotic media described by different autocorrelation functions. Calculations are limited to 1-D acoustic media with constant density. However, studied media range from very smooth to fractal like and incidence is not limited to be vertical. I then compare the theoretical results with estimates of the same quantities obtained from numerical simulations.

It may be concluded that: 1) Theoretical predictions and numerical simulations agree in nearly the whole frequency domain for angles of incidence  $\leq 30^\circ$  and relative standard deviations of the fluctuations of the incompressibility  $\leq 30\%$ . 2) An inspection of the inverse quality factor confirms that the apparent attenuation is strongest in the domain of Mie scattering except for fractal-like media. In such media no particular ratio of the wavelength to the typical scale-length of heterogeneities is preferred as no such typical scale-length exists. Hence, the inverse quality factor is constant over a large frequency band. 3) Obtained group and phase velocities agree with the effective medium theory and the Kramers-Krönig-relations. That is, both converge to the effective medium velocity and the geometric velocity in respectively the low and high frequency domain. However, for intermediate frequencies, the exact behaviour strongly depends on the type of medium. Differences concern mainly the number of extrema and Airy phases.

#### III.2.1 Introduction

Much has already been written about wave scattering and in particular the influence of small- and large-scale heterogeneities on the dispersion and the apparent attenuation of a wavefront traversing a heterogeneous medium. Sato and Fehler (1998), for instance, give a good overview of theoretical and practical developments of the last 20 years, albeit in a seismological context only.

Although the wave scattering problem is completely described by linear partial differential equations, non-linearities are introduced as the solutions depend in a non-linear way on the coefficients of the differential equations. These coefficients are naturally random in random media (Frisch, 1968). Hence, for heterogeneous media, exact solutions can only be established in some rare and extremely simplified cases. Yet, most of these solutions are formulated in terms of infinite series, which do not always converge easily. Further complications arise due to the

<sup>1</sup>Article soumis à Geophysical Journal International en 1999.



enormous number of variables which has to be dealt with, especially if a deterministic approach is chosen for. Moreover, these mathematical complications make it very hard to improve our physical understanding of wave scattering in all its aspects.

One possible method to simplify the problem is not to opt for a deterministic, but a statistical approach. Whereas the first approach requires a full description of each single heterogeneity (magnitude, position, shape, orientation, etc.), in the latter approach media are completely described by their spatial autocorrelation functions, as these functions determine the characteristic magnitudes and scale-lengths of the present heterogeneities and their respective roughness or smoothness. Such a statistical description is of particular interest if both the principal wavelength and the characteristic scale-length of the heterogeneities are largely inferior to the dimensions of the random medium (or the considered path length). A sound mathematical review of the problem of wave propagation in random media can be found in Frisch (1968), whereas Herraiz and Espinosa (1987) and Sato and Fehler (1998) deal with more practical issues.

Thus, the use of a statistical description of the random medium results in simplified mathematics and can, therefore, help to improve our understanding of the influence of heterogeneities on the dispersion and apparent attenuation of a wavefront. Unfortunately, however, no guarantee exists that observed and theoretically predicted mean values coincide. That is, fluctuations may dominate in single realizations of the medium. Hence, the true mean can only be estimated if several configurations and thus independent measurements exist. This drawback makes statistical theories very hard to use for inversion purposes, especially if only a single realization exists like e.g. in the problem of wave propagation in the real Earth.

The solution to this problem is to use self-averaging quantities only. Such quantities assume their mean values in each single realization with unit probability provided that the wave passes through a sufficiently large medium, thereby endowing the mean values with the physical meaning of real observables (Gredeskul and Freilikher, 1990; Shapiro and Zien, 1993).

In this article, I will use the theory of wave localization which deals with such self-averaging quantities only. In addition, this theory is particularly suited for strong scattering problems and works, therefore, well for large propagation distances (exceeding largely the mean free path) in contrast to other statistical approaches which will be described below.

The origin of the wave localization theory stems from the quantum theory of disordered solids and in particular from the seminal article of Anderson (1958). In this discipline, it is now of main stream interest (see e.g. the review articles of Ramakrishnan (1987), Souillard (1987) and Van Tiggelen (1999)). However, the application of results developed in quantum mechanics to the theory of wave propagation in random media is of a more recent origin. Interest in predictions of localization theory for multiple scattering of acoustic waves started only 20 to 25 years after the 1958 paper of Anderson.

Wave localization implies that a wave has an exponential fall-off for large distances from its maximum, causing the envelope of a pulse to decrease exponentially for large propagation distances in 1-D random media. Or to put it differently, the transmission coefficient  $T$  behaves like  $|T| = \lim_{L \rightarrow \infty} \exp(-\gamma/L)$  (for almost any frequency and almost any realization of the medium) with  $\gamma$  the so-called Lyapunov exponent.

The mathematical proof of such an exponential decay comes from the discipline of random matrix theory. This discipline tries to predict asymptotic values for certain systems described by (transfer) matrix multiplications and can, therefore, be applied to wave propagation problems in 1-D media. Hence, random matrix theory can establish theoretical predictions for certain problems described by linear partial differential equations with random coefficients like some multiple scattering problems.

The principal theorems needed to prove the above statement are those of Fürstenberg (1963)

and Oseledec (1968) for stationary and ergodic random media. Sometimes Virster's (1979) theorem is also invoked which yields a generalization of the first two theorems. Fürstenberg's theorem states that the Lyapunov exponent  $\gamma$  exists and is positive for almost any frequency and is non-random for almost any realization of the random medium. Oseledec's theorem, on the other hand, states that  $\gamma$  exists pairwise (Souillard, 1987). That is, for 2x2 matrices, there exists both a positive and negative  $\gamma$  with equal absolute magnitudes. Hence, mathematically both an exponentially increasing and decreasing wave exists with unit probability. Although the physically non-realistic exponentially growing wave dominates for random boundary conditions, posing the right boundary conditions (in particular finite radiation conditions at  $\pm$  infinity) will eliminate it. Proofs can also be found in Lifshits et al. (1988). Delyon et al. (1983) give a simplified proof holding for 2x2 matrices only.

The Lyapunov exponent  $\gamma$  satisfies the above-mentioned statements almost surely since more general results cannot be obtained. Nevertheless, it is a self-averaging quantity, i.e. it reaches its predicted value with unit probability for large distances as only a finite, i.e. countable, number of exceptions will exist. One exception is related to the fact that for  $f \rightarrow 0$  no wave scattering occurs, i.e. in the long wavelength limit the medium is effectively homogeneous and  $\gamma$  becomes zero as will be shown later. Other exceptions are mainly related to the occurrence of resonances e.g. due to periodic layering (Souillard, 1987). Notice, however, that the above theorems do imply that mean values may deviate from the theoretically predicted ones if they are dominated by atypical realizations. However, the probability for such an atypical realization to occur is zero.

Naturally, other theories describing the apparent decay of amplitude and phase distortions due to multiple scattering do exist. However, none of these use self-averaging quantities and most do not yield simple analytical expressions. The best known theories are based on the Born and Rytov approximations (Chernov, 1960; Tatarskii, 1961) and the mean-wave formalism (Karal and Keller, 1964). These theories work well for low frequencies, but display an unrealistic behavior in the high frequency domain (Herraiz and Espinosa, 1987; Sato and Fehler, 1998). Some more recent ameliorations (Sato, 1982; Wu, 1982) do produce more realistic results for high frequencies. Unfortunately, however, these remain single scattering approximations. Hence, they are only valid for a weak scattering medium or more appropriately put if the propagation distance is small compared to the mean free path, which controls the energy transferred from the primary (incident) to the scattered waves throughout the traveled path (Herraiz and Espinosa, 1987). Multiple scattering theories are needed if this condition is violated.

A method which is known to produce good results in a strongly scattering medium for large propagation distances is the diffusion approach, as a strong multiple scattering process can be described by a diffusion process (Morse and Feshbach, 1953). However, energy is spreading in front of the wavefront, thereby violating causality (Sato and Fehler, 1998). Thus, it can only be applied to coda analysis. Moreover, Margerin et al. (1998) showed that care should be taken in applying the diffusion equation in layered media.

Another approach which is gaining an increasing popularity (Wu, 1985; Sato and Fehler, 1998; Margerin et al., 1998) is the radiative transfer theory of Chandrasekhar (1960). This is a phenomenological theory which deals only with energy densities and intensities. Its basic assumption is that wave scattering results in random fluctuations of phase. Hence, no interference effects have to be accounted for, thereby considerably simplifying mathematics as we only have to treat conservation of energy. Naturally, the theory can not handle phenomena in which wave interference predominates as e.g. the coherent backscattering effect. Nevertheless, in media with discrete scatterers, the theory can accurately handle any order of scattering. Unfortunately, however, it does not produce simple analytical expressions for random media described



by spatial autocorrelation functions.

In this article, equations will be derived for the dispersion and apparent attenuation of acoustic wave propagation in 1-D random media using first principles only and the above described exponential decay of the transmission coefficient due to wave localization. First, the mathematical problem of wave propagation in chaotic media will be formulated and some reasons will be given to explain the different behavior of quantum mechanical and classical wave localization. Then, it will be shown that a medium becomes effectively homogeneous in the long wavelength limit. Next, to obtain and analyze the frequency dependent apparent attenuation and dispersion due to wave scattering, the second order perturbation expansion of Shapiro and Zien (1993) will be used. Finally, numerical simulations will be performed to determine to what extent theoretical predictions and practice agree. Data is analyzed using the wavelet transform which produced considerably more accurate results using single realizations only than those obtained in other studies and numerical tests.

### III.2.2 Theory

#### III.2.2.1 Posing the problem: wave propagation through random media

In this paper wave propagation in random media will be considered. The chaotic parts of the media will be described by means of statistical fluctuations of the background parameters which are characterized by a standard deviation and a particular autocorrelation function. The standard deviation determines the characteristic magnitude and the autocorrelation function the typical scale-length of the heterogeneities and the amount of smoothness or on the contrary roughness present.

As a first physical insight is most easily obtained from a simplified approach, several assumptions will be made to simplify mathematics. Nevertheless, large parts of the full mathematical 1-D problem are already known and will be referred to in due time. In addition, the influence of dropping these simplifications will be briefly alluded to.

First of all, only acoustic wave propagation in perfect 1-D media without energy absorption (dissipation) will be considered. Although the analysis is limited to 1-D media, non-vertical incidence is allowed for. Moreover, only acoustic waves will be dealt with to prevent complications due to the co-existence of both P and S waves and their coupling. Second, fluctuations are assumed to be static and stationary. However, relative standard deviations of up to 30 % can be handled rather well by the theory. Such standard deviations are largely sufficient to account for most 1-D models. Finally, the medium is assumed to be ergodic.

Scales are all determining in wave scattering problems. The two most important scales are  $ka$  and  $kL$  with  $a$  the typical scale-length of the heterogeneities,  $L$  the thickness of the random part of the medium and  $k$  the wavenumber ( $k = 2\pi/\lambda$  with  $\lambda$  the wavelength of the probing pulse). Wave localization occurs only if  $kL \gg 1$  and  $a \ll L$  which places some constraints on  $ka$ .

A ratio of scales often used by R. Burrige and G. Papanicolaou and co-workers (see e.g. Burrige *et al.* (1994)) is  $a : \lambda : L = \varepsilon^2 : \varepsilon : 1$  with  $\varepsilon$  a small parameter tending to zero. Hence, the wavelength exceeds the typical scale-length of the heterogeneities, but remains smaller than the thickness of the random medium. It is, therefore, a low frequency approach. I will adopt this ratio for the moment as it leads to an increased physical insight in the problem. It will, however, be dropped later on and  $ka$  is allowed to vary up to  $a \approx \lambda$ . The advantage of this predefined ratio is that it allows us to define 3 typical scale-lengths, namely a microscopic scale associated with random fluctuations, a macroscopic scale associated with propagation distances and finally the intermediate scale of the pulse width which will be called the mesoscopic scale.

Another advantage of this predefined ratio is that a very large class of fluctuating density  $\rho(z)$  and compressibility  $\kappa(z)$  profiles are allowed for. These profiles can be described by

$$\begin{cases} \rho(z) = \rho_0(z)[1 + \sigma_\rho(z, z/\varepsilon^2)] \\ \kappa^{-1}(z) = \kappa_0^{-1}(z)[1 + \sigma_\kappa(z, z/\varepsilon^2)], \end{cases} \quad (\text{III.2.1})$$

where  $z$  is positive with depth. Thus, the fluctuations are described by an average profile varying only on the macroscopic scale  $z$  and relative perturbations  $\sigma(z, z/\varepsilon^2)$ , which may vary both on the macroscopic and the microscopic scale, i.e. respectively on the slow variable  $z$  and the fast variable  $z/\varepsilon^2$ . Furthermore, both  $\sigma_\rho$  and  $\sigma_\kappa$  have an expectancy, i.e. average, of zero. For instance, profiles are allowed which have relative perturbations diminishing with depth superimposed on a linear gradient. The pulse width, i.e. the mesoscopic scale, will be described by  $z/\varepsilon$ . Hence, the constant  $\varepsilon$  serves as an amplification factor representing the different scales.

This ratio is employed to show that a medium becomes effectively homogeneous in the long wavelength limit. However, results can only be obtained in the low frequency domain. Thus, thereafter, a more restricted class of profiles with only stationary fluctuations will be considered, permitting however a much broader frequency range.

Another means to simplify mathematics is to deal with two different formulations of the acoustic wave equation, namely either as two coupled first order differential equations or as a single second order differential equation. To show similarity, both will be derived from the equation of motion, i.e.

$$\rho \partial_{tt} u_i = \partial_j \sigma_{ij} + f_i, \quad (\text{III.2.2})$$

which balances forces and tractions on a small volume element. The displacements  $u_i$  in directions  $x_i$ , the stress tensor  $\sigma_{ij}$  and the body force component  $f_i$  (acting per unit volume) are all defined as perturbations with respect to the static situation (Nolet, 1987). Moreover, the Einstein convention is used implying summation over repeated indices, save  $\partial_{tt}$  which stands for  $\partial^2/\partial t^2$ .

Since only acoustic waves are considered, the stress tensor can be replaced by

$$\sigma_{ij} = -p \delta_{ij}. \quad (\text{III.2.3})$$

Furthermore, pressure  $p$  is directly proportional to the dilatation  $\partial_i u_i$ , namely

$$-p = \kappa \partial_i u_i. \quad (\text{III.2.4})$$

Equations (III.2.2) to (III.2.4) suffice to derive the two desired representations of the wave equation. In both cases, equation (III.2.3) can be used to eliminate the stress tensor in equation (III.2.2).

**Pressure – particle displacement** The first formulation can then be obtained if we note that the divergence of  $u_i$  can be expressed in terms of pressure, equation (III.2.4). Therefore, dividing equation (III.2.2) by the density, applying the divergence operator and inserting equation (III.2.4) results in

$$\kappa^{-1} \partial_{tt} p = \partial_i (\rho^{-1} \partial_i p) - \partial_i (\rho^{-1} f_i) \quad (\text{III.2.5})$$

for stationary  $\kappa$ . In the absence of body forces, i.e. excitation forces, equation (III.2.5) reduces to the well known wave equation for a constant density profile, i.e.

$$\partial_{tt} p = c^2 \partial_{ii} p, \quad (\text{III.2.6})$$



where the velocity  $c$  is defined as  $c = (\kappa/\rho)^{1/2}$  and  $\partial_{ii}$  represents  $\sum_i \partial^2/\partial x_i^2$ .

Since computers are better in solving first order differential equations, equation (III.2.5) is also often formulated as a system of coupled first order differential equations in frequency - wave number domain. Namely, if a time harmonic dependence of the form

$$p(x, y, z, t) = \left(\frac{1}{2\pi}\right)^3 \int \hat{p}(k_x, k_y, z, \omega) e^{i(k_x x + k_y y - \omega t)} dk_x dk_y d\omega \quad (\text{III.2.7})$$

is assumed for the pressure in stratified media described by system (III.2.1) with the  $y$ -axis oriented perpendicular to the direction of wave propagation then equation (III.2.5) transforms to

$$\left[\frac{\omega^2}{\kappa} - \frac{k_x^2}{\rho}\right] \hat{p} + \partial_z (\rho^{-1} \partial_z \hat{p}) = 0 \quad (\text{III.2.8})$$

after a Fourier Transformation of the equation, since  $k_y = 0$  in this case. By defining the displacement  $\hat{w}$  in frequency domain as  $\rho^{-1} \omega^{-2} \partial_z \hat{p}$ , equation (III.2.8) can also be written as

$$\frac{d}{dz} \begin{pmatrix} \hat{p} \\ \hat{w} \end{pmatrix} = \begin{pmatrix} 0 & \rho\omega^2 \\ \frac{k_x^2}{\rho\omega^2} - \frac{1}{\kappa} & 0 \end{pmatrix} \begin{pmatrix} \hat{p} \\ \hat{w} \end{pmatrix}. \quad (\text{III.2.9})$$

**Pressure - particle velocity** Alternatively, wave propagation can also be described using two coupled differential equations in time domain. Namely, after defining the particle velocity  $\partial_t u_i$  as  $\dot{u}_i$ , insertion of equation (III.2.3) in (III.2.2) and application of time differentiation to equation (III.2.4) yields

$$\rho \partial_t \dot{u}_i + \partial_i p = f_i \quad (\text{III.2.10})$$

and

$$\kappa^{-1} \partial_t p + \partial_i \dot{u}_i = 0, \quad (\text{III.2.11})$$

which are also known as the equations of momentum and mass conservation.

### III.2.2.2 Quantum mechanical and classical waves

Localization is a phenomenon characteristic for waves in random media. It is well known that, for (infinite) systems described by one or two spatial dimensions, both quantum mechanical and classical waves are localized if an infinitesimal small amount of randomness is present over sufficiently large propagation distances. This phenomenon was first predicted by the scaling theory of localization (Abrahams et al., 1979). Although the basic physics of localization are common for both types of waves, specific behaviors are expected to differ due to different dispersion relations and boundary conditions. Examples of different behavior can be found in wave-propagation characteristics and in the frequency dependence of the localization length (Sheng et al., 1986a). Physically, this can directly be deduced from the fact that quantum mechanical waves are described by the Schrödinger equation, whereas classical waves are best described by the Helmholtz equation.

This fact can be illustrated by means of the one dimensional classical wave equation, equation (III.2.6), for vertical incidence. Now formulated as

$$\partial_{zz} u(z, t) - \frac{1}{c^2(z)} \partial_{tt} u(z, t) = 0, \quad (\text{III.2.12})$$

where  $\partial_{zz}$  stands for  $\partial^2/\partial z^2$ . Technically, this formulation implies constant  $\kappa$ , but that is not relevant for the present analysis.

Fourier transforming equation (III.2.12) and decomposing the velocity as

$$\frac{1}{c^2(z)} = \frac{1}{c_0^2} (1 - V(z)) \quad (\text{III.2.13})$$

results in

$$\partial_{zz} u(z, \omega) + k_0^2 u(z, \omega) = k_0^2 V(z) u(z, \omega), \quad (\text{III.2.14})$$

where the dispersion relation  $k = \omega/c$  has been used. This decomposition has already been used by Snieder (1990) and Dorren and Snieder (1995) to demonstrate similarities and dissimilarities between classical and quantum mechanical waves. It should be noted that  $(1 - V)$  does not only represent the squared relative slowness perturbation, but may also be identified with either the dielectric constant or the squared index of refraction, two variables which are more frequently encountered in optics.

If we compare equation (III.2.14) with the 1-D Schrödinger equation, we can see that the right-hand side in the classical wave equation is not constant, but directly proportional to the wave energy  $k_0^2$ . This in contrast to the Schrödinger equation, given by

$$\partial_{zz} u(z, \omega) + E u(z, \omega) = V(z) u(z, \omega). \quad (\text{III.2.15})$$

Hence, a quantum mechanical particle becomes more and more localized in disordered systems as its energy  $E$  decreases, since the potential  $V(z)$  gains in importance. Thus, localization of quantum mechanical particles (e.g. electrons) is always possible in the low frequency domain, since the energy of quantum particles is also proportional to frequency.

In contrast to this, in the low frequency domain, the influence of the potential barrier  $V(z)$  decreases with frequency for classical waves. As a matter of fact, a low frequency classical wave tends to perceive the disordered system as an essentially homogeneous medium. Thus, in the low frequency domain, no localization exists for classical waves, which will be shown hereafter using an approach leading to the effective medium theory of Backus (1962).

### III.2.2.3 Effective medium theory: homogenization

To show that no localization exists in the low frequency limit, a pulse propagating through the random medium is considered. However, the propagation distance is limited to several wavelengths. In fact, it is better to speak of the long wavelength limit than the low frequency limit, since it is the ratio  $L/\lambda$  which has to tend to zero (Sheng, 1995).

Since only propagation distances of the order of the wavelength are considered, the random medium (III.2.1) simplifies to

$$\begin{cases} \rho(z) = \rho_0 [1 + \sigma_\rho(z/\varepsilon^2)] \\ \kappa^{-1}(z) = \kappa_0^{-1} [1 + \sigma_\kappa(z/\varepsilon^2)], \end{cases} \quad (\text{III.2.16})$$

due to the fact that both profiles do not change on the mesoscopic scale  $z/\varepsilon$ . Thus, both  $\rho$  and  $\kappa$  are only functions of  $z/\varepsilon^2$ .

Now, a perturbation expansion of  $p$  and  $u_z$  will reveal the behavior of a probing pulse propagating in a 1-D medium for normal incidence as shown by Burrigge et al. (1994). If the following variables are introduced:  $w = u_z$ ,  $\tilde{z} = z/\varepsilon$ ,  $\tilde{t} = t/\varepsilon$  and  $y = z/\varepsilon^2 = \tilde{z}/\varepsilon$  then equations (III.2.10) and (III.2.11) can also be written as

$$\rho \partial_{\tilde{t}} \dot{w} + \partial_{\tilde{z}} p = 0 \quad (\text{III.2.17})$$



and

$$\kappa^{-1}\partial_{\bar{t}}p + \partial_{\bar{z}}\dot{w} = 0 \quad (\text{III.2.18})$$

in the absence of absorption and external body forces.

A perturbation expansion of both pressure and particle velocity up to the second order will yield

$$p(\bar{z}, y, \bar{t}) = p_0(\bar{z}, y, \bar{t}) + \varepsilon p_1(\bar{z}, y, \bar{t}) + \dots \quad (\text{III.2.19})$$

and

$$\dot{w}(\bar{z}, y, \bar{t}) = \dot{w}_0(\bar{z}, y, \bar{t}) + \varepsilon \dot{w}_1(\bar{z}, y, \bar{t}) + \dots \quad (\text{III.2.20})$$

Insertion of expressions (III.2.19) and (III.2.20) in (III.2.17) and (III.2.18) will then result in

$$\rho\partial_{\bar{t}}(\dot{w}_0 + \varepsilon\dot{w}_1) + (\partial_{\bar{z}} + \varepsilon^{-1}\partial_y)(p_0 + \varepsilon p_1) = 0 \quad (\text{III.2.21})$$

and

$$\kappa^{-1}\partial_{\bar{t}}(p_0 + \varepsilon p_1) + (\partial_{\bar{z}} + \varepsilon^{-1}\partial_y)(\dot{w}_0 + \varepsilon\dot{w}_1) = 0 \quad (\text{III.2.22})$$

if  $\partial_{\bar{z}}$  is replaced by  $\partial_{\bar{z}} + \varepsilon^{-1}\partial_y$  (Burrige et al., 1994). These last two expressions have to equal zero independent of the exact value of  $\varepsilon$ . Thus, the coefficients of the consecutive powers of  $\varepsilon$  have to equal zero. By collecting the terms of  $\varepsilon^{-1}$  and  $\varepsilon^0$ , we will obtain

$$\begin{cases} \partial_y \dot{w}_0 = 0 \\ \partial_y p_0 = 0 \end{cases} \quad (\text{III.2.23})$$

and

$$\begin{cases} \rho\partial_{\bar{t}}\dot{w}_0 + \partial_{\bar{z}}p_0 + \partial_y p_1 = 0 \\ \kappa^{-1}\partial_{\bar{t}}p_0 + \partial_{\bar{z}}\dot{w}_0 + \partial_y \dot{w}_1 = 0. \end{cases} \quad (\text{III.2.24})$$

Therefore, system (III.2.23) learns us that both  $p_0$  and  $\dot{w}_0$  are independent of  $y = z/\varepsilon^2$ , yielding that these variables are not sensitive to microscopic perturbations, but only to large scale features.

The second system, equation (III.2.24), can be simplified if it is assumed that the dependence of  $\kappa^{-1}$  and  $\rho$  on  $y$  is periodic. In principle, periodicity of  $y$  is not required, but it simplifies the mathematics considerably (Asch et al., 1991). In that case, both third terms will vanish after averaging over  $y$  and the expressions can be replaced by

$$\begin{cases} \langle \rho \rangle \partial_{\bar{t}}\dot{w}_0 + \partial_{\bar{z}}p_0 = 0 \\ \langle \kappa^{-1} \rangle \partial_{\bar{t}}p_0 + \partial_{\bar{z}}\dot{w}_0 = 0, \end{cases} \quad (\text{III.2.25})$$

yielding that probing pulses in random media travel as if through a homogeneous medium for propagation distances of the order of a wavelength, since  $\langle \rho \rangle$  and  $\langle \kappa^{-1} \rangle$  are constants on the order of a wavelength. A phenomenon also known as the effective medium theory of Backus (1962).

Thus, any medium will appear locally homogeneous, at least for propagation distances of the order of a wavelength. However, for larger distances the pulse will be attenuated and broadened due to multiple scattering. Nevertheless, the theory does imply that, in the long wavelength limit, no localization of energy can occur, since the medium will appear homogeneous to the pulse, which will, therefore, remain constant in amplitude and shape.

Another interesting feature of the effective medium theory is that it shows how averaging has to be done. Namely, the local vertical velocity  $c_v$  is given by

$$c_v^2 = \langle \kappa^{-1} \rangle^{-1} \langle \rho \rangle^{-1}. \quad (\text{III.2.26})$$

Thus, the local compressibility has to be calculated using the harmonic mean, whereas density is given by its arithmetic average to obtain the effective vertical velocity. Furthermore, the more elaborate analysis of Asch et al. (1991) shows that the horizontal velocity  $c_h$  is given by

$$c_h^2 = \langle \kappa^{-1} \rangle^{-1} \langle \rho^{-1} \rangle, \quad (\text{III.2.27})$$

thereby indicating the existence of anisotropy due to fine scale layering. A fact also stated by Backus (1962).

Equations (III.2.26) and (III.2.27) are very interesting as they learn us how the phase velocity at zero frequency for vertical incidence has to be evaluated. It follows that  $c_0$ , the effective velocity, is given by

$$c_0 \equiv \langle \rho_0 / \kappa_0 \rangle^{-1/2} = \langle c^{-2} \rangle^{-1/2}. \quad (\text{III.2.28})$$

Its high frequency equivalent, i.e. the geometric velocity, can be calculated by averaging travel-times over layer thicknesses, i.e.

$$c_\infty = \langle c^{-1} \rangle^{-1}, \quad (\text{III.2.29})$$

which results in

$$c_\infty \approx c_0 \left[ 1 - \frac{1}{8} \langle \sigma_\kappa^2 \rangle^{-1} + O(\sigma_\kappa^3) \right] \approx c_0 \left[ 1 + \frac{1}{8} \langle \sigma_\kappa^2 \rangle \right] + O(\sigma_\kappa^3) \quad (\text{III.2.30})$$

for a constant density profile and vertical incidence. Equation (III.2.30) is most easily established using a second order Taylor expansion of (III.2.1) and remembering that  $\langle \sigma_\kappa \rangle = 0$ . However, it is only valid for small values of  $\sigma_\kappa$ , constant  $\rho_0$  and  $\kappa_0$  and stationary  $\sigma_\kappa$ .

It should be noted that the predicted homogenization of the effective medium theory of Backus (1962) holds true for any medium, be it periodic or chaotic, whereas wave localization can only occur in non-periodic media. As a direct consequence, in this analysis  $a$  should not be seen as the typical correlation length of the heterogeneities, but as the typical length scale of these heterogeneities (like e.g. the average layer thickness in 1-D problems). Otherwise  $a$  will tend to infinity for periodic media. The non-existence of wave localization for periodic media will be shown below.

#### III.2.2.4 Amplitude and phase: two self-averaging quantities

**Amplitude** Localization in quantum mechanics implies that energy is bounded in a specific region of a disordered medium and has an exponential decay outside this region. As was already mentioned above, a similar phenomenon occurs if a classical wave enters a disordered medium: the energy of the probing pulse (and thus its amplitude) will decay exponentially with penetration depth, due to multiple wave scattering. Hence, the transmission coefficient decays as

$$|T| = \lim_{z \rightarrow \infty} e^{-\gamma z} \Leftrightarrow \gamma = \lim_{z \rightarrow \infty} -\frac{1}{z} \ln |T|, \quad (\text{III.2.31})$$

where  $\gamma$  is known as the Lyapunov exponent and depends both on frequency and the characteristics of a medium. The reciprocal  $\gamma^{-1}$  of the Lyapunov exponent is equivalent to the (maximum)



localization length  $l(f)$  and defines the characteristic attenuation length or penetration depth of a specific medium. Note that here the definition of amplitude decay has been used.

To obtain an expression for  $l(f)$  a so-called matched-medium is considered, i.e. the chaotic part of the medium, equation (III.2.1), has a thickness  $L$  and is sandwiched between 2 homogeneous half-spaces with matching density and compressibility. Moreover, it will be assumed that  $\rho_0(z)$  and  $\kappa_0^{-1}(z)$  are constant and  $\sigma_\kappa$  stationary. In addition, as a further simplification, the density is assumed to be constant in the whole medium, i.e.  $\sigma_\rho = 0$ . Thus, the random medium is described by

$$\kappa^{-1}(z) = \begin{cases} \kappa_0^{-1} & z < 0, z > L \\ \kappa_0^{-1}[1 + \sigma_\kappa(z/\varepsilon^2)] & 0 \leq z \leq L \end{cases} \quad (\text{III.2.32})$$

$$\rho(z) = \rho_0 \quad \text{everywhere}$$

To determine the behavior of wave propagation in such a system, the wave equation in frequency - wave number domain, equation (III.2.9), will be used. A plane-wave traveling downwards in the homogeneous region  $z < 0$  is described by

$$\hat{p} = r \sin \{k_0 \sin(\phi_i)x + k_0 \cos(\phi_i)z - \omega t\} \quad (\text{III.2.33})$$

$$\hat{w} = \frac{1}{\rho_0 \omega^2} \partial_z \hat{p} = \frac{k_0 \cos(\phi_i)}{\rho_0 \omega^2} r \cos \{k_0 \sin(\phi_i)x + k_0 \cos(\phi_i)z - \omega t\}$$

with  $\phi_i$  denoting the angle of incidence and  $k_x = k_0 \sin(\phi_i)$  and  $k_z = k_0 \cos(\phi_i)$ . In the chaotic part of the medium, it will then be described by

$$\hat{p} = r_s(z) \sin \vartheta_s(z) \quad (\text{III.2.34})$$

$$\hat{w} = \frac{k_0 \cos(\phi_i)}{\rho_0 \omega^2} r_s(z) \cos \vartheta_s(z),$$

where the pressure  $\hat{p}$  and vertical displacement  $\hat{w}$  have been separated in amplitude  $r_s(z)$  and phase  $\vartheta_s(z)$  factors. Moreover, it is assumed that the horizontal slowness and wavenumber remain constant.

If such a wave impinges from above on the disordered system, the following system of equations is obtained (Shapiro and Zien, 1993)

$$\partial_z \vartheta_s(z) = k_0 \cos(\phi_i) + \frac{k_0 \sigma_\kappa(z)}{\cos(\phi_i)} \sin^2 \vartheta_s(z) \quad (\text{III.2.35})$$

and

$$\ln r_s(z) = \ln r_s(0) - \frac{k_0}{2 \cos(\phi_i)} \int_0^z \sigma_\kappa(z') \sin 2\vartheta_s(z') dz' \quad (\text{III.2.36})$$

by inserting equations (III.2.34) and (III.2.32) in (III.2.9). Both expressions are only valid for small angles of incidence (mathematically to prevent division by zero, but physically to prevent P-S conversions and other complications due to the occurrence of turning points and inhomogeneous waves). Since  $T$  is defined as  $r_s(z)/r_s(0)$ , we get (using equation (III.2.31) with  $l^{-1} = \gamma$ )

$$l^{-1} = \pm \lim_{z \rightarrow \infty} \frac{1}{z} \frac{k_0}{2 \cos(\phi_i)} \int_0^z \sigma_\kappa(z') \sin 2\vartheta_s(z') dz'. \quad (\text{III.2.37})$$

The plus-minus sign enters the equation, due to the fact that system (III.2.9) has zero trace, yielding that a second, linearly independent solution exists, namely a wave which grows with depth,  $e^{z/l}$ . This is also illustrated by the fact that  $\hat{p} = r_c \cos \vartheta_c$  constitutes also a solution to the wave equation and has already been discussed in the introduction. Imposing the right boundary conditions will eliminate this wave (Sheng et al., 1986a; Sheng et al., 1990; Shapiro and Zien, 1993).

Equation (III.2.37) learns us that the obtained estimation for  $l(f)$  may fluctuate strongly around its true value, due to the sinus in the integrand. Only if the phase  $\vartheta(z)$  stabilizes, an accurate estimation may be obtained. It is proven in Lifshits et al. (1988) that  $l(f)$  is self-averaging for  $z \gg l$ . Thus, the phase becomes stationary in this limit.

If we replace  $z$  with  $L$ , i.e. the thickness of the disordered medium, then, due to this self-averaging property, the limit of  $L \rightarrow \infty$  of a single realization may be replaced by its expected value. That is, equation (III.2.37) transforms to

$$l^{-1} = \pm \lim_{L \rightarrow \infty} \frac{k_0}{2L \cos(\phi_i)} \int_0^L E[\sigma_\kappa(z') \sin 2\vartheta_s(z')] dz', \quad (\text{III.2.38})$$

where 'E' denotes expected value or expectation (Shapiro and Zien, 1993).

By using a perturbation expansion of equation (III.2.35) up to the second order of  $\sigma_\kappa$  and the fact that  $\sigma_\kappa$  is stationary and has zero mean, equation (III.2.37) transforms to

$$l^{-1} = \mp \frac{k_0^2}{4 \cos^2(\phi_i)} \int_0^\infty E[\sigma_\kappa(z) \sigma_\kappa(z + \xi)] \cos(2k_0 \cos(\phi_i) \xi) d\xi \quad (\text{III.2.39})$$

$$= \mp \frac{k_0^2}{8 \cos^2(\phi_i)} \int_{-\infty}^\infty E[\sigma_\kappa(z) \sigma_\kappa(z + \xi)] e^{-2ik_0 \cos(\phi_i) \xi} d\xi$$

if the higher moments of  $\sigma_\kappa$  decrease sufficiently fast (Lifshits et al., 1988; Shapiro and Zien, 1993). A detailed derivation can be found in Shapiro and Hubral (1999). Since only the exponentially decaying pulse exists physically, just the positive integrand is retained in equation (III.2.39).  $E[\sigma_\kappa(z) \sigma_\kappa(z + \xi)]$  represents the spatial autocorrelation function of the random medium.

It is not difficult to show that expression (III.2.39) leads to the well-known O'Doherty-Anstey relation as the right-hand side of this expression equals  $R(k)$ , the power spectrum of the reflection coefficient series (Shapiro and Zien, 1993). Therefore, using this result and equation (III.2.31) yields

$$|T(k_0)| \sim e^{-LR(k_0)}, \quad (\text{III.2.40})$$

being the O'Doherty-Anstey (1971) relation. Hence, their relation can be understood as a wave localization phenomenon (Shapiro and Zien, 1993).

**Phase** Lifshits et al. (1988) proved that the Lyapunov exponent is not the only self-averaging quantity. Also the phase of the pulse divided by  $L$  is self-averaging (as it represents the imaginary part of the signal and the pulse has to remain causal). Hence, the second self-averaging quantity  $\psi$  is defined by

$$\psi = \lim_{L \rightarrow \infty} \frac{\vartheta_s(L)}{L}. \quad (\text{III.2.41})$$



A similar derivation as for the Lyapunov coefficient  $\gamma$  yields (Shapiro et al., 1994; Shapiro and Hubral, 1999)

$$\psi = k_0 \cos(\phi_i) - \frac{k_0^2}{4 \cos^2(\phi_i)} \int_0^\infty E[\sigma_\kappa(z)\sigma_\kappa(z+\xi)] \sin(2k_0 \cos(\phi_i)\xi) d\xi. \quad (\text{III.2.42})$$

Notice that  $\psi L$  represents the unwrapped phase. Therefore,  $\psi$  equals the effective vertical wavenumber.

**Ergodicity** In geophysical or seismological applications only a single realization of the Earth exists. Therefore, ergodicity has to be invoked, i.e. ensemble (or configurational) averaging can be replaced by spatial averaging:

$$E[\sigma_\kappa(z)\sigma_\kappa(z+\xi)] = \langle \sigma_\kappa(z)\sigma_\kappa(z+\xi) \rangle \equiv \lim_{z \rightarrow \infty} \frac{1}{z} \int_0^z \sigma_\kappa(z')\sigma_\kappa(z'+\xi) dz'. \quad (\text{III.2.43})$$

### III.2.2.5 Frequency dependent scattering and dispersion

To determine the behavior of a pulse probing a chaotic medium and thereby the influence of small-scale heterogeneities on the frequency dependent apparent attenuation and dispersion, 3 different spatial autocorrelation functions will be considered. These are the Gaussian, exponential and Von Kármán correlation functions. Gaussian correlation functions are related to very smooth media (they are infinitely differentiable), whereas exponential autocorrelation functions are only piecewise continuous and contain, therefore, first order discontinuities. The Von Kármán function (Tatarskii, 1961), on the other hand, describes a very large class of correlation functions ranging from functions smoother than the exponential function to functions having discontinuities on any scale and displaying, therefore, fractal characteristics.

To analyze the influence of multiple scattering, the frequency dependent self-averaged amplitude and phase will be calculated. With these, the frequency dependent attenuation and vertical wavenumber will be estimated. However, as the quality factor  $Q$ , the phase velocity  $c$  and the group velocity  $U$  are more commonly used in seismology, I will also provide these. The scattering coefficient  $Q^{-1}$  is related to the Lyapunov exponent by

$$Q^{-1} = \frac{2\gamma}{k_0 \cos(\phi_i)}. \quad (\text{III.2.44})$$

and the vertical phase velocity  $c_z$  is given by

$$c_z(f) = 2\pi f \psi^{-1} = \frac{c_0}{\cos(\phi_i)} \{1 + \delta c\}, \quad (\text{III.2.45})$$

where  $\delta c$  represents the relative perturbation of the phase velocity. The vertical group velocity  $U_z$  is associated with the actual propagation velocity of energy and can be calculated from

$$U_z(f) = \left[ \frac{\partial \psi}{\partial \omega} \right]^{-1} = \frac{c_0}{\cos(\phi_i)} \{1 + \delta U\}, \quad (\text{III.2.46})$$

with  $\delta U$  the frequency dependent part of the group velocity.

To simplify expressions, let  $k'_0$  denote  $k_0 \cos(\phi_i)$ , the wavenumber corrected for the angle of incidence. All expressions are easily translated to the frequency domain by using the dispersion relation, i.e.  $k_0 = 2\pi f/c_0$ .

**Gaussian correlation function** If smooth transitions are assumed then the medium is best described using a Gaussian correlation function, i.e.

$$E[\sigma_\kappa(z)\sigma_\kappa(z+\xi)] = \langle \sigma_\kappa^2 \rangle e^{-\xi^2/a^2}. \quad (\text{III.2.47})$$

Substitution of this equation in (III.2.39) and (III.2.42) yields

$$l_G = \frac{8 \cos^4(\phi_i) a}{\pi^{1/2} \langle \sigma_\kappa^2 \rangle} (k'_0 a)^{-2} e^{-(k'_0 a)^2} \quad (\text{III.2.48})$$

for the localization length and

$$\psi_G = k'_0 \left[ 1 - \frac{(k'_0 a)^2 \langle \sigma_\kappa^2 \rangle}{4 \cos^4(\phi_i)} {}_1F_1(1; 1.5; -(k'_0 a)^2) \right] \quad (\text{III.2.49})$$

for the self-averaged phase. Expression (III.2.49) can be obtained using the table of integrals of Gradshteyn and Ryzhik (1980).  ${}_1F_1(\cdot)$  represents the degenerate or confluent hypergeometric function. Furthermore, the inverse quality factor  $Q^{-1}$  is expressed by (using equations (III.2.44) and (III.2.48))

$$Q_G^{-1} = \frac{\pi^{1/2} \langle \sigma_\kappa^2 \rangle}{4 \cos^4(\phi_i)} k'_0 a e^{-(k'_0 a)^2}. \quad (\text{III.2.50})$$

The frequency dependent part of the dispersion, i.e.  $\delta c$  in expression (III.2.45), equals

$$\delta c_G \approx \frac{(k'_0 a)^2 \langle \sigma_\kappa^2 \rangle}{4 \cos^4(\phi_i)} {}_1F_1(1; 1.5; -(k'_0 a)^2) + O(\sigma_\kappa^4). \quad (\text{III.2.51})$$

Finally, using equations (III.2.46) and (III.2.49) and the relation (Gradshteyn and Ryzhik, 1980)

$$\frac{{}_1F_1(\alpha; \gamma; z)}{dz} = \frac{\alpha}{\gamma} {}_1F_1(\alpha+1; \gamma+1; z), \quad (\text{III.2.52})$$

the vertical group velocity can be established to be

$$U_{G,z} = \frac{c_0}{\cos(\phi_i)} \left[ 1 - \frac{3 \langle \sigma_\kappa^2 \rangle}{4 \cos^4(\phi_i)} (k'_0 a)^2 {}_1F_1(1; 1.5; -(k'_0 a)^2) + \frac{1 \langle \sigma_\kappa^2 \rangle}{3 \cos^4(\phi_i)} (k'_0 a)^4 {}_1F_1(2; 2.5; -(k'_0 a)^2) \right]^{-1}. \quad (\text{III.2.53})$$

Hence,  $\delta U_G$  equals

$$\delta U_G \approx \frac{3 \langle \sigma_\kappa^2 \rangle}{4 \cos^4(\phi_i)} (k'_0 a)^2 {}_1F_1(1; 1.5; -(k'_0 a)^2) - \frac{1 \langle \sigma_\kappa^2 \rangle}{3 \cos^4(\phi_i)} (k'_0 a)^4 {}_1F_1(2; 2.5; -(k'_0 a)^2) + O(\sigma_\kappa^4). \quad (\text{III.2.54})$$

Analysis of equations (III.2.48) to (III.2.54) in both the high and low frequency limits learns us that (see also Figure III.2.1):

- The localization length behaves as  $l_G \sim k_0^{-2}$  in the low frequency domain and as  $l_G \rightarrow \infty$  for very high frequencies. Hence, no localization of energy exists in either limit. Physically,



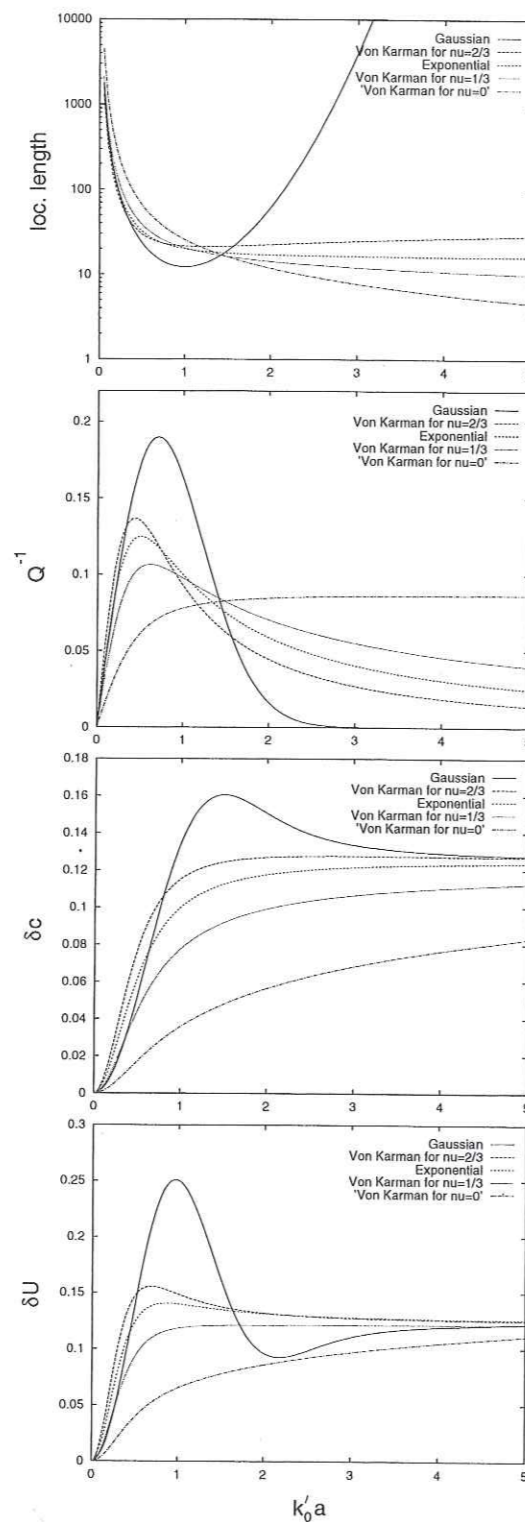


Figure III.2.1: Localization length, apparent attenuation and dispersion as a function of  $k_0 a$ . Vertical scales have been normalized with  $\langle \sigma_\kappa^2 \rangle / \cos^4(\phi_i)$ . In addition, units of the localization length are expressed in typical scale-lengths  $a$ . The constant  $\Gamma(0)$  in the Von Kármán function for  $\nu = 0$  has been replaced by the constant 9.0 to display the general trend. Notice the strong dependence of all quantities on the angle of incidence, indicating e.g. the presence of anisotropy due to fine-scale layering.

this is due to the fact that in the low frequency domain (long wavelength limit) the medium becomes effectively homogeneous and that no scattering occurs either in the high frequency limit as no discontinuities are present in smooth media in that limit. The  $f^2$ -dependence of  $l_G^{-1}$  is explained by the  $f^2$ -frequency variation of Rayleigh scattering for stratified media or generally  $f^{d+1}$  for  $d$ -dimensional media (White et al., 1990).

Finally, a minimum localization length exists for  $k_0 a = 1$  or  $\lambda = 2\pi a$  for vertical incidence, thereby favoring Mie scattering.

- The scattering coefficient  $Q_G^{-1}$  displays a similar behavior. No scattering occurs in either the low or high frequency limit. Furthermore, a maximum in the apparent attenuation due to scattering occurs for  $k_0 a = \sqrt{2}/2$ , yielding again that scattering is most efficient for wavelengths of the same order as the correlation length of the heterogeneities, i.e. due to Mie scattering.

- The vertical wavenumber  $\psi_G$  behaves as

$$\lim_{k_0 \rightarrow 0} \psi_G = k_0' \quad (III.2.55)$$

in the low frequency limit and as

$$\lim_{k_0 \rightarrow \infty} \psi_G = k_0' \left[ 1 - \frac{\langle \sigma_\kappa^2 \rangle}{8 \cos^4(\phi_i)} \right] \quad (III.2.56)$$

in the high frequency limit. This last expression can be found using the relation (Abramowitz and Stegun, 1972)

$${}_1F_1(a; b; z) = \frac{\Gamma(b)}{\Gamma(b-a)} (-z)^{-a} \left[ 1 + O(|z|^{-1}) \right], \quad |z| \rightarrow \infty \text{ and } z < 0, \quad (III.2.57)$$

where  $\Gamma(\cdot)$  represents the Gamma function. Equation (III.2.55) clearly shows the occurrence of the homogenization phenomenon in the low frequency domain.

Finally, using equation (III.2.45), we obtain for  $\delta c_G$  in the high frequency limit

$$\lim_{k_0 \rightarrow \infty} \delta c_G = \frac{\langle \sigma_\kappa^2 \rangle}{8 \cos^4(\phi_i)} \quad (III.2.58)$$

if higher order moments of  $\sigma_\kappa$  are neglected. Equations (III.2.55) and (III.2.56) or (III.2.58) confirm equations (III.2.28) and (III.2.30) for constant density profiles and vertical incidence. Thus, the phase velocity satisfies the established values for both the effective ( $c_0$ ) and the geometric ( $c_\infty$ ) velocities. Equation (III.2.28) results from equations (III.2.55) and (III.2.45) and the dispersion relation and equation (III.2.30) from equation (III.2.56) and again (III.2.45) and the dispersion relation.

Finally, the phase velocity shows an overshoot with a maximum at approximately  $k_0 a = 1.5$  (Figure III.2.1).

- It can easily be shown that the group velocity does also satisfy the established values for the effective ( $c_0$ ) and geometric ( $c_\infty$ ) velocities. Moreover, Airy phases (i.e. extrema of the group velocity curve) occur at  $k_0 a = 0, 0.96$  and  $2.18$  and for  $k_0 a \rightarrow \infty$ . Airy phases



are normally associated with strong arrivals as waves are allowed to interfere constructively over a large frequency band due to the presence of an extremum (because of the absence of dispersion). Naturally, Airy phases occur in both frequency limits, since no dispersion exists in these extremes. However, the second Airy phase is situated in the domain of Mie scattering, rendering it hard to detect. Nevertheless, the third one may be detectable as it is characterized by a very high quality factor (Figure III.2.1).

**Exponential correlation function** If an exponential spatial correlation between heterogeneities exists then the ensemble averaging can be replaced by

$$E[\sigma_\kappa(z)\sigma_\kappa(z+\xi)] = \langle \sigma_\kappa^2 \rangle e^{-|\xi|/a}. \quad (\text{III.2.59})$$

Thus, substitution of this equation in (III.2.39) and (III.2.42) yields

$$l_{exp} = \frac{4 \cos^4(\phi_i) a}{\langle \sigma_\kappa^2 \rangle} [4 + (k'_0 a)^{-2}] = \frac{16 \cos^4(\phi_i) a}{\langle \sigma_\kappa^2 \rangle} + \frac{c_0^2 \cos^2(\phi_i)}{\pi^2 \langle \sigma_\kappa^2 \rangle a} f^{-2}, \quad (\text{III.2.60})$$

$$\psi_{exp} = k'_0 \left[ 1 - \frac{\langle \sigma_\kappa^2 \rangle (k'_0 a)^2}{2 \cos^4(\phi_i) 4(k'_0 a)^2 + 1} \right] \quad (\text{III.2.61})$$

and for the apparent attenuation due to scattering (using equation (III.2.44))

$$Q_{exp}^{-1} = \frac{\langle \sigma_\kappa^2 \rangle k'_0 a}{2 \cos^4(\phi_i) 4(k'_0 a)^2 + 1}. \quad (\text{III.2.62})$$

Moreover,  $\delta c_{exp}$  can be found to be

$$\delta c_{exp} \approx \frac{\langle \sigma_\kappa^2 \rangle (k'_0 a)^2}{2 \cos^4(\phi_i) 4(k'_0 a)^2 + 1} + O(\sigma_\kappa^4) \quad (\text{III.2.63})$$

using expression (III.2.45). Finally, we find for the vertical group velocity  $U_{exp,z}$  (with equations (III.2.46) and (III.2.61))

$$U_{exp,z} = \frac{c_0}{\cos(\phi_i)} \left[ 1 - \frac{1}{2} \frac{\langle \sigma_\kappa^2 \rangle (k'_0 a)^2 (4(k'_0 a)^2 + 3)}{\cos^4(\phi_i) (4(k'_0 a)^2 + 1)^2} \right]^{-1} \quad (\text{III.2.64})$$

and for  $\delta U_{exp}$

$$\delta U_{exp} \approx \frac{1}{2} \frac{\langle \sigma_\kappa^2 \rangle (k'_0 a)^2 (4(k'_0 a)^2 + 3)}{\cos^4(\phi_i) (4(k'_0 a)^2 + 1)^2} + O(\sigma_\kappa^4). \quad (\text{III.2.65})$$

Analysis of equations (III.2.60) to (III.2.65) learns us that (see also Figure III.2.1):

- The localization length  $l_{exp}$  behaves again as  $l_{exp} \sim f^{-2}$  in the low frequency domain. However, it saturates to a constant for  $f \rightarrow \infty$ , since exponential models are associated to piece-wise continuous media. Hence, above a certain frequency, all wavelengths will 'feel' all discontinuities in exactly the same way.

The obtained expression for the localization length has the same form  $l_{exp}(f) = c_1 + c_2 f^{-2}$  as proposed by Sheng et al. (1990) and White et al. (1990). However, the constant  $c_2$  of White et al. (1990) equals  $c_0^2/(2\pi^2 \langle \sigma_\kappa^2 \rangle a)$  for an exponential correlation function and vertical incidence. This is probably due to a misprint in White et al. (1990) as can easily be checked from expressions (32), (34b) and (35a) in Sheng et al. (1990).

- The scattering coefficient  $Q_{exp}^{-1}$  shows a different behavior from  $l_{exp}$ . It is more similar to  $Q_G^{-1}$  as  $Q_{exp}^{-1}$  goes to zero for either frequency limit. Moreover, the apparent attenuation is maximal for  $k_0 a = 1/2$ , thereby again favoring Mie scattering.
- The effective vertical wavenumber  $\psi_{exp}$  shows a similar behavior as  $\psi_G$  in both frequency limits, i.e. an effective homogeneous medium for  $k_0 \rightarrow 0$  given by expression (III.2.55) and a speed up of velocity for  $k_0 \rightarrow \infty$  given by expressions (III.2.56) and (III.2.58). However,  $\delta c_{exp}$  is monotonically rising (Figure III.2.1).
- The group velocity satisfies again both velocity limits. However, now, only 3 Airy phases exist in naturally the low and high frequency limit and for  $k'_0 a = \sqrt{3}/2$ . Strangely, the second Airy phase occurs again in the domain of Mie scattering. Hence, it cannot be associated with a strong arrival.

**Von Kármán correlation function** The third kind of correlation function to be examined is the Von Kármán functions which describes a very large class of models, ranging from smooth functions to fractal-like models. It is given by

$$E[\sigma_\kappa(z)\sigma_\kappa(z+\xi)] = \frac{2^{1-\nu} \langle \sigma_\kappa^2 \rangle}{\Gamma(\nu)} \left( \frac{|\xi|}{a} \right)^\nu K_\nu \left( \frac{|\xi|}{a} \right), \quad (\text{III.2.66})$$

where  $K_\nu(\cdot)$  represents the third modified Bessel function of order  $\nu$ , also known as the MacDonald function. Autocorrelation functions described by it include the Kolmogorov turbulence function ( $\nu = 1/3$ ), the exponential correlation function ( $\nu = 1/2$ ) and a function with fractal characteristics ( $\nu = 0$ ) having discontinuities on any scale. In general,  $\nu$  is limited to range between 0 and 1. For  $\nu > 1/2$  functions are smoother than the exponential function and rougher for  $\nu < 1/2$ . In equation (III.2.66),  $K_0(0)$  is defined as  $\Gamma(0)/2$  to produce a value equal to the variance for zero lag ( $\xi = 0$ ). The solutions to the integrals of equations (III.2.39) and (III.2.42) with a Von Kármán correlation function can be found using the table of integrals of Gradshteyn and Ryzhik (1980), resulting in

$$l_{VK} = \frac{4 \cos^4(\phi_i) a \Gamma(\nu)}{\langle \sigma_\kappa^2 \rangle \sqrt{\pi} \Gamma(\nu + 1/2)} \frac{(4(k'_0 a)^2 + 1)^{\nu+1/2}}{(k'_0 a)^2} \quad (\text{III.2.67})$$

for  $a^{-1} > 0$ ,  $k_0 > 0$  and  $\nu > -1/2$  and for  $\psi_{VK}$  in

$$\psi_{VK} = k'_0 \left[ 1 - \frac{\langle \sigma_\kappa^2 \rangle \Gamma(\nu + 1)}{\cos^4(\phi_i) \Gamma(\nu)} (k'_0 a)^2 {}_2F_1(\nu + 1, 1; 1.5; -4(k'_0 a)^2) \right] \quad (\text{III.2.68})$$

for  $\nu > -1$ . Notice that  $\Gamma(\nu + 1)/\Gamma(\nu)$  may be replaced by  $\nu$  for  $\nu \neq 0$ . The function  ${}_2F_1(\cdot)$  is known as Gauss' hypergeometric function. Furthermore, the apparent attenuation due to scattering is given by

$$Q_{VK}^{-1} = \frac{\langle \sigma_\kappa^2 \rangle \sqrt{\pi} \Gamma(\nu + 1/2)}{2 \cos^4(\phi_i) \Gamma(\nu)} \frac{k'_0 a}{(4(k'_0 a)^2 + 1)^{\nu+1/2}} \quad (\text{III.2.69})$$

and  $\delta c_{VK}$  by

$$\delta c_{VK} \approx \frac{\langle \sigma_\kappa^2 \rangle \Gamma(\nu + 1)}{\cos^4(\phi_i) \Gamma(\nu)} (k'_0 a)^2 {}_2F_1(\nu + 1, 1; 1.5; -4(k'_0 a)^2) + O(\sigma_\kappa^4). \quad (\text{III.2.70})$$



Finally, the group velocity can be calculated using equations (III.2.46) and (III.2.68) and the relation (Abramowitz and Stegun, 1972)

$$\frac{d_2 F_1(a, b; c; z)}{dz} = \frac{ab}{c} {}_2F_1(a+1, b+1; c+1; z). \quad (\text{III.2.71})$$

This results in

$$U_{VK,z} = \frac{c_0}{\cos(\phi_i)} \left[ 1 - \frac{3\langle\sigma_\kappa^2\rangle\Gamma(\nu+1)}{\cos^4(\phi_i)\Gamma(\nu)} (k'_0 a)^2 {}_2F_1(\nu+1, 1; 1.5; -(k'_0 a)^2) + \frac{16(\nu+1)\langle\sigma_\kappa^2\rangle\Gamma(\nu+1)}{3\cos^4(\phi_i)\Gamma(\nu)} (k'_0 a)^4 {}_2F_1(\nu+2, 2; 2.5; -(k'_0 a)^2) \right]^{-1} \quad (\text{III.2.72})$$

and

$$\delta U_{VK} \approx \frac{3\langle\sigma_\kappa^2\rangle\Gamma(\nu+1)}{\cos^4(\phi_i)\Gamma(\nu)} (k'_0 a)^2 {}_2F_1(\nu+1, 1; 1.5; -(k'_0 a)^2) - \frac{16(\nu+1)\langle\sigma_\kappa^2\rangle\Gamma(\nu+1)}{3\cos^4(\phi_i)\Gamma(\nu)} (k'_0 a)^4 {}_2F_1(\nu+2, 2; 2.5; -(k'_0 a)^2) + O(\sigma_\kappa^4) \quad (\text{III.2.73})$$

Expressions (III.2.67) to (III.2.73) are more difficult to analyze than their equivalents for Gaussian or exponential correlation functions. Nevertheless, it can be concluded that (see also Figure III.2.1)

- The localization length is always proportional to  $f^{-2}$  in the low frequency domain due to Rayleigh scattering. However, in the high frequency domain,  $l \sim f^{2\nu-1}$ , implying a divergence for  $\nu > 1/2$  (smooth models), the convergence to a constant for  $\nu = 1/2$  (as it equals the exponential case) and a convergence to zero for  $\nu < 1/2$ . This last feature is due to the existence of discontinuities on any scale for  $\nu < 1/2$ . Therefore, no energy penetrates in the high frequency limit for these values of  $\nu$ . Therefore, the statement of Sheng et al. (1986a) that a minimum localization length exists is not necessarily true for media displaying fractal-like characteristics.
- The scattering coefficient  $Q^{-1}$  always equals 0 for  $f \rightarrow 0$  due to the homogenization of the medium. Furthermore, in the high frequency limit,  $Q^{-1}$  is proportional to  $f^{-2\nu}$ , implying again a convergence to zero for any  $\nu$  except  $\nu = 0$ . Moreover, a maximum is present (except for  $\nu = 0$ ). Such maxima occur always in the domain of Mie scattering, i.e.  $k'_0 a \approx 1$ . However, they tend to shift to lower values of  $k'_0 a$  for increasing  $\nu$ . On the other hand, for  $\nu = 0$ ,  $Q^{-1}$  is monotonically rising and converges to a constant, i.e.

$$\lim_{f \rightarrow \infty} Q_{VK,\nu=0}^{-1} = \frac{\langle\sigma_\kappa^2\rangle\pi}{4\cos^4(\phi_i)\Gamma(0)}, \quad (\text{III.2.74})$$

which is effectively zero due to the division by  $\Gamma(0)$ . A nearly constant  $Q^{-1}$  over a large frequency range may be characteristic for fractal-like distributions as fractals are self-affine. Hence, no typical scale-length exists and thus no maximum occurs for  $Q^{-1}$ .

- Expression (III.2.68) is difficult to analyze analytically due to the presence of the hypergeometric function. Nevertheless, numerical calculations agree with previously obtained results, i.e. limits (III.2.55) and (III.2.56) are satisfied for any  $\nu$  ( $0 \leq \nu < 1$ ). Moreover, the smoother models ( $\nu > 1/2$ ) tend to show an 'overshoot' (although this is not visible on the scale of Figure III.2.1), whereas rougher models ( $\nu < 1/2$ ) are monotonically rising. Nevertheless, the limit (III.2.58) for  $\delta c$  is always satisfied in the high frequency domain.

- Numerical calculations of  $\delta U_{VK}$  reveal the following picture. Like the phase velocity, the group velocity satisfies both the limits for the effective and the geometric velocities. However, in between these limits, the exact behavior depends strongly on the value of  $\nu$ . For  $\nu < 1/3$   $\delta U$  is monotonically rising, for  $1/3 < \nu < 0.40$  a transition zone occurs in which a maximum (overshoot) is followed by a minimum (in analogy with the Gaussian case), for  $0.40 < \nu < 1/2$  only the overshoot remains and for  $\nu > 1/2$  the overshoot is again followed by a minimum (not visible on the scale of Figure III.2.1). Bounds have only been established approximately.

As a consequence, the number and positions of Airy phases depends strongly on the exact value of  $\nu$ . Nevertheless, a comparison of the established curves for  $Q^{-1}$  and  $\delta U$  in Figure III.2.1 clearly shows the existence of some direct relation between these two parameters. For instance, the maxima (or corners) in  $Q^{-1}$  and  $\delta U$  seem to be separated by a constant shift and both move to higher values of  $k'_0 a$  for decreasing  $\nu$ .

A final note concerning the particular case of  $\nu = 0$ . The factor  $\Gamma(0)$  is needed to normalize the Von Kármán function. However, in non-perfect realizations (e.g. due to finite layer thicknesses), it can be replaced by a constant which is much smaller than infinity. Nevertheless, its localization length seems to be much larger and  $c$  and  $U$  approach the geometric velocity for much higher frequencies than for the other autocorrelation functions.

**Kramers-Krönig relations** A recent controversy has arisen concerning wave localization theory in that estimates of the effective ( $c_0$ ) and the geometric ( $c_\infty$ ) velocities calculated from sonic well logs do not coincide with those derived from measurements of the localization length obtained by means of numerical wave propagation in these same well logs and the Kramers-Krönig relations. This might indicate a discrepancy between effective medium theory and wave localization theory (Scales, 1993).

Nevertheless, both theories should give the same values for  $c_0$  and  $c_\infty$  as for linear, passive and causal systems,  $l^{-1}(f)$  and  $c(f)$  are not independent of each other, but related by Hilbert transforms (Beltzer, 1989). These relations are known as the Kramers-Krönig relations. For instance,  $c_0$  and  $c_\infty$  are related to  $l(f)$  by (Beltzer, 1989)

$$c_\infty = c_0 \left[ 1 - \frac{c_0}{\pi^2} \int_0^\infty \frac{df}{l(f)f^2} \right]^{-1}. \quad (\text{III.2.75})$$

I will show that this discrepancy is not due to a lack in theory, but does constitute an important issue for practical inversion problems.

White et al. (1990) were the first to try to estimate the frequency dependent localization length in realistic numerical simulations using real sonic well log data as random models. They calculated the transmission coefficient  $|T|$  of a synthetic wave passing through a single stack of layers, i.e. without using configurational averaging. Scales (1993) showed, however, that their estimates of  $l(f)$  were in contradiction with effective medium theory. The two limits of the dispersion curve,  $c_0$  and  $c_\infty$ , given by equations (III.2.28) and (III.2.29) directly applied on the sonic well logs were not identical to the same limits but now as predicted by the estimated  $l(f)$  and the Kramers-Krönig relations, thereby violating causality. Errors in the estimation of  $l(f)$  were attributed by Scales (1993) to a possible lack of low-frequency information in seismic data.

If the wave localization theory satisfies the Kramers-Krönig relations then relations between  $c_0$  and  $c_\infty$  as predicted by equation (III.2.75) should correspond independently to a) expressions (III.2.55), (III.2.56) and (III.2.58) for the effective wavenumber in these limits and to b) equations (III.2.28) and (III.2.30) given by the effective medium theory. It has already been shown



that equations (III.2.55), (III.2.56) and (III.2.58) do confirm equations (III.2.28) and (III.2.30) for Gaussian, exponential and Von Kármán autocorrelation functions. Moreover, insertion of equations (III.2.48) and (III.2.60) in expression (III.2.75) and performing the integration yields for both cases

$$c_{\infty} = c_0 \left[ 1 - \frac{1}{8} \langle \sigma_{\kappa}^2 \rangle \right]^{-1} \quad (\text{III.2.76})$$

for vertical incidence. Thus, at least for exponential and Gaussian random media, wave localization theory in combination with the Kramers-Krönig relations and effective medium theory do yield the same limits for the dispersion curve up to the second order. Unfortunately, again, no explicit analytical results could be obtained for the Von Kármán function.

Nevertheless, this does indicate that the overshoot in the relative dispersion curve for Gaussian media (Figure III.2.1) is required for the traversing pulse to remain causal.

**Absence of localization** In the introduction it was mentioned that no wave localization occurs for a finite number of frequencies and a set of realizations with zero probability, which are mainly related to periodic layering and the occurrence of resonances. Before showing that the obtained equations for the localization length, i.e. (III.2.48), (III.2.60) and (III.2.67), are in agreement with this statement, let us first notice that the localization length is inversely proportional to the variance of the characteristic magnitude of the heterogeneities. This indicates that naturally no localization of energy can occur in the absence of heterogeneities. However, it does also indicate that an infinitesimal small amount of randomness suffices to cause wave localization in 1-D media, which is in agreement with the scaling theory of localization of Abrahams et al. (1979).

Besides on the amount of randomness present, the localization length does also directly depend on the ratio  $k'_0 a$  and on the characteristic scale-length  $a$  of the heterogeneities itself. These two variables have to be seen as two independent quantities and can be analyzed separately. However, for fixed  $a$ , the ratio  $k'_0 a$  depends only on the frequency  $f$  or wavelength considered. Let us first examine the influence of  $k'_0 a$  and  $f$  on the localization length.

It was shown that independent of the type of correlation function used, i.e. Gaussian, exponential or Von Kármán, no wave localization occurs in the low frequency domain (long wavelength limit). This is due to the occurrence of Rayleigh scattering and the homogenization of the medium in this limit. Moreover, for smooth models, i.e. for Gaussian and Von Kármán functions with  $\nu > 1/2$ , no localization occurs either for high frequencies (geometric limit), since no discontinuities are present for smooth models in this limit. Therefore, for the 3 types of autocorrelation functions used, only 2 frequencies form an exception to the statement that wave localization always occurs in 1-D media if an infinitesimal small amount of randomness is present.

The localization length is not only related to  $k'_0 a$ , but it is also directly proportional to the typical scale-length  $a$  of the heterogeneities. Hence, no localization occurs either if  $a$  tends to infinity. Infinite typical scale-lengths (for finite wavelengths) are associated with periodic media for which it is known that no wave localization can occur due to perfect constructive and destructive interferences, i.e. certain frequencies are not attenuated at all, whereas others are immediately attenuated. However, the probability of a periodic layering (realization) to occur is zero.

$L$	$z_{source}$	$c_0$	$\rho_0$	$\Delta z$	$\Delta t$	$f_{p,Ricker}$	$a$
4000 m	4000 m	4000 m/s	2.2 g cm <sup>-3</sup>	1 m	1 ms	150 Hz	10 m

Table III.2.1: Medium parameters with  $L$ : thickness of random layer,  $z_{source}$ : source depth,  $c_0$ : average velocity,  $\rho_0$ : density,  $\Delta z$ : discrete layer thickness,  $\Delta t$ : sample time,  $f_{p,Ricker}$ : peak frequency of Ricker source wavelet.

### III.2.3 Numerical simulations

#### III.2.3.1 Strategy and measure method

To check to what degree theory and practice agree, numerical simulations have to be done. Therefore, with this aim in mind, measurements of  $l(f)$ ,  $Q^{-1}(f)$ ,  $\delta c(f)$  and  $\delta U(f)$  are made and compared with their theoretical predictions over a large range of  $k'_0 a$  and for media described by different autocorrelation functions. Although such comparisons have been made before, either measurements were averaged over a large number of realizations (Sheng et al., 1986a) or only highly fluctuating estimates could be obtained (White et al., 1990; Shapiro et al., 1994; Shapiro and Hubral, 1999) which are not as convincing as they could be. In this section, I will present a method which is able to produce accurate estimates of the desired quantities with only a limited number of fluctuations for a single realization of the medium by means of the wavelet transform.

In this paper, only the predictions of wave localization theory for a pulse consisting of plane-waves which traverses a random medium has been described. As a consequence, only the transmission problem will be dealt with in the numerical simulations. Hence, a source is placed in the homogeneous halfspace beneath the random part of the medium and the first arriving energy is recorded at the surface for 2 angles of incidence, namely 0° and 30°, respectively. This will be done for 4 independent realizations each of 3 different autocorrelation functions with a typical scale-length  $a$  of the heterogeneities equaling 10 m and relative standard deviations  $\sigma_{\kappa}$  of 15 % and 30 %. The relative standard deviations of  $\kappa$  correspond to relative standard deviations of 7.5 % and 15 %, respectively, of the velocity. Table III.2.1 displays the remaining constant parameters of the medium.

Some additional remarks have to be made. An identical  $c_0$  is used in both the random part of the medium and in the matching halfspace underneath it. No free surface is used to prevent back-scattering of energy from it. Finally, fluctuations are slightly damped at the edges to prevent energy from not entering the random medium. Moreover, a point source is used instead of plane waves using a so-called Ricker wavelet (second derivative of a Gaussian) as source wavelet. Simulations are done with an acoustic version of the method of Dietrich (1988) in which the generalized reflection and transmission matrix method of Kennett (1974) and the discrete wavenumber summation method of Bouchon (1981) have been combined.

To analyze the attenuation and dispersion of the primary (first arrival), traces will be transformed to the frequency domain. However, this is not done using a Fourier transform, but a wavelet transform. The inconvenience of the Fourier transform is that an optimal window length centered around the primary has to be chosen, which may be problematic. On the one hand, results obtained by using long analysis windows will be biased due to the presence of later arriving energy. On the other hand, behavior at low frequencies cannot be accurately measured using short window lengths. To circumvent this problem, use will be made of a wavelet transform as it analyzes the low frequency content of data with long windows and the high frequencies with short ones. Hence, it forms a natural bridge between frequency and time analyses of non-stationary data.



The wavelet transform is most easily understood by drawing an analogy with the Short Term Fourier Transform which is simply a repeated Fourier transform using a sliding window. It is expressed as

$$\mathcal{F}[s](f, \tau) = \int s(t)g^*(t - \tau)e^{2\pi ift} dt \quad (\text{III.2.77})$$

with  $s(t)$  the recorded time signal and  $g^*(t)$  the complex conjugate of the employed window  $g(t)$  which has a fixed length. The continuous wavelet transform, on the other hand, is given by

$$\mathcal{W}[s](a_w, \tau) = a_w^{-1/2} \int s(t)w^*\left(\frac{t - \tau}{a_w}\right) dt, \quad (\text{III.2.78})$$

where  $w(\cdot)$  represents the shifted and scaled mother wavelet and  $a_w$  denotes scale, which is inversely proportional to frequency. The normalization factor  $a_w^{-1/2}$  ensures that wavelets have identical energy for any scale. The mother wavelet represents that type of wavelet used for signal analysis. A comparison of equations (III.2.77) and (III.2.78) directly shows that in the wavelet transform, window length is not kept fixed, but does display the desired trade-off: it is dilated ( $a_w > 1$ ) or contracted ( $a_w < 1$ ) by changing the value of  $a_w$ . For more background, the interested reader is referred to the review articles of Rioul and Vetterli (1991) and Kumar and Foufoula-Georgiou (1997).

In this application, use will be made of the classical Morlet mother wavelet to analyze data. It is given by (with  $t' = t/\Delta t$ )

$$w(t) = \frac{\pi^{1/4}}{c_w^{1/2}} e^{-i\pi t'} e^{\frac{1}{2}(\pi t'/c_w)^2}. \quad (\text{III.2.79})$$

Hence, it is simply a real cosine plus complex sine weighted with a Gaussian with variance  $c_w/\pi$  to produce a local support. Expressing it in this form has the convenience that  $a_w = 1$  equals the Nyquist frequency. Consequently increasing  $a_w$  yields lower frequencies, i.e.  $f = 1/(2\Delta t a_w)$ . The constant  $c_w$  has been put to 5.5 in this study.

Like any complex number  $\mathcal{W}[s](a_w, \tau)$  can be expressed in the polar form  $r(a_w, \tau) \exp\{i\theta(a_w, \tau)\}$ , where  $r$  depends on the source wavelet and its radiation pattern and attenuation and  $\theta$  on the initial phase of the source wavelet and dispersion.

To prevent that source correction terms have to be estimated for obtaining accurate measurements of attenuation and dispersion, a totally homogeneous medium was used as a reference. Nevertheless, this medium had the same setup as the other simulations, i.e. same number of layers, source and receiver positions, etc. Using such a reference has several advantages. Not only are source correction terms not needed, but in addition results will not be biased by numerical dispersion. Moreover, spherical divergence does not have to be corrected for either, thereby allowing for simulations using more complicated models.

Determination of the desired frequency dependent apparent attenuation and dispersion is now straightforward. Let  $r_i^{max}(a_w, \tau_i^{max})$  denote the maximum amplitude of the wavelet coefficients for a given  $a_w$  or frequency measured at time  $\tau_i^{max}$ . In general,  $r_i^{max}$  will represent the peak amplitude of the primary for a specific frequency and  $\tau_i^{max}$  its arrival time, unless scattering has become too strong, i.e. the main energy arrives after the primary. Moreover, let  $i = 1$  define a realization of some heterogeneous model and  $i = 0$  the homogeneous case. The localization length can now be estimated using equation (III.2.39) with  $|T| = r_1^{max}/r_0^{max}$  and  $z = L/\cos(\phi_i)$ , i.e.

$$l(a_w) = \frac{-L}{\cos(\phi_i) \ln(r_1^{max}/r_0^{max})}. \quad (\text{III.2.80})$$

In addition,  $Q^{-1}$  is again given by expression (III.2.44).

The dispersion  $\delta c$  can be obtained in a similar way using the difference of the arguments  $\theta_1^{max}$  and  $\theta_0^{max}$  which should equal

$$\delta\theta = \theta_1^{max} - \theta_0^{max} = (k_{z,1} - k_{z,0})L - \omega(\tau_1^{max} - \tau_0^{max}). \quad (\text{III.2.81})$$

The horizontal wavenumber is not involved as it was assumed to be constant. Moreover, as the vertical wavenumber  $k_z$  equals  $\psi$ , the previous expression yields

$$\delta c = \frac{-s_c}{1 + s_c} \text{ with } s_c = \frac{c_0 [\delta\theta + \omega(\tau_1^{max} - \tau_0^{max})]}{\omega L \cos(\phi_i)} \quad (\text{III.2.82})$$

by using equation (III.2.45). Finally, it should be noted that  $\delta\theta$  represents the unwrapped phase. Phase unwrapping is best done starting at the lowest frequencies, i.e. high values of  $a_w$ .

The corresponding equation for  $\delta U$  can be found by putting

$$\frac{\partial \delta\theta}{\partial \omega} = 0, \quad (\text{III.2.83})$$

yielding

$$\delta U = \frac{-s_u}{1 + s_u} \text{ with } s_u = \frac{(\tau_1^{max} - \tau_0^{max})c_0}{L \cos(\phi_i)} \quad (\text{III.2.84})$$

by using equations (III.2.81) and (III.2.46).

Measured quantities using equations (III.2.80), (III.2.44), (III.2.82) and (III.2.84) are compared with their theoretical counterparts, i.e. expressions (III.2.48), (III.2.50), (III.2.51) and (III.2.54) for Gaussian media, (III.2.60), (III.2.62), (III.2.63) and (III.2.65) for exponential media and (III.2.67), (III.2.69), (III.2.70) and (III.2.73) for Von Kármán media, respectively.

### III.2.3.2 Results

In the following the numerical simulations will be described and discussed. Figure III.2.2 shows an identical realization for the 3 types of autocorrelation functions used, namely Gaussian, exponential and Von Kármán for  $\nu = 0$ . It can be clearly seen that the Gaussian is the smoothest profile and this particular Von Kármán function the roughest one. The exponential function is somewhere in between these extremes. Moreover, although the typical scale-length of the heterogeneities is 10 m, the profiles display clearly fluctuations on any scale-length.

Media with particular autocorrelation functions are usually created in the wavenumber domain. First, a random sequence is created with a Gaussian distribution function and an autocorrelation function equaling a Dirac spike. Next, the sequence is transformed to the wavenumber domain, multiplied with the square root of the Fourier transform of the desired autocorrelation function and back-transformed to the space domain. The square root is needed to ensure that it is the autocorrelation function of the resulting sequence which has the desired form and not the created sequence.

Obviously, discrete representations of the desired media are employed as a finite layer thickness is used. Unfortunately, this makes that perfectly smooth models and media displaying fluctuations on any scale cannot be handled well. Moreover, due to such finite layer thicknesses, variances will be biased. The variances of discrete profiles should be inferior to their continuous counterparts (Frankel and Clayton, 1986). Theoretical predictions will, therefore, be slightly biased as this phenomenon is neglected in the numerical simulations and the used profiles.

Figure III.2.3 displays parts of the resulting traces for the realization shown in Figure III.2.2 for 2 angles of incidence ( $0^\circ$  and  $30^\circ$ ) and 2 different values of  $\sigma_\kappa$  (15 % and 30 %). Every



second trace represents the case of  $\phi_i = 30^\circ$ . The first two traces show the homogeneous case, the next 6 traces  $\sigma_\kappa = 15\%$  and the last 6 traces  $\sigma_\kappa = 30\%$ , for Gaussian, exponential and Von Kármán media, respectively. Notice, that the last 5 traces are plotted on a different scale. No time dependent amplitude corrections have been applied to clearly show the amount of later arriving energy.

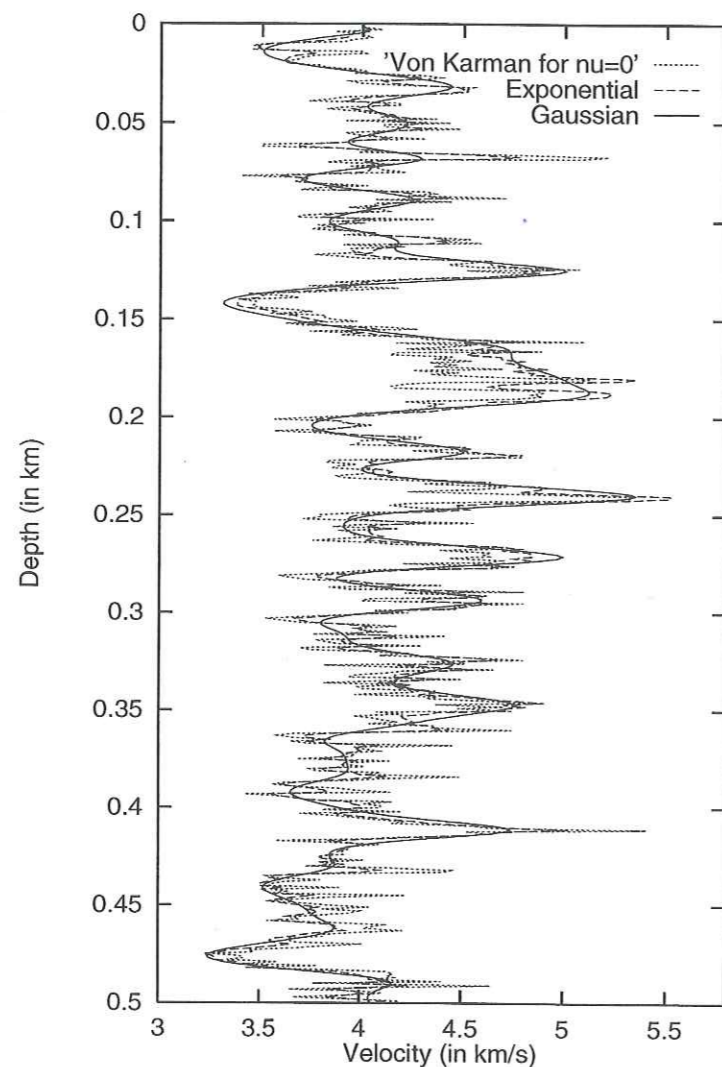


Figure III.2.2: Top parts of the used velocity profiles for a single realization of Gaussian, exponential and Von Kármán ( $\nu = 0$ ) autocorrelation functions.

Attenuation, phase distortions and dispersions of the primary purely due to scattering can be clearly detected by a simple comparison with the original source wavelet (as shown in the first 2 traces). The Gaussian correlation function for  $\sigma_\kappa = 30\%$  is a particularly interesting case as its primary for vertical incidence is only weakly attenuated (plotted on true scale), whereas the primary for  $\phi_i = 30^\circ$  is nearly completely attenuated (amplitudes multiplied by 20). This phenomenon is due to the strongly frequency and angle of incidence dependent attenuation factor for Gaussian media (see Figure III.2.1).

Figure III.2.4 displays the measured values for  $l$ ,  $Q^{-1}$ ,  $\delta c$  and  $\delta U$  for 4 different realizations each of the 3 types of autocorrelation functions used for both angles of incidence and a standard

deviation of 15%. Figure III.2.5 displays the same quantities for a standard deviation of 30% of the fluctuations. Moreover, the theoretical predictions have been shown in both figures. Again, it should be noted that  $\Gamma(0)$  used in the Von Kármán function has been replaced by a constant (13.0) chosen to display the general trend.

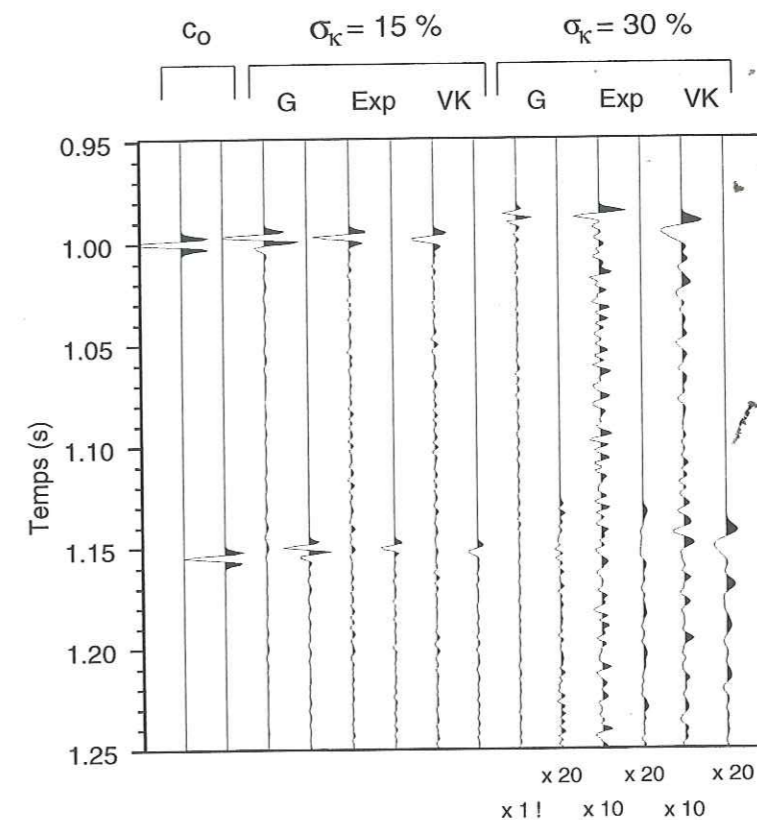


Figure III.2.3: Primaries for the random media shown in Figure III.2.2. Every second trace shows the case of  $\phi_i = 30^\circ$ . First 2 traces display the homogeneous case, next 6 traces  $\sigma_\kappa = 15\%$  and the last 6 traces  $\sigma_\kappa = 30\%$ . Notice different scale on last 5 traces. G stands for Gaussian, E for exponential and VK for Von Kármán autocorrelation functions.

A close examination of Figures III.2.4 and III.2.5 reveals that theory and practice agree well especially in the low to intermediate frequency range. For instance,  $l$  is proportional to  $f^{-2}$  at low frequencies and  $Q^{-1}$ ,  $\delta c$  and  $\delta U$  all tend to zero in this limit, thereby confirming the existence of the homogenization phenomenon. In addition,  $\delta c$  and  $\delta U$  tend always to the geometric velocity in the high frequency domain independent of the type of autocorrelation function used.

In the high frequency domain, however, some discrepancies exist, although the general trend is confirmed. Most remarkable are discrepancies for  $l(f)$  and  $\delta c$  as the localization length does not seem to display the  $e^{f^2}$ -dependence for Gaussian media and all theoretical predictions are systematically too low for  $\delta c$ . Due to the blocky character of the measurements of  $\delta U$ , it cannot be concluded if its theoretical predictions have also been underestimated. Nevertheless, it should be noted that part of such discrepancies are due to the fact that a finite layer thickness has been used as discussed above. Hence, profiles are not smooth and theoretical variances may have been underestimated.

Finally, it should be noted that few fluctuations exist in the estimates of the various quantities



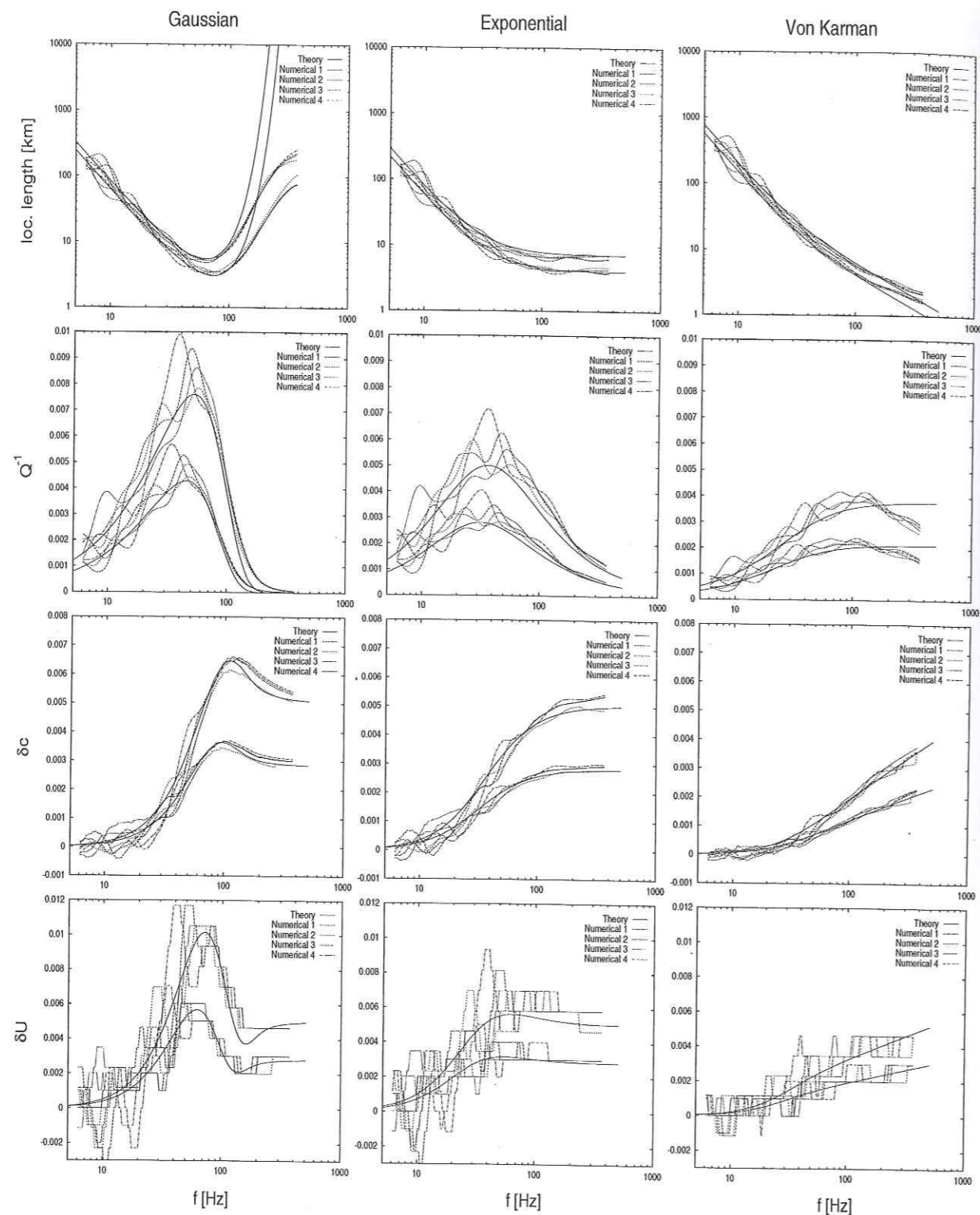


Figure III.2.4: Measured localization length, apparent scattering coefficient and dispersion for 4 realizations of the 3 types of used correlation functions for both angles of incidence for  $\sigma_\kappa = 15\%$ . Theoretical predictions are plotted too. The lowest theoretical curve always presents  $\phi_i = 0$  and the other one  $\phi_i = 30^\circ$ , except for the localization length where the order is inverted. The blocky character of  $\delta U$  is due to the used sampling rate (1 ms). Notice, however, that the values of  $\delta c$  have been calculated with a precision higher than the sampling rate.

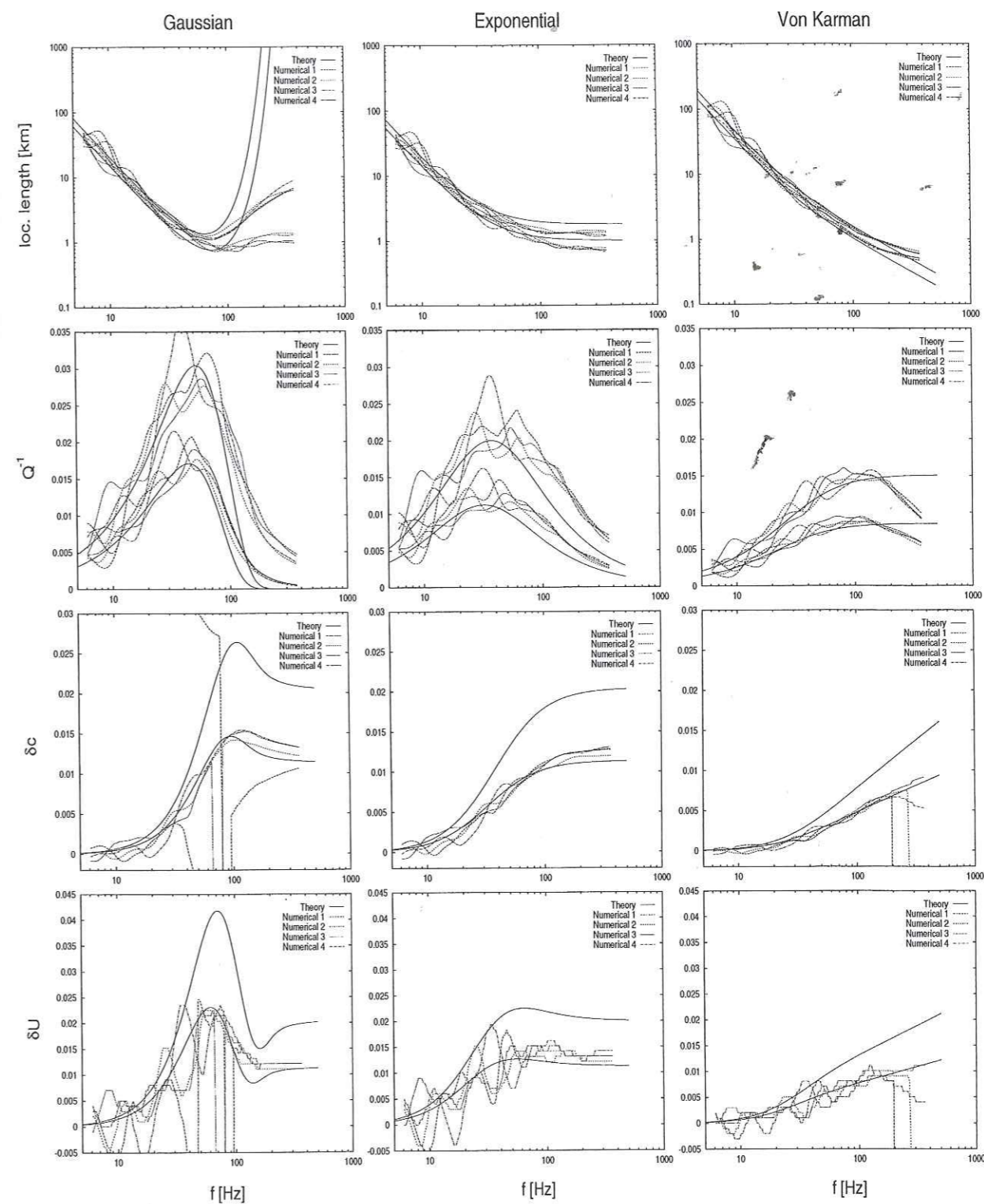


Figure III.2.5: Identical to Figure III.2.4 only now for  $\sigma_\kappa = 30\%$ . Group and phase velocity measurements have not been displayed for  $\phi_i = 30^\circ$  as phase unwrapping became completely unstable and the primary could not be automatically detected (main energy arrived after the primary). Notice changed vertical scales with respect to Figure III.2.4.



and that they seem to diminish with frequency (in particular for  $\delta c$  and  $l$ , the self-averaging quantities). Estimates of  $Q^{-1}$  and  $\delta U$  are more strongly fluctuating than those for  $l$  and  $\delta c$  as these are not self-averaging quantities.

### III.2.4 Discussion

Before discussing the obtained results in connection to other known results and before putting 1-D wave localization theory in a wider perspective, let me shed some more light on self-averaging and its implications as it is one of the most important features of the above described theory. Self-averaging of certain quantities allows for applying a statistical approach to the wave scattering problem, thereby significantly simplifying mathematics, while at the same time producing real observables which can be measured and e.g. inverted for. The main characteristic of self-averaging variables is that they display a Gaussian distribution centered around their expected values, at least in some macroscopic limit (Gredeskul and Freilikher, 1990). Moreover, its variance is inversely proportional to the thickness of the chaotic medium for 1-D systems. Thus, in the limit of an infinitely thick medium, the variance will reduce to zero resulting in a probability distribution resembling a Dirac positioned at the expected value. That is, the expected value has unit probability in this limit for (almost) any realization of the medium, which is the strict requirement for self-averaging. Finally, it should be noted that convergence of self-averaging quantities is assumed to be fast, once the thickness of the random medium exceeds the localization length (Souillard, 1987; Van Tiggelen, 1999).

As can be seen in Figures III.2.4 and III.2.5, fluctuations of the localization length and the relative velocity dispersion are decreasing with increasing frequency. Two features will contribute to this phenomenon. Firstly, a Ricker wavelet has been used with a peak frequency of 150 Hz and a limited bandwidth. Such a source wavelet is not very energetic at low frequencies, thereby rendering an exact validation more difficult. On the other hand, fluctuations are nearly absent for high frequencies, while it does not have much high frequencies either. Secondly, fluctuations of self-averaging quantities should increase if  $kL$  drops, which is the case for decreasing frequencies.

Using a wavelet transform for signal analysis has tremendously reduced fluctuations, as can be concluded from a comparison with results of e.g. Sheng et al. (1986a), White et al. (1990), Shapiro et al. (1994) and Shapiro and Hubral (1999). These reduced fluctuations are probably due to 2 characteristics of the wavelet transform. Firstly, it uses an optimal analysis window length for each considered frequency as later arriving energy produces a minimal distortion of results and energy of low frequencies is not underestimated due to too short window lengths. Secondly, estimation quality of self-averaging quantities can be increased by averaging over some other variable over a mesoscopic length-scale like e.g. the pulse length (Gredeskul and Freilikher, 1990). As the wavelet transform analyzes data over several periods, such an averaging is explicitly applied, thereby increasing quality of estimates. Interestingly, the wavelet transform has been concised by Morlet et al. (1982) in the context of sampling theory and wave scattering, albeit due to periodic layering.

Therefore, the use of the wavelet transform has allowed to obtain clear indications of the validity range of the developed theory. Theory and numerical simulations closely agree up to 100 Hz for both angles of incidence and the 3 types of considered autocorrelation functions, i.e.  $k_0 a \approx 1.5$  and  $k_0 L \approx 600$ . Minor discrepancies can be seen above this value. Two reasons may explain these deviations. Firstly, the above-described multiple scattering theory of Shapiro and Zien (1993) is a second order theory. Hence, results may become less accurate for large perturbations of parameters. Secondly, numerical simulations have been done using discrete profiles with exactly the desired standard deviations. However, discrete sequences should have

standard deviations slightly inferior to their continuous equivalents. Hence, theoretical standard deviations may have been underestimated. Indeed, slightly increasing the theoretical  $\sigma_\kappa$  does diminish the existing gaps, but is not able to explain all deviations.

Nevertheless, the validity range does cover the most interesting frequencies and wavelengths. For instance, theory is accurate in the domain of Mie scattering, where wavelengths and typical scale-lengths of heterogeneities are comparable. Therefore, theoretical predictions of the localization length hold true in a much broader frequency range than those published by Burridge et al. (1994) and Clouet and Fouque (1994), which are only valid in the low-frequency domain (Rayleigh scattering). In addition, expressions are given for determining directly the frequency dependent dispersion. Nevertheless, predictions should coincide in the low frequency domain.

Some of the phenomena which are shown to exist had already been predicted before. Sheng et al. (1986a) showed the  $f^{-2}$ -dependence of the localization length in the low frequency domain for some discontinuous and a smooth type of random media. Moreover, they showed the possible approach to a constant or divergence of  $l$  at high frequencies and the existence of a minimum localization length, which largely exceeds the correlation length of the inhomogeneities. However, the last phenomenon is not necessarily true for fractal-like media and, in addition, no explicit simple formulae were obtained. Shapiro and Zien (1993) were the first to obtain a simple expression for the localization length in exponential media and Shapiro et al. (1994) gave the corresponding expression for the velocity dispersion. Beltzer (1989) obtained similar expressions for  $\delta c$  for Gaussian and exponential media using the Kramers-Krönig relations and the theoretical predictions of the apparent attenuation due to scattering of Wu (1982) resulting from a first order theory. In particular, the existence of the overshoot of the phase velocity in Gaussian media had been predicted by Beltzer (1989). However, first order approximations are not valid for long propagation distances in strongly scattering media. Moreover, accuracy of predictions had not been confirmed by numerical simulations.

None, however, has given explicit expressions for the localization length and phase velocity for large propagation distances in strongly scattering media described by Gaussian and Von Kármán autocorrelation functions. In addition, numerical simulations have confirmed theoretical predictions and it has been shown that wave localization theory agrees with effective medium theory and the Kramers-Krönig relations, at least for certain types of autocorrelation functions.

In this paper, wave localization has been primarily presented to increase our understanding of strong multiple wave scattering. However, to reduce mathematical complications and to emphasize some implications for multiple scattering of waves, several simplifications have been made and the theory has been described in its simplest form. On the other hand, the full range of implications of wave localization is not yet known. In particular, the question if wave localization occurs in real 3-D media is unresolved, although experiments with light of Wiersma et al. (1997) do point in this direction.

Let me briefly point to some known results. Two simplifications can easily be dropped. Firstly, in Shapiro et al. (1994) equations are obtained for the localization length and the vertical phase velocity for random density profiles. Fluctuations in density will increase the apparent attenuation due to scattering because of increased impedance contrasts. However, as it is often assumed that velocity and density fluctuations are correlated in the real Earth (see e.g. Sato and Fehler (1998)), it might be interesting to express the self-averaging quantities directly in terms of impedance fluctuations instead of velocity, compressibility and density fluctuations.

Secondly, dissipation of energy or intrinsic attenuation is often thought to exist independent of apparent attenuation due to scattering (Herraiz and Espinosa, 1987; Beltzer, 1989). This implies that besides the Lyapunov exponent  $\gamma_{sc}$  for attenuation due to wave scattering, a second 'Lyapunov exponent'  $\gamma_{in}$  for intrinsic scattering can be introduced and that the 'total



Lyapunov exponent  $\gamma_{tot}$  is given by a simple summation of  $\gamma_{sc}$  and  $\gamma_{in}$ , thereby increasing again attenuation.

Further complications arise if non-stationary fluctuations, elastic waves or higher dimensions are considered. Non-stationary fluctuations may even prevent the localization of waves. For instance, the existence of the Lyapunov exponent is not guaranteed if fluctuations decrease faster than a certain limit with increasing depth (Souillard, 1987) simply due to leakage into the deeper parts of the model which become increasingly more homogeneous. However, smooth macroscopic variations in the background velocity and density superposed on stationary fluctuation functions do not influence the attenuation, but do influence phase velocities. Fortunately, the resulting phase shifts can easily be corrected for by raytracing (Burrige et al., 1988; Shapiro and Hubral, 1999). Furthermore, the low frequency approximation of Burrige et al. (1994) does allow for macroscopic variations of fluctuations as only the integral over the autocorrelation function is involved. On the other hand, small non-stationarities can probably also be handled in the low frequency domain in the above-described theory by averaging the autocorrelation function over depth, equation (III.2.43).

Two theories have been developed describing wave localization in elastic 1-D media. In such media, a second Lyapunov exponent occurs due to the co-existence of P and SV waves and their coupling. Shapiro et al. (1996) extended the method of Shapiro and Zien (1993) and Shapiro et al. (1994) to the elastic domain and showed that localization lengths exist for both P and SV waves. However, their results are only valid for propagation distances which are of the same order as the localization lengths. Phenomena occurring for larger propagation distances have been described by Kohler et al. (1996). In their low frequency approximation, only a single localization length remains which is the reciprocal of the smallest Lyapunov exponent. The reciprocal of the largest Lyapunov exponent denotes a scale-length beyond which P and SV energies are in equilibrium.

Let me finally say some words concerning wave localization in higher dimensions and for non-perfect layering. Kohler et al. (1999) showed that characteristics of wave localization remain robust if small non-random 3-D variations in the background are superposed on slowly undulating layering, i.e. layers are only locally plane. Moreover, first order effects occur only for non-vertical incidence. That is, the resulting localization length can be determined as if a perfect 1-D medium exists for vertical incidence.

The study of Sheng and Zhang (1995) indicated that the vertical localization length changes but does exist if a small amount of isotropic scatterers are incorporated in a layered medium. In this case, wave localization characteristics are somewhere in between those for purely 1- and 2-D media. Finally, Kirkpatrick (1985) showed that wave localization exists in 2-D media except at the two frequency extremes and also in 3-D media for some frequency range in between intermediate and high frequencies. These last two results agree with the scaling theory of localization (Abrahams et al., 1979) which predicted existence of wave localization for 1- and 2-D media and possibly in 3-D. See also Souillard (1987). However, it should be noted that in 2- and 3-D models, wave localization theory may not be the best adapted theory to study multiple scattering as the localization length may become very large unless the randomness is extremely high. Diffusion and radiative transfer approaches may be more suited.

### III.2.5 Conclusions

Multiple wave scattering is a complex phenomenon. Using a statistical description of the chaotic part of the medium can reduce the involved mathematics considerably. However, such an approach does usually not result in real observables unless self-averaging quantities are considered.

Wave localization permits such a statistical approach and produces, in addition, two self-averaging quantities, namely the Lyapunov exponent and the effective vertical wavenumber. The frequency dependent localization length, i.e. penetration depth, the inverse quality factor and the dispersion (phase and group velocities) all purely due to wave scattering can be calculated with these two observables. These variables have been calculated using the second order perturbation expansion of Shapiro and Zien (1993) for several chaotic media described by different autocorrelation functions. Used media range from very smooth to fractal-like.

Analysis of results shows that the localization length is always proportional to  $f^{-2}$  in the low frequency domain independent of the type of medium. This is characteristic for Rayleigh scattering in 1-D media. In the high frequency domain, it either diverges or converges to zero. Only for exponential autocorrelation functions associated with piecewise continuous models does it approach a constant. Hence, minimal localization lengths as inferred by Sheng et al. (1986a) do not necessarily exist for very rough and fractal-like media.

Moreover, it is confirmed that waves become always localized in 1-D media if a small amount of randomness is present, save for a countable number of exceptions. These exceptions include the long wavelength limit in which the medium becomes effectively homogeneous as implied by effective medium theory and in smooth media in the short wavelength or geometric limit in which no heterogeneities (discontinuities) exist either.

Inspection of the inverse quality factor indicated that Mie scattering yields the most effective apparent attenuation due to scattering. On the other hand, fractal-like media form an exception as the reciprocal of the quality factor rises monotonically with frequency and approaches a constant, making it effectively constant over a large range of frequencies. Therefore, no particular ratio  $ka$  is favored for such media.

The last quantity which has been examined is the frequency dependent dispersion due to scattering. It has been shown that predictions of wave localization theory for the apparent attenuation and the phase velocities are consistent with the Kramers-Krönig relations. Therefore, the theory respects causality. Moreover, it is confirmed that anisotropy exists due to the existence of fine-scale layering.

Both the phase and group velocities converge always independent of the type of medium to the same limits for both high and low frequencies. That is, they approach the effective medium velocity in the long wavelength limit and the geometric velocity in the high frequency limit. However, for intermediate frequencies, large differences occur. For instance, unlike rough models where the phase velocity is monotonically rising with frequency, in smooth models an overshoot occurs. This implies that the geometric velocity is not the highest encountered phase velocity, but an intermediate range of frequencies exist which will arrive before the highest frequencies in smooth models. This particular phenomenon occurs to prevent that causality is violated. In addition, the exact shape of the group velocity curves and the number of Airy phases also strongly depend on the type of autocorrelation function. Moreover, Airy phases often occur in the domain of Mie scattering. Hence, they cannot always be associated to strong arrivals; contrary to the common assumption. Furthermore, there seems to exist some direct relation between the group velocity and the inverse quality factor.

Finally, numerical simulations have clearly confirmed the theoretical predictions as theory and practice coincide for values of  $ka$  up to 1.5, thereby including both Rayleigh and Mie scattering, and for very large propagation distances, i.e.  $kL \approx 600$ . In addition, above this value, the general trend is affirmed. Discrepancies for the higher frequencies are probably due to the use of discretized models for the numerical simulations and the fact that a second order theory has been applied. In contrast to previous simulations, highly accurate estimates of all parameters could be made by means of the wavelet transform which damped out nearly all



perturbing fluctuations.

### III.2.6 Acknowledgements

The author would like to thank Michel Bouchon for discussions concerning the implications of the theoretical predictions and Michel Dietrich for providing his numerical code. Moreover, he is indebted to Sergei Shapiro who was always willing to discuss and to explain details of his theoretical development and wave localisation theory in general.

## III.3 Essais d'inversions des caractéristiques statistiques d'un milieu

### Résumé

La diffusion forte est un phénomène compliqué. Toutefois, la théorie de la localisation d'onde décrit bien l'atténuation apparente et la dispersion d'une impulsion traversant un milieu stratifié très hétérogène. Dans cette étude, nous appliquons cette théorie à l'inférence statistique afin de déterminer les paramètres statistiques décrivant le milieu hétérogène. Nous résolvons le problème inverse en mesurant la dépendance fréquentielle des rapports spectraux de plusieurs réflexions enregistrées à des temps différents.

Dans un premier temps, nous faisons des tests sur des sismogrammes synthétiques. Ces tests nous montrent que les rapports spectraux mesurés sont contaminés par des réflexions multiples tardives. Ces réflexions ne sont pas anticipées par la théorie de la localisation d'onde, puisqu'il s'agit d'une théorie en transmission. Pour la diffusion faible, l'influence des réflexions multiples tardives est probablement négligeable. Par contre, pour la diffusion forte, leur effet immédiat est d'augmenter l'énergie totale enregistrée et ainsi de réduire l'atténuation apparente. Par conséquent, la magnitude typique des hétérogénéités et la réciproque du facteur de qualité de l'atténuation intrinsèque seront sous-estimées. Nous pouvons estimer la dépendance fréquentielle de leur influence en la mesurant sur des parties de traces ne contenant pas de réflexions majeures. Mais malheureusement, leur magnitude absolue est inconnue à cause des corrections imparfaites de la divergence sphérique. Par conséquent, il faut rechercher des méthodes théoriques prédisant leur influence. D'autre part, il n'est pas clair si leur influence n'est pas négligeable dans des données réelles où d'autres facteurs contaminant les rapports spectraux mesurés peuvent dominer (comme le bruit et l'énergie des signaux voisins).

Ensuite, nous appliquons la méthode d'inversion sur des données de l'expérience Ecos 2.5-D afin d'estimer la magnitude  $\sigma_c$  et la dimension  $a$  typiques des fluctuations de vitesses et le facteur de qualité de l'atténuation intrinsèque  $Q_{in}$ . Les résultats nous montrent que la croûte supérieure est caractérisée par  $a \sim 15$  m,  $\sigma_c \leq 7.5$  % et  $Q_{in} \geq 350$  et la croûte inférieure par  $a \sim 40$  à  $50$  m,  $\sigma_c \leq 5.5$  % et  $Q_{in} \geq 1600$ . Ceci nous montre que dans cette région la croûte supérieure est plus hétérogène et plus atténuante que la croûte inférieure et que les dimensions typiques  $y$  sont plus petites.

### III.3.1 Introduction

Nous avons vu dans la section précédente que la théorie de la localisation d'onde prédit bien l'atténuation et la dispersion d'une onde traversant un milieu hétérogène stratifié. Dans la suite, nous voulons examiner si la théorie peut être utilisée afin d'estimer les caractéristiques statistiques d'un milieu hétérogène en mesurant l'atténuation d'une telle onde en fonction de la fréquence.

Afin de parvenir à ce but, nous ferons d'abord appel à des simulations numériques et finalement à des données réelles. Nous voulons surtout savoir sous quelles conditions la théorie peut être appliquée et quels sont les problèmes rencontrés. Nous ne nous intéressons pas au problème en transmission, mais à celui de la réflexion. En effet, si nous réussissons à appliquer la théorie sur des données en réflexion, la méthode deviendra très intéressante pour les expériences de sismique réflexion profonde, car les distances parcourues  $y$  sont vite de l'ordre de plusieurs dizaines de kilomètres. Ces grandes distances sont nécessaires pour que le mécanisme auto-moyennant puisse se manifester. De plus, des données en réflexion sont plus faciles à acquérir que des données en transmission.

Nous commençons par décrire brièvement les méthodes d'inversion appliquées et puis les facteurs qui influencent les amplitudes enregistrées. Ensuite nous ferons des inversions avec des



données synthétiques et enfin avec des données réelles.

Dans la suite, nous nous limitons à l'incidence verticale pour deux raisons principales. Premièrement, une impulsion à incidence normale dans un milieu stratifié ne subit pas de conversions P-S et/ou inversement et il n'y a ni des ondes inhomogènes ni des refractions. Par conséquent, nous pouvons faire appel à une méthode acoustique simple (si le milieu est stratifié). Deuxièmement, Kohler et al. (1999) ont démontré que, pour une onde acoustique à incidence normale, les longueurs de localisation d'un milieu stratifié et d'un milieu n'ayant que de faibles perturbations dans les vitesses moyennes et dans la stratification sont identiques. Par conséquent, si nous nous limitons à des données de sismique réflexion verticale, la théorie devrait donner de bonnes prédictions.

### III.3.2 Les facteurs influençant l'atténuation apparente

Si le milieu traversé est linéaire, passif et causal, nous pouvons représenter le signal enregistré par une série de convolutions en temps ou par des multiplications en fréquence. Ceci nous permet d'analyser séparément les facteurs influençant l'atténuation apparente et enfin de combiner leurs effets.

Par exemple, O'Doherty et Anstey (1971) nous donnent les termes suivants: la divergence sphérique  $\hat{r}_{ds}$ , l'atténuation à cause des hétérogénéités présentes et les diffractions résultantes ('scattering' en anglais)  $\hat{r}_{sc}$  et l'absorption intrinsèque  $\hat{r}_{in}$ . De plus, la directivité de la source et l'ondelette de la source  $\hat{r}_s$  jouent un rôle. Pour le problème en réflexion, il nous faut encore ajouter deux autres termes. D'abord, un terme décrivant l'influence des réflexions multiples tardives  $\hat{r}_{rmt}$ . De plus, les mesures de l'atténuation de l'onde seront faites sur une réflexion profonde. Donc, il nous faut aussi inclure un terme de correction pour le coefficient de réflexion effectif  $\hat{r}_{cr}$ .

Les termes  $\hat{r}_{sc}$ ,  $\hat{r}_{in}$  et  $\hat{r}_{rmt}$  dépendent à la fois de la fréquence  $f$  et du temps  $t$  auquel on fait la mesure. Ceci s'exprime par  $\hat{r}(f|t)$ . Cependant, au premier ordre et pour des angles d'incidence constants, les expressions  $\hat{r}_{ds}$  et  $\hat{r}_{cr}$  ne dépendent que du temps de mesure. Donc, ces paramètres-là sont des constantes en fréquence:  $\hat{r}(f|t) = \hat{r}(t)$ . Finalement,  $\hat{r}_s$  ne dépend pas du temps de mesure:  $\hat{r}_s = \hat{r}_s(f)$ .

Par conséquent, l'amplitude d'une onde en transmission  $\hat{r}_{tot,t}$  enregistrée au temps  $t$  en fonction de la fréquence  $f$  peut être exprimée comme

$$\hat{r}_{tot,t}(f|t) = \hat{r}_{ds}(t) \cdot \hat{r}_{sc}(f|t) \cdot \hat{r}_{in}(f|t) \cdot \hat{r}_s(f) \quad (III.3.1)$$

et l'amplitude d'une onde réfléchie  $\hat{r}_{tot,r}$  par

$$\hat{r}_{tot,r}(f|t) = \hat{r}_{cr}(t) \cdot \hat{r}_{ds}(t) \cdot \hat{r}_{sc}(f|t) \cdot \hat{r}_{in}(f|t) \cdot \hat{r}_s(f) + \hat{r}_{rmt}(f|t) \cdot \hat{r}_s(f). \quad (III.3.2)$$

Maintenant que nous connaissons les facteurs qui influencent les amplitudes, nous pouvons essayer d'en trouver les approximations. Dans la suite, nous les traitons une par une (le tableau III.3.1 nous donne les abréviations des facteurs différentes).

**Divergence sphérique:**  $\hat{r}_{ds}$  Il est bien connu que dans un milieu homogène la décroissance des amplitudes est linéaire avec le temps. Malheureusement, cette relation simple n'est plus valable si les vitesses changent avec la profondeur. Mais, en général, nous pouvons corriger de la divergence sphérique de façon analytique ou empirique, car le phénomène est bien compris. Par exemple, Newman (1973) nous donne une expression correctrice pour la divergence sphérique pour des milieux stratifiés. Sinon, nous pouvons toujours utiliser des méthodes empiriques (Claerbout, 1985) soit en corrigeant en  $t^2$ , soit en normalisant l'énergie dans une fenêtre d'analyse. En tous

Variable	Signification
$\hat{r}_{ds}$	Divergence sphérique
$\hat{r}_{sc}$	Effets des hétérogénéités
$\hat{r}_{in}$	Absorption intrinsèque
$\hat{r}_s$	Effets de la source
$\hat{r}_{cr}$	Coefficient de réflexion
$\hat{r}_{rmt}$	Réflexions multiples tardives

TAB. III.3.1 - Facteurs influençant les amplitudes et leurs abréviations

cas,  $\hat{r}_{ds}$  ne dépend pas de la fréquence et peut être approximé par une constante. Par conséquent, nous l'incluons de cette façon dans une inversion.

**Effets des hétérogénéités:**  $\hat{r}_{sc}$  Dans la section III.2 nous avons montré que la théorie de la localisation d'onde décrit bien les effets des hétérogénéités sur l'atténuation apparente et la dispersion d'une onde traversant un milieu aléatoire. Donc, de nouveau, nous utilisons cette théorie. Par conséquent,  $\hat{r}_{sc}$  est égal à (expression (III.2.31))

$$\hat{r}_{sc}(f|t) = |T| = \exp[-\gamma_{sc}z(t)] \quad (III.3.3)$$

avec  $\gamma_{sc}$  l'exposant de Lyapunov. Cet exposant est la réciproque de la longueur de localisation donnée par les expressions (III.2.48), (III.2.60) ou (III.2.67) dépendant du type de fonction d'autocorrélation spatiale. La variable  $z(t)$  est obtenue par une simple conversion de temps en profondeur (bien entendu à condition que nous connaissions le profil des vitesses avec la profondeur: une première estimation de ce profil peut être obtenue par une analyse de la courbure apparente des réflecteurs). Finalement,  $z(t)$  représente la distance parcourue qui est égale à deux fois la profondeur du réflecteur analysé.

Les expressions (III.2.48), (III.2.60) et (III.2.67) donnent de bons résultats si le milieu parcouru est caractérisé par une seule fonction d'autocorrélation. Par contre, quelle expression faut-il utiliser si le milieu est divisé en plusieurs morceaux indépendants?

Cette expression est facile à obtenir en notant que les distributions statistiques dans les couches indépendantes sont non-corrélées. Dans ce cas, nous pouvons obtenir une approximation de l'exposant de Lyapunov moyen  $\langle \gamma_{sc} \rangle$  par

$$\langle \gamma_{sc} \rangle = \frac{\sum_i \delta L_i \gamma_{sc,i}}{\sum_i \delta L_i} \quad (III.3.4)$$

avec  $\gamma_{sc,i}$  l'exposant de Lyapunov dans la couche  $i$   
 $\delta L_i$  l'épaisseur de la couche  $i$

Notons que l'expression (III.3.4) reste valable même si les couches ne sont pas indépendantes mais identiques. Puisque  $\sum_i \delta L_i$  est égal à  $z(t)$ , nous obtenons finalement pour l'équation (III.3.3)

$$\hat{r}_{sc,tot}(f|t) = \exp \left[ - \sum_i \delta L_i \gamma_{sc,i} \right]. \quad (III.3.5)$$

Naturellement, les variables  $\delta L_i$  sont doublées pour des données en réflexion.



**Absorption intrinsèque:**  $\hat{r}_{in}$  Afin de compenser de l'absorption intrinsèque, nous utilisons la loi la plus simple, c'est-à-dire que nous supposons que la perte d'énergie est constante par longueur d'onde parcourue. Par conséquent, le facteur de qualité  $Q_{in}$  de l'absorption intrinsèque est une constante. Ceci aboutit à l'expression

$$\hat{r}_{in}(f|t) = \exp \left[ -\frac{1}{2} k_0 z(t) Q_{in}^{-1} \right] \quad (\text{III.3.6})$$

pour l'incidence normale. Le facteur 1/2 apparaît, puisque dans l'équation précédente nous ne décrivons pas l'atténuation de l'énergie, mais des amplitudes.

Si nous désirons introduire plusieurs couches indépendantes, la moyenne de  $Q_{in}^{-1}$  est de nouveau calculée par

$$\langle Q_{in}^{-1} \rangle = \frac{\sum_i \delta L_i Q_{in,i}^{-1}}{\sum_i \delta L_i} \quad (\text{III.3.7})$$

et l'atténuation totale par

$$\hat{r}_{in,tot}(f|t) = \exp \left[ -\frac{1}{2} k_0 \sum_i \delta L_i Q_{in,i}^{-1} \right] \quad (\text{III.3.8})$$

où  $Q_{in,i}^{-1}$  représente la réciproque du facteur de qualité de l'absorption dans la couche  $i$ . De nouveau, les variables  $\delta L_i$  sont doublées pour le problème en réflexion.

**Les effets de la source:**  $\hat{r}_s$  Les équations (III.3.1) et (III.3.2) nous montrent clairement que la source a une influence directe sur les spectres d'amplitude. Ceci a pour conséquence qu'il faut avoir une très bonne estimation de l'ondelette de la source afin de pouvoir bien corriger de tous les effets de source. De plus, ces équations nous montrent aussi qu'il est impossible de déduire quoi que ce soit au-delà de la gamme de fréquences de la source.

Pour des données synthétiques, ceci ne pose aucun problème. Par contre, il est évident que les résultats obtenus en utilisant des données réelles seront déformés à cause d'erreurs d'estimation. Néanmoins, il existe plusieurs techniques pour estimer l'ondelette de source soit en utilisant des méthodes théoriques (Robinson, 1996), soit en prenant une première arrivée supposant qu'elle lui ressemble beaucoup. Dans ce dernier cas, toutes les mesures sont faites par-rapport à la première arrivée.

**Coefficient de réflexion:**  $\hat{r}_{cr}$  Le coefficient de réflexion entre deux couches homogènes pour l'incidence verticale est définie par (pour la pression)

$$\hat{r}_{cr} = \frac{\rho_2 c_2 - \rho_1 c_1}{\rho_1 c_1 + \rho_2 c_2} \quad (\text{III.3.9})$$

où  $\rho_i$  et  $c_i$  représentent respectivement la densité et la vitesse dans la couche  $i$  et où l'indice  $i$  augmente avec la profondeur.

Mais comment faut-il définir ces variables dans un milieu hétérogène? Pour connaître la réponse, il nous faut examiner à nouveau la théorie de l'homogénéisation (section III.2.2.3). Cette théorie nous dit qu'une perturbation locale des paramètres n'a presque aucune influence sur une onde, même si elle est caractérisée par une très grande magnitude. Seules des perturbations soutenues (ou systématiques) seront 'ressenties' par l'onde. Par conséquent, les paramètres effectifs  $\rho_i$  et  $c_i$  peuvent être estimés en calculant leurs moyennes locales à l'échelle de l'impulsion.

La dernière question posée est maintenant: 'Nous faut-il utiliser les moyennes harmoniques ou arithmétiques?' L'équation (III.2.26) nous en donne la réponse (pour l'incidence verticale): la

densité effective est définie par la moyenne arithmétique locale et le carré des vitesses effectives par la réciproque de la moyenne harmonique locale de la compressibilité divisée par la densité effective. Par conséquent,  $\rho_i c_i$  est donné par

$$\rho_i c_i = \langle \rho \rangle_i^{1/2} \langle \kappa^{-1} \rangle_i^{-1/2}. \quad (\text{III.3.10})$$

**Réflexions multiples tardives:**  $\hat{r}_{rmt}$  Pour des données en réflexion, il reste un dernier facteur influençant directement les amplitudes enregistrées. Ce facteur est l'énergie des réflexions multiples tardives arrivant au même temps que la réflexion profonde analysée. Malheureusement, c'est le facteur le plus difficile à prédire, puisque les réflexions multiples tardives interfèrent de façon constructive et/ou destructive entre elles et avec la réflexion profonde obscurcissant ainsi les mesures de l'atténuation apparente. De plus, les interférences exactes changent avec la réalisation (configuration) du milieu et au mieux nous ne pouvons en obtenir qu'une prédiction statistique. Dans la suite nous traitons 3 méthodes pour obtenir cette estimation désirée.

Une des rares estimations théoriques est celle due aux développements mathématiques de Sheng et al. (1986b), White et al. (1987) et Sheng et al. (1990) Ils ont calculé la fonction de corrélation du signal réfléchi en supposant que ce signal est composé d'une convolution de l'ondelette de source et la réflectivité. D'après eux, la valeur attendue de l'énergie  $E[p^2(t)]$  au temps  $t$  est égale à

$$E[p^2(t)] = \int |\hat{r}_s|^2 t^{-1} \mu_i df \quad (\text{III.3.11})$$

avec  $\mu_i$  la densité locale du spectre de puissance (dans leurs articles ce terme est aussi connu sous le nom de spectre de bruit). L'indice  $i$  permet la distinction entre les deux variantes existantes:  $i = 0$  pour un milieu sans surface libre ('matched medium' conditions) où  $r_{cr,surface} = 0$  et  $i = 1$  pour un milieu avec une surface libre ('pressure-release' conditions) où  $r_{cr,surface} = 1$ . Pour le premier type de milieu,  $\mu_i$  est donnée par l'expression

$$\mu_0 = \frac{x}{(1+x)^2} \quad (\text{III.3.12})$$

et pour le deuxième type

$$\mu_1 = 4x, \quad (\text{III.3.13})$$

où  $x$  est égal à

$$x = \frac{t c_0}{l(f)}. \quad (\text{III.3.14})$$

Il nous est maintenant possible d'obtenir une expression pour  $\hat{r}_{rmt}$  en utilisant le deuxième terme de l'équation (III.3.2), le théorème de Parseval et l'inégalité de Cauchy-Schwarz. Ceci résulte en

$$E[p^2(t)] \leq \int |\hat{r}_s|^2 |\hat{r}_{rmt}|^2 df. \quad (\text{III.3.15})$$

Puisque cette relation est valable indépendamment de l'ondelette de source exacte  $\hat{r}_s$  et de sa distribution en fréquence, nous obtenons pour  $\hat{r}_{rmt}$  en utilisant l'équation (III.3.11)

$$|\hat{r}_{rmt}| \geq \sqrt{t^{-1} \mu_i}. \quad (\text{III.3.16})$$

Notons que les équations (III.3.12) et (III.3.13) montrent que l'énergie des réflexions multiples tardives est liée à l'exposant de Lyapunov  $\gamma_{sc}$ .



Il est possible d'obtenir une deuxième estimation de  $|\hat{r}_{rmt}|$  en notant qu'elle ne peut représenter qu'une fraction de l'énergie totale perdue par l'impulsion et donc réfléchi  $|R|^2$ . Cette variable est donnée par

$$|R|^2 = 1 - |T|^2. \quad (\text{III.3.17})$$

En utilisant le fait que  $|T| = \hat{r}_{sc} = \exp[-\gamma_{sc}z]$ , expression (III.3.3), nous obtenons finalement

$$|\hat{r}_{rmt}| = \alpha |R| = \alpha \sqrt{1 - \exp[-2\gamma_{sc}z]} \text{ avec } \alpha < 1, \quad (\text{III.3.18})$$

où  $\alpha$  représente la fraction de l'énergie totale réfléchi arrivant avec les réflexions multiples tardives au temps  $t$ . Malheureusement, nous n'avons trouvé aucune estimation théorique du paramètre  $\alpha$ . Par contre, des simulations numériques (sans surface libre) nous ont montré qu'elle est de l'ordre de 0.10.

Il est important de noter que les estimations théoriques fournies sont faites en négligeant l'influence de l'absorption. De plus, il est difficile d'estimer clair comment celle-ci influence ces estimations. Par conséquent, nous sommes obligés de supposer qu'il est possible de négliger cette influence.

Notons que, de nouveau, nous avons vu que l'énergie des réflexions multiples tardives est proportionnelle à l'exposant de Lyapunov. Par exemple pour un coefficient de transmission  $|T|$  proche de 1, le milieu est effectivement homogène et, par conséquent, il n'y a pas de réflexions multiples tardives et vice versa. Ceci est aussi la raison pour laquelle, dans les articles de Burridge et al. (1994) et Lewicki et Burridge (1996), il n'apparaît aucun terme correctif pour les réflexions multiples tardives. Dans ces articles, les auteurs prédisent la forme et l'amplitude d'une réflexion profonde dans un milieu aléatoire. Aucun terme concernant les réflexions multiples tardives n'est incorporé dans leurs résultats finaux, puisqu'ils utilisent une approximation à basses fréquences. Dans cette gamme de fréquence, l'exposant de Lyapunov est proche de zéro et donc l'énergie des réflexions multiples tardives est négligeable.

La dernière manière pour estimer  $\hat{r}_{rmt}$  est simplement de mesurer l'énergie enregistrée avant ou après l'arrivée de la réflexion analysée. Donc, il faut mesurer le spectre de bruit en fonction de la fréquence. Afin d'obtenir des estimations correctes de l'énergie des réflexions multiples tardives, il est nécessaire d'avoir bien corrigé de la divergence sphérique, puisque ces estimations sont faites à des temps différents.

Par conséquent, nous voyons que l'approximation faite de  $\hat{r}_{rmt}$  dépend non seulement de la qualité des mesures ou des prédictions théoriques, mais aussi de la qualité de l'estimation de la divergence sphérique et peut-être même de l'absorption. De plus, malheureusement, toutes ces méthodes ne donnent qu'une estimation statistique de l'influence des réflexions multiples tardives. Ceci ne suffit pas, parce qu'il nous manque des informations sur le spectre de phase afin de pouvoir bien prédire toutes les interférences constructives et destructives dans l'équation (III.3.2). Seulement, en supposant que les diffractions multiples aboutissent à une phase aléatoire, nous pouvons le négliger. Dans ce cas, nous pouvons approximer l'énergie totale  $|\tilde{r}_{tot,r}|^2$  enregistrée en temps  $t$  par une simple addition des énergies

$$|\tilde{r}_{tot,r}(f|t)| = \left\{ |\hat{r}_{cr}(t) \cdot \hat{r}_{ds}(t) \cdot \hat{r}_{sc}(f|t) \cdot \hat{r}_{in}(f|t) \cdot \hat{r}_s(f)|^2 + |\tilde{r}_{rmt}(f|t) \cdot \hat{r}_s(f)|^2 \right\}^{1/2} \quad (\text{III.3.19})$$

où  $\tilde{r}_{rmt}$  est l'approximation faite de l'influence des réflexions multiples tardives.

Il est évident que les réflexions multiples tardives causent beaucoup de problèmes. Le problème principal n'est même pas le fait qu'il est difficile de les corriger, mais le fait que le phénomène de la localisation d'onde perd son caractère auto-moyennant. Donc, pour le problème en réflexion, le phénomène de la localisation d'onde perd beaucoup de son intérêt, puisqu'il s'agit

maintenant d'une théorie statistique 'normale'. Seulement, si l'on reste dans le domaine de la diffusion de type 'Rayleigh' (à basses fréquences), le phénomène reste auto-moyennant, puisque nous pouvons négliger l'influence obscurcissante des réflexions multiples tardives. Malheureusement, à basses fréquences, nous ne pouvons estimer par inversion que le produit de l'échelle caractéristique des hétérogénéités  $a$  et leur variance  $\langle \sigma_a^2 \rangle$  (voir équations (III.2.48), (III.2.60) et (III.2.67)).

### III.3.3 Méthodes d'inversion

Dans cette partie nous décrivons brièvement les méthodes d'inversion utilisées. Elles sont décrites plus en détails dans les annexes III.A et III.B. Afin d'estimer par inversion pour les paramètres statistiques décrivant le milieu, nous utilisons non seulement une méthode d'inversion locale, mais aussi une technique globale. Nous faisons appel à une technique globale, puisqu'il est non seulement notre intention de retrouver le meilleur modèle, mais aussi d'examiner l'espace total des solutions afin de reconstruire toutes les corrélations existantes entre les diverses paramètres. Ceci nous aidera à mieux découvrir les problèmes empêchant une bonne mise en œuvre d'une méthode d'inversion. Bien qu'une méthode d'inversion locale puisse aboutir à une meilleure estimation des paramètres décrivant le problème, elle ne nous donne pas les corrélations non-linéaires entre les paramètres.

**La méthode d'inversion globale** Au lieu d'utiliser une technique de Monte Carlo pure, nous ferons appel à une méthode globale adaptive à savoir l'algorithme Vegas (Lepage, 1978). L'avantage d'une méthode adaptive est qu'elle converge plus vite qu'une technique non-adaptive, sans perdre sa capacité de reconstruire l'espace totale des solutions (aussi complexe soit-elle avec par exemple des corrélations asymétriques et/ou des multiples solutions optimales).

L'algorithme Vegas est à l'origine une technique d'intégration concentrant l'échantillonnage aux amplitudes contribuant le plus à l'intégrale, c'est-à-dire aux amplitudes maximales. Cet algorithme a été adapté par Van der Baan (1996) et Muzyert (1998) afin qu'elle puisse résoudre des problèmes inverses. L'algorithme adapté utilise le fait qu'une erreur minimale entre les observations et les prédictions théoriques correspond à une vraisemblance maximale de la solution proposée. Par conséquent, une minimisation des erreurs de prédiction correspond à une optimisation des vraisemblances. Donc, nous pouvons utiliser l'algorithme Vegas afin de trouver les solutions ayant les vraisemblances maximales, puisque cet algorithme cherche justement les solutions qui attribuent le plus au calcul intégrale. De plus, l'échantillonnage de l'espace est proportionnel à la contribution relative au calcul de l'intégrale. Donc, l'échantillonnage est concentré dans les parties ayant les vraisemblances maximales sans que la méthode néglige les parties restantes. Par conséquent, il peut trouver les solutions correspondant aux vraisemblances maximales beaucoup plus vite que des méthodes échantillonnant l'espace de solutions de façon uniforme.

Comme avantage additionnel, Vegas peut nous fournir des fonctions marginales individuelles et combinées de la densité de probabilité. Le premier type de fonction nous donne la vraisemblance qu'un paramètre individuel du modèle se trouve dans un certain intervalle, tandis que le deuxième type représente toutes les corrélations existantes entre deux paramètres d'inversion. Ces dernières correspondent à des matrices de corrélation généralisées (Muzyert, 1998). L'annexe III.B contient une description mathématique de l'algorithme.

**La méthode d'inversion locale** Nous avons non seulement utilisé une méthode d'inversion globale, mais aussi une méthode locale, puisqu'une telle méthode est mieux capable de trouver



$L$	2000 m	$z_{source}$	0 m
$c_0$	4000 m/s	$\rho_0$	$2.2 \text{ g cm}^{-3}$
$\Delta z$	1 m	$\Delta t$	2 ms
$f_{p,Ricker}$	80 Hz	$a$	10 m
$\sigma_\kappa$	5 et 15 %	$\phi_i$	$0^\circ$
type	exponentiel	$Q_{in}$	$10^6$

TAB. III.3.2 – Paramètres constants des simulations numériques.  $L$ : épaisseur du milieu,  $z_{source}$ : profondeur de la source,  $c_0$ : vitesse moyenne,  $\rho_0$ : densité,  $\Delta z$ : épaisseur des couches,  $\Delta t$ : pas d'échantillonnage,  $f_{p,Ricker}$ : fréquence principale de l'ondelette de Ricker,  $a$ : échelle caractéristique des hétérogénéités,  $\sigma_\kappa$ : magnitude caractéristique des hétérogénéités,  $\phi_i$ : angle d'incidence, type: type de fonction d'autocorrélation et  $Q_{in}$ :  $Q$  intrinsèque.

une solution optimale du problème considéré. Ce type de méthodes aboutit à une telle amélioration des paramètres du modèle en linéarisant l'influence d'une perturbation de ces paramètres sur le désaccord entre les valeurs observées (mesurées) et prédites. Cette linéarisation se fait autour d'une estimation initiale des paramètres du modèle. Ensuite, une nouvelle estimation des paramètres d'inversion est obtenue en minimisant ce désaccord selon une certaine norme (dans notre cas, il s'agit de la norme  $l_2$  qui correspond à une minimisation de l'énergie).

L'avantage d'une méthode d'inversion locale est qu'elle produit des paramètres expliquant mieux les observations. Par contre, une méthode locale ne peut nous donner qu'une seule nouvelle estimation. Par conséquent, il existe toujours le risque de tomber dans un minimum local entre les observations et les prédictions au lieu de trouver le minimum global désiré. Le minimum obtenu dépend directement de l'estimation initiale des paramètres d'inversion, d'où l'intérêt d'utiliser plusieurs estimations initiales. De plus, un algorithme global (comme Vegas) peut nous fournir ces estimations initiales. Par conséquent, une approche combinée peut être très utile pour des problèmes compliqués.

Les équations mathématiques nécessaires pour inverser les données par une méthode locale sont contenues dans l'annexe III.A.

### III.3.4 Inversion de données synthétiques

#### III.3.4.1 Stratégie d'inversion

D'abord, nous utilisons des données synthétiques afin de pouvoir bien étudier tout problème empêchant la mise en œuvre d'une méthode d'inversion. Nous examinons surtout l'effet des réflexions multiples tardives, puisque, dans le paragraphe III.3.2, nous avons vu qu'elles peuvent compliquer les mesures de l'atténuation apparente.

Pour analyser ce problème, nous ferons à nouveau appel à la stratégie utilisée dans la section III.2. Donc, nous ferons deux simulations numériques: la première pour le milieu hétérogène et l'autre pour le milieu homogène. De nouveau, les simulations numériques sont faites avec la méthode de Dietrich (1988). Le tableau III.3.2) nous montre les paramètres constants des simulations numériques.

Comme nous avons deux mesures, à savoir  $\hat{r}_{tot,r}^{hom}$  (cas homogène) et  $\hat{r}_{tot,r}^{hét}$  (cas hétérogène), nous pouvons directement calculer le rapport spectral  $\hat{r}_{rsp,r}$ . Puisque  $\hat{r}_{tot,r}^{hom}$  est donné par (tableau III.3.1

$$\hat{r}_{tot,r}^{hom} = \hat{r}_{cr} \cdot \hat{r}_{ds} \cdot \hat{r}_s \quad (III.3.20)$$

et  $\hat{r}_{tot,r}^{hét}$  est approximé par l'équation III.3.19, nous obtenons pour le rapport spectral  $\hat{r}_{rsp,r}$

$$\hat{r}_{rsp,r} = \frac{\hat{r}_{tot,r}^{hét}}{\hat{r}_{tot,r}^{hom}} = \left\{ |\hat{r}_{sc} \cdot \hat{r}_{in}|^2 + \left| \frac{\tilde{r}_{rmt}}{\hat{r}_{cr} \cdot \hat{r}_{ds}} \right|^2 \right\}^{1/2} \quad (III.3.21)$$

Malheureusement, il y aura toujours des erreurs dans l'estimation de  $\hat{r}_{tot,r}^{hom}$  (seulement pour le terme de divergence sphérique  $\hat{r}_{ds}$  dans un cas synthétique, mais dans un cas réel pour tous les trois paramètres  $\hat{r}_{cr}$ ,  $\hat{r}_{ds}$  et  $\hat{r}_s$ ). Nous avons vu dans la théorie que  $\hat{r}_{cr}$  et  $\hat{r}_{ds}$  sont des constantes en fréquence (pour  $t$  fixé). Par conséquent, nous pouvons les estimer à un facteur constant de multiplication  $c_{st}$  près:

$$\hat{r}_{cr} \cdot \hat{r}_{ds} = c_{st} \cdot \tilde{r}_{cr} \cdot \tilde{r}_{ds} \quad (III.3.22)$$

où le tilde représente l'approximation faite. Donc, l'équation (III.3.21) est transformée en

$$\begin{aligned} \hat{r}_{rsp,r} &= \left\{ \left| \frac{c_{st}^{hét}}{c_{st}^{hom}} \cdot \tilde{r}_{sc} \cdot \tilde{r}_{in} \right|^2 + \left| \frac{\tilde{r}_{rmt}}{c_{st}^{hom} \cdot \tilde{r}_{cr} \cdot \tilde{r}_{ds}} \right|^2 \right\}^{1/2} \\ &= \left\{ |c_{st} \cdot \tilde{r}_{sc} \cdot \tilde{r}_{in}|^2 + \mu \left| \frac{\tilde{r}_{rmt}}{\tilde{r}_{cr} \cdot \tilde{r}_{ds}} \right|^2 \right\}^{1/2} \end{aligned} \quad (III.3.23)$$

Nous serons obligés de négliger l'erreur faite en approximant  $\hat{r}_s$  en espérant qu'elle n'est pas systématique.

L'équation (III.3.23) nous montre un problème considérable. Même si nous sommes capables d'estimer l'influence des réflexions multiples tardives, il existera toujours une ambiguïté entre les premier et second termes de cette équation. Ceci est exprimé par les constantes de multiplication  $c_{st}$  et  $\mu$ . Bien sûr, nous pouvons estimer les deux constantes simultanément, mais une telle stratégie ne peut qu'échouer. Une solution à ce problème est de considérer la variable  $\mu$  comme une constante de multiplication Lagrangienne. De cette façon, nous n'attribuons qu'une fonction correctrice au deuxième terme. Toutefois, toute erreur faite sur l'estimation de la valeur de  $\mu$  aura des conséquences sur la qualité des résultats obtenus après inversion. Dans le reste de cette étude, la valeur de  $\mu$  est prise égale à 1.0.

La suite de cette section est divisée en trois parties. D'abord nous montrons en détails les effets exacts des réflexions multiples tardives. Nous verrons à la fois leur caractère aléatoire dans une seule réalisation du milieu et leurs effets systématiques en moyennant sur une trentaine de réalisations. Ceci nous montrera bien leur caractère statistique. Nous verrons aussi qu'elles introduisent des erreurs systématiques dans les inversions, si nous ne corrigeons pas de leurs effets. De plus, nous examinons les corrélations existantes entre les différents paramètres en utilisant l'algorithme Vegas.

Dans la deuxième partie, nous testons les méthodes proposées pour prédire l'influence des réflexions multiples tardives. Nous voulons surtout obtenir une idée de leurs avantages et inconvénients, ainsi que de la qualité de leurs prédictions.

Finalement, dans la dernière partie, nous essayons de corriger des effets des réflexions multiples tardives afin d'obtenir des estimations correctes des paramètres statistiques décrivant le milieu hétérogène.

#### III.3.4.2 Le problème des réflexions multiples tardives

Avant d'illustrer les effets à la fois aléatoires et systématiques des réflexions multiples tardives sur l'atténuation apparente d'une impulsion traversant un milieu hétérogène, nous montrons



quelques sismogrammes obtenus dans des milieux homogènes et hétérogènes. Tous les milieux homogènes et hétérogènes sont identiques à part le coefficient de la réflexion profonde.

Ces sismogrammes sont montrés sur la figure III.3.1. Comme sur la figure III.2.3, les traces alternent entre une réalisation homogène et hétérogène qui ont néanmoins le même coefficient de réflexion effectif. Le premier ensemble de deux traces montre le cas en transmission et le deuxième ensemble le cas en réflexion avec un réflecteur parfait, puis  $\hat{r}_{cr}$  diminue ( $\hat{r}_{cr}$  devient respectivement 1/2, 1/3 et finalement 1/5). Le cas en transmission est montré puisque nous l'analyserons plus tard. Le milieu hétérogène utilisé pour calculer le cas en transmission est construit en utilisant le milieu du cas en réflexion suivi de son symétrique. De cette façon, l'atténuation apparente en transmission devrait égaler celle d'une réflexion sur un réflecteur parfait. L'inversion de la polarité entre le premier ensemble de 2 traces et les suivants est causée par le fait que les sismogrammes montrent le déplacement vertical et non la pression. Par conséquent, le coefficient de réflexion devient l'opposé de celui de l'équation (III.3.9) (Pain, 1983).

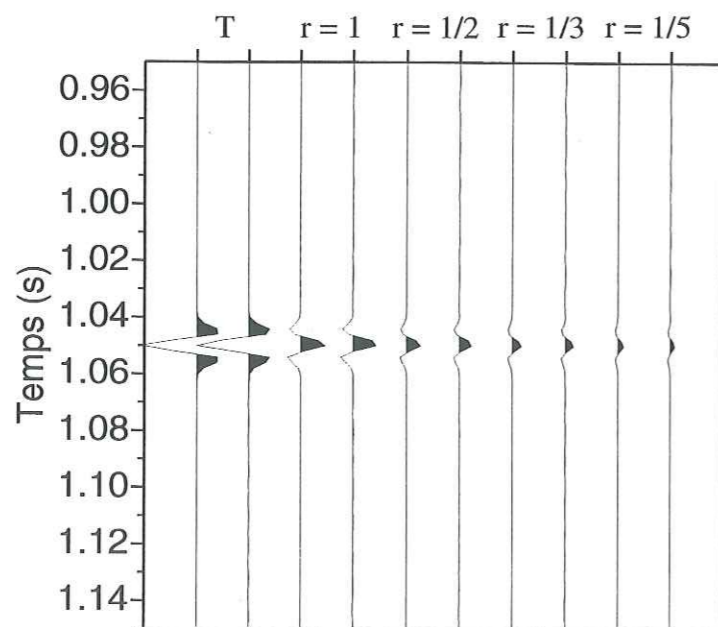


FIG. III.3.1 – Sismogrammes synthétiques obtenus pour une réalisation particulière du milieu avec  $\sigma_{\kappa} = 5\%$ . Chaque ensemble de 2 traces est composé d'un sismogramme pour un milieu homogène et ensuite pour un milieu hétérogène. Les ensembles correspondent de gauche à droite à un cas en transmission et plusieurs cas en réflexion où le coefficient de réflexion baisse (d'un réflecteur parfait jusqu'à  $\hat{r}_{cr} = 0.20$ ).

**Mesures** Même si nous n'avons pas encore corrigé de la divergence sphérique, la figure nous montre bien que les amplitudes des réflexions multiples tardives sont largement inférieures à celles de la réflexion profonde. Ceci nous donne l'idée trompeuse que nous pouvons les négliger. Par contre, la figure III.3.2 nous indique clairement que cette idée est fautive, puisqu'elle nous présente les mesures faites de la longueur de localisation en fonction du coefficient de réflexion. Cette figure nous montre que l'influence des réflexions multiples tardives augmente lorsque le coefficient de réflexion diminue. Heureusement, il n'est pas difficile à comprendre pourquoi ceci est le cas

en examinant l'équation (III.3.23) où nous voyons que le deuxième terme (correspondant à l'influence de réflexions multiples tardives) est inversement proportionnel à  $\hat{r}_{cr}$ .

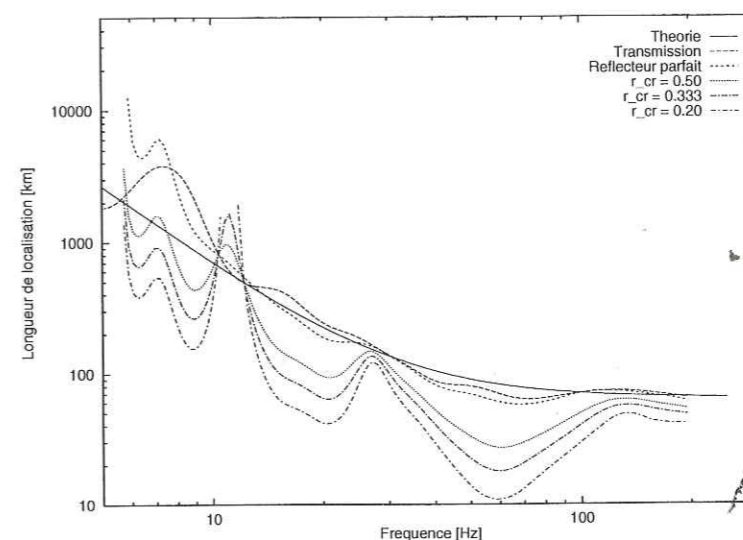


FIG. III.3.2 – Mesures de la longueur de localisation des sismogrammes montrés sur la figure III.3.1.

De plus, nous voyons que les estimations de la longueur de localisation faites pour le cas en transmission et pour le cas d'un réflecteur parfait sont presque identiques. Donc, l'approximation que l'atténuation d'une réflexion correspond à deux fois la transmission à travers du milieu est correcte.

La figure III.3.2 est aussi intéressante, puisque, pour cette réalisation particulière du milieu, la longueur de localisation diminue lorsque le coefficient de réflexion diminue. Ceci indique que, dans ce cas particulier, l'atténuation apparente croît alors que la part d'énergie due aux réflexions multiples tardives augmente. Ce cas est contraire à la logique, car normalement les réflexions multiples tardives font justement diminuer l'atténuation apparente, puisqu'elles font croître l'énergie totale enregistrée (voir équation (III.3.23)). Donc, cet exemple nous montre bien le caractère aléatoire des interférences constructives et destructives causées par ces réflexions multiples. En effet, elles peuvent même faire que certaines fréquences ont des amplitudes supérieures à celles du cas homogène!

Les déviations systématiques ne se révèlent qu'en moyennant sur plusieurs réalisations. Ainsi, sur les figures III.3.3 et III.3.4 sont représentés les rapports spectraux et les écarts-types associés en utilisant 30 réalisations et 2 valeurs de  $\sigma_{\kappa}$ , 5 et 15 %. Sur les figures III.3.3.a et III.3.4.a nous avons représenté les mesures pour  $\hat{r}_{cr} = 1$  et sur les III.3.3.b et III.3.4.b celles pour  $\hat{r}_{cr} = 0.2$ , ainsi que les valeurs théoriques. De nouveau, nous voyons que non seulement l'influence des réflexions multiples tardives augmente lorsque  $\hat{r}_{cr}$  diminue, mais aussi qu'il y a des déviations systématiques par-rapport à la valeur théorique pour  $\hat{r}_{cr} = 0.2$  (pour  $\sigma_{\kappa} = 15\%$ , figure III.3.4.b), confirmant ainsi ce que nous avons dit ci-dessus.

**Inversions des mesures** Il est facile de montrer d'une autre manière comment les réflexions multiples tardives introduisent des erreurs systématiques, simplement en essayant d'inverser les courbes et écarts-types montrés sur les figures III.3.3 et III.3.4. De plus, ces essais nous montreront les corrélations existantes entre tous les paramètres. Par conséquent, ils aideront à mieux comprendre le problème posé et peut-être même à trouver une stratégie d'inversion



permettant d'estimer correctement ces paramètres.

Pour parvenir à ces buts, nous commencerons par inverser les mesures montrées sur la figure III.3.4.a pour un réflecteur parfait en utilisant l'algorithme Vegas<sup>2</sup> et ensuite pour le cas  $\hat{r}_{cr} = 0.2$  (figure III.3.4.b). Nous ne montrons que les résultats pour  $\sigma_\kappa = 15\%$ , car ici les effets systématiques des réflexions multiples tardives sont plus facilement visibles que pour  $\sigma_\kappa = 5\%$  (figure III.3.3).

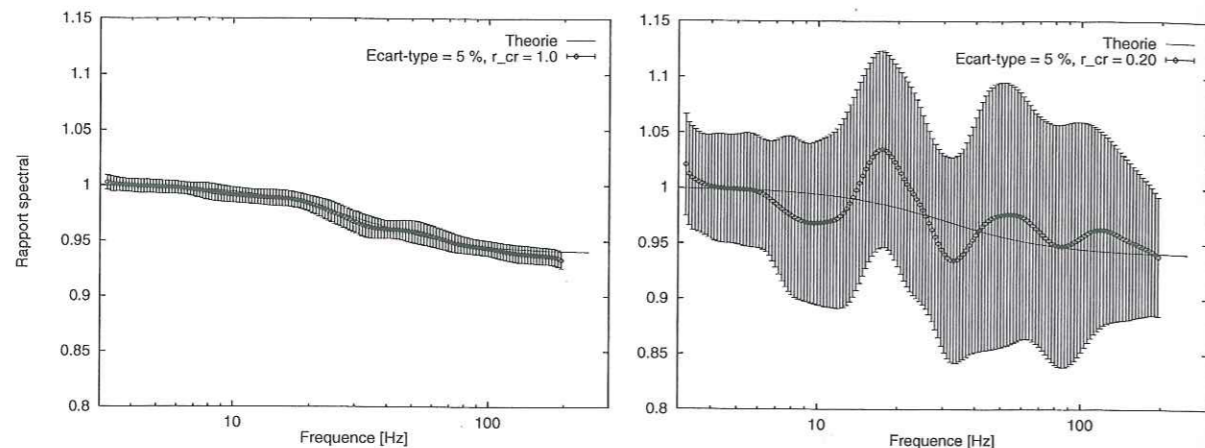


FIG. III.3.3 – Mesures du rapport spectral pour  $\sigma_\kappa = 5\%$  et a) un réflecteur parfait et b)  $\hat{r}_{cr} = 0.20$ .

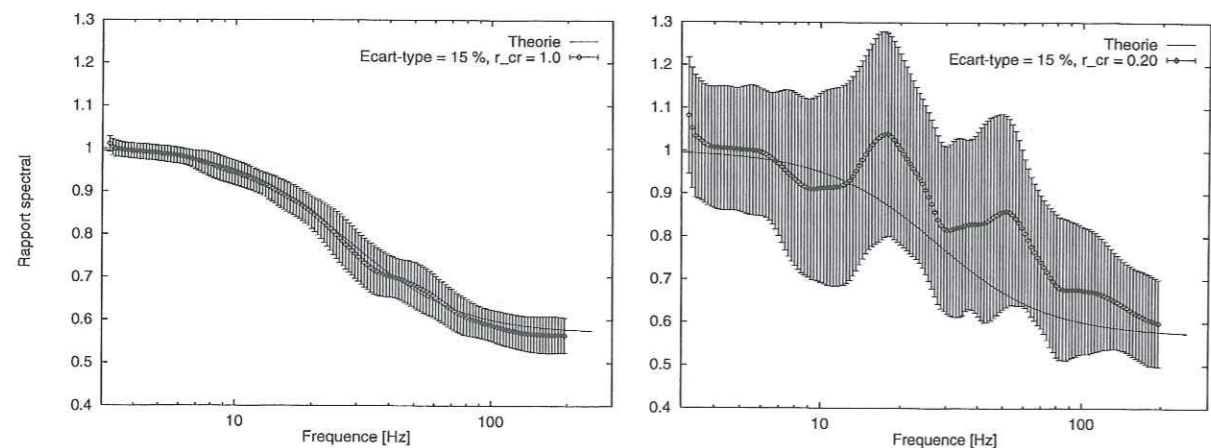


FIG. III.3.4 – Comme la figure précédente, sauf que  $\sigma_\kappa = 15\%$ . a) réflecteur parfait. b)  $\hat{r}_{cr} = 0.20$ .

La figure III.3.5 nous montre les fonctions marginales individuelles de la densité de probabilité pour la constante de multiplication  $c_{st}$ ,  $\log(a)$ ,  $\sigma_\kappa$  et  $\log(Q)$ . Ces figures nous donnent les vraisemblances individuelles qu'un certain paramètre se trouve dans un intervalle particulier indépendant des valeurs des autres paramètres. Cette vraisemblance s'obtient en intégrant la fonction entre les deux limites de l'intervalle. Donc, nous voyons que les valeurs individuelles les plus probables sont  $c_{st} = 1.0$ ,  $\log(a) = 1.1$ ,  $\sigma_\kappa = 13\%$  et  $\log(Q) = 2.8$ . Par contre, le modèle à vraisemblance maximale trouvé par l'algorithme Vegas est  $c_{st} = 1.00$ ,  $\log(a) = 1.00$ ,  $\sigma_\kappa = 15.65\%$  et  $\log(Q) = 6.56$ . Les différences entre ces valeurs sont dues à l'existence de corrélations entre les paramètres. Néanmoins, ces valeurs sont très proches des vraies valeurs (tableau III.3.2).

2. Vegas a utilisé 4 itérations de 20000 échantillons chacune.

De plus, les figures nous donnent les résolutions que nous pouvons atteindre si les écarts-types du rapport spectral en fonction de la fréquence sont petits (figure III.3.4.a). Dans ce cas, on voit que nous aurons une bonne estimation de la constante de multiplication  $c_{st}$  et peut-être du facteur de qualité  $Q_{in}$  (ce dernier en utilisant le modèle à vraisemblance maximale). Seulement, il sera plus difficile de retrouver les vraies valeurs pour  $\log(a)$  et  $\sigma_\kappa$ .

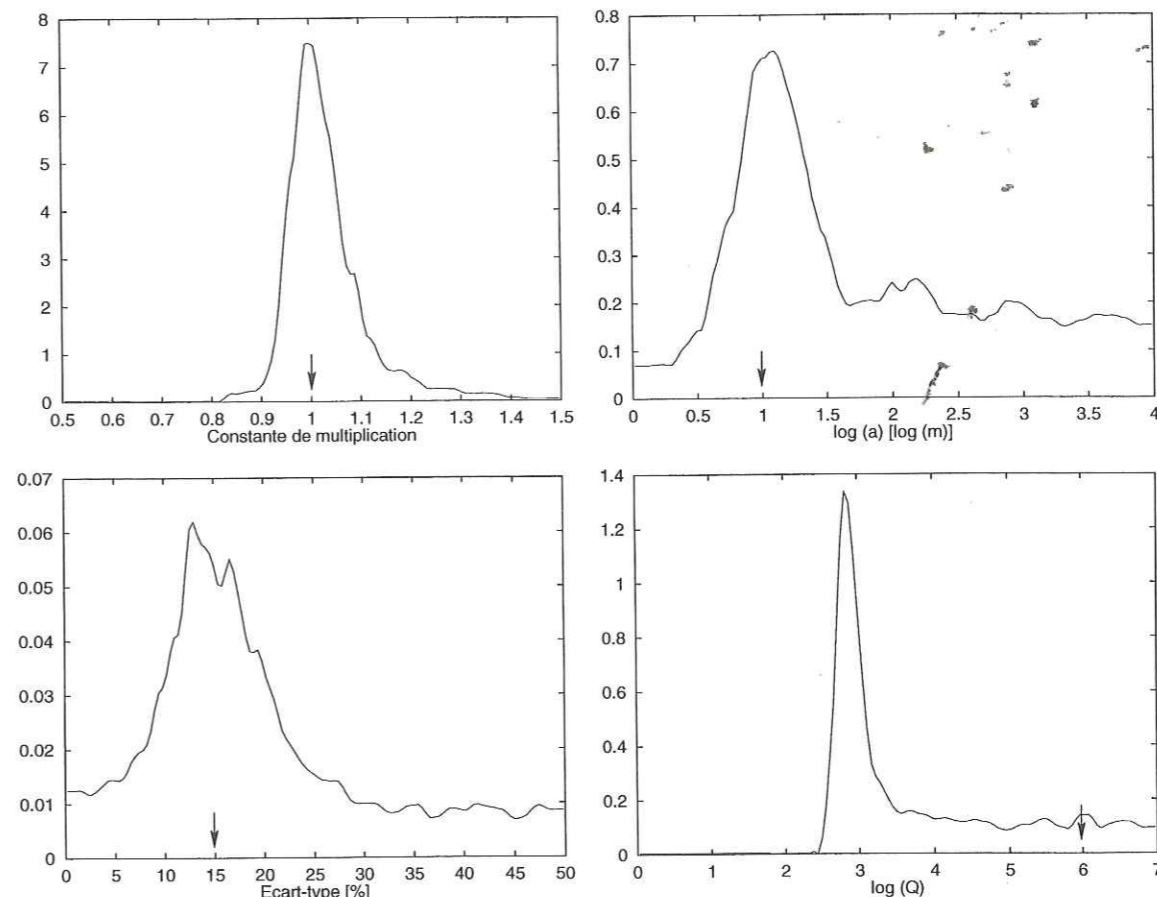


FIG. III.3.5 – Fonctions marginales individuelles de la densité de probabilité pour  $\sigma_\kappa = 15\%$  et un réflecteur parfait (figure III.3.4.a). L'«Ecart-type» est  $\sigma_\kappa$  et les flèches indiquent les vraies valeurs des paramètres.

Comme nous l'avons indiqué ci-dessus, Vegas peut aussi nous fournir des fonctions marginales combinées de la densité de probabilité. Ces fonctions nous décrivent toutes les corrélations existantes entre deux ou plusieurs paramètres. De plus, elles nous donnent la vraisemblance qu'une combinaison de paramètres se trouve dans un certain intervalle. Sur la figure III.3.6 nous les avons présentées pour le cas d'un réflecteur parfait et  $\sigma_\kappa = 15\%$ . Une inspection de ces fonctions nous montre qu'il n'y a presque aucune corrélation entre  $c_{st}$  et  $\log(Q)$ . Par contre, il existe de légères corrélations positives entre  $\log(a)$  et  $\sigma_\kappa$ ,  $\log(a)$  et  $c_{st}$  et donc aussi entre  $\sigma_\kappa$  et  $c_{st}$ . C'est à cause de cette corrélation positive entre  $\log(a)$  et  $\sigma_\kappa$  qu'il sera difficile de retrouver leurs vraies valeurs dans une inversion.

De plus, la figure III.3.6 nous montre la relation inhabituelle entre  $\log(Q)$  et  $\log(a)$  et  $\log(Q)$  et  $\sigma_\kappa$ . Il nous semble que cette première relation peut être expliquée par une corrélation positive entre  $Q^{-1}$  et  $\log(a)$  (et donc  $a$ ). Il est possible d'expliquer une telle relation si nous nous rappelons que  $a$  et le coefficient de Lyapunov  $\gamma_{sc}$  sont inversement reliés (voir 'Absence of localization' dans



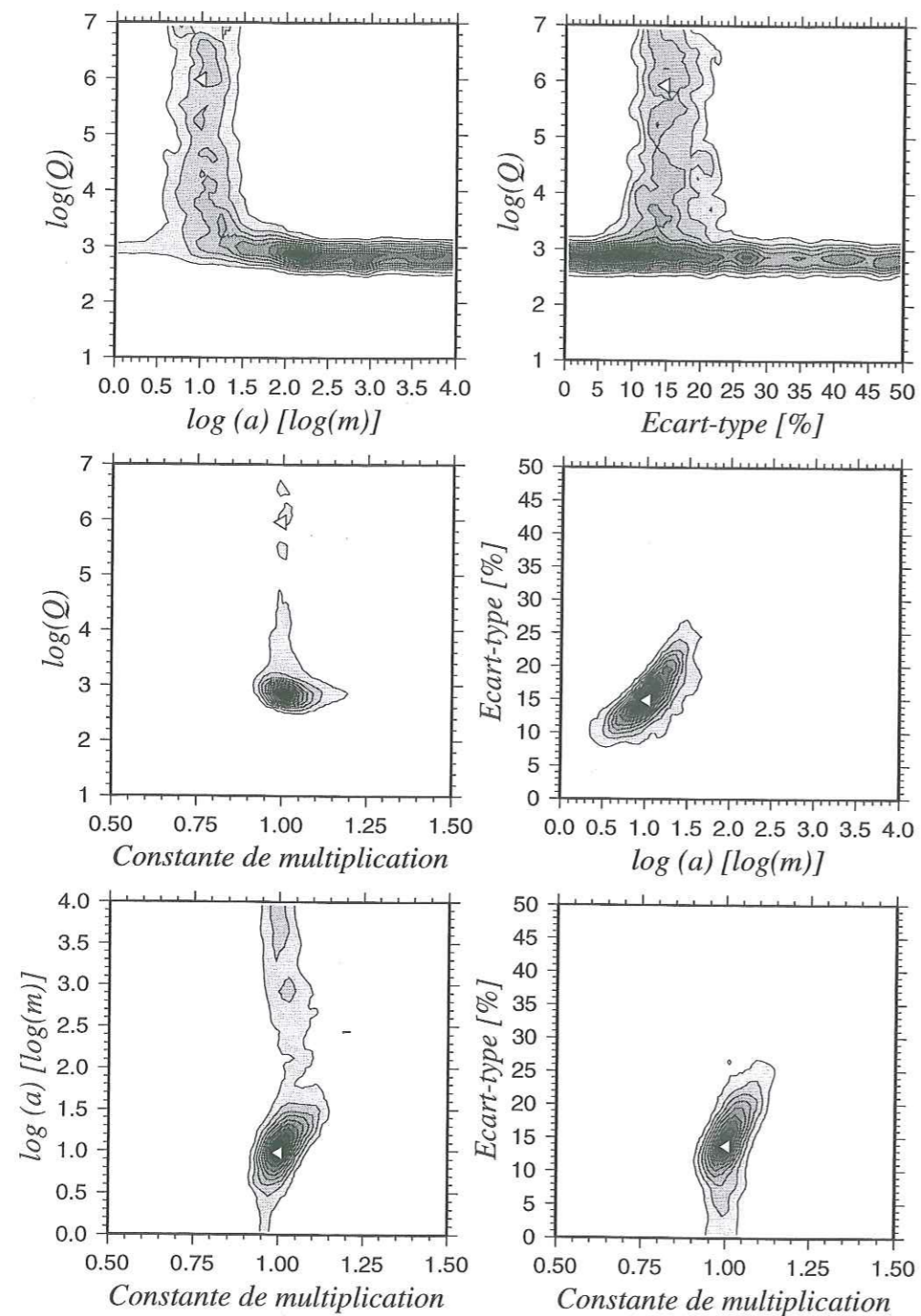


FIG. III.3.6 – Fonctions marginales combinées de la densité de probabilité pour  $\sigma_\kappa = 15\%$  et un réflecteur parfait (figure III.3.4.a). Les triangles blancs indiquent les vraies valeurs des paramètres.

la section III.2). Par conséquent, nous pouvons légèrement compenser une réduction de  $\hat{r}_{sc}$  en diminuant  $Q^{-1}$ , puisque dans ce cas  $\hat{r}_{in}$  augmentera (voir les équations (III.3.3) et (III.3.6)). Donc, le rapport spectral ne changera pas. Par contre, le fait que  $\hat{r}_{sc}$  et  $\hat{r}_{in}$  sont bornés à 1 fait que cette relation devient non-linéaire et asymétrique.

Ensuite, nous avons inversé le rapport spectral mesuré pour le cas d'un réflecteur imparfait ( $\hat{r}_{cr} = 0.20$ , figure III.3.4.b). Les fonctions marginales de la densité de probabilité obtenues sont montrées sur les figures III.3.7 (fonctions marginales individuelles) et III.3.8 (fonctions marginales combinées). En les comparant avec les figures III.3.5 et III.3.6, nous pouvons observer plusieurs faits.

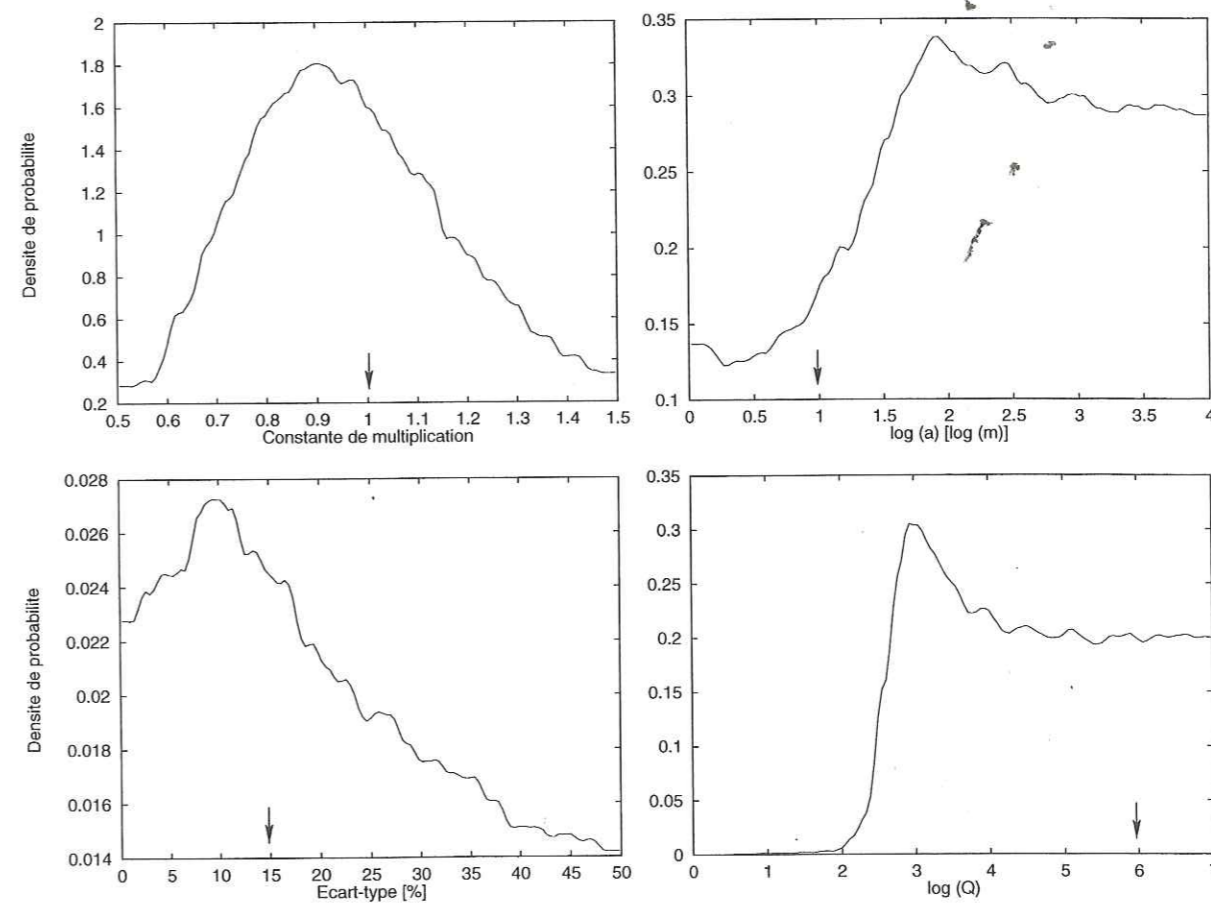


FIG. III.3.7 – Fonctions marginales individuelles de la densité de probabilité pour  $\sigma_\kappa = 15\%$  et un réflecteur imparfait ( $\hat{r}_{cr} = 0.20$ , figure III.3.4.b). À comparer avec la figure III.3.5.

Premièrement, elles nous montrent clairement que les incertitudes et les corrélations des paramètres ont augmenté considérablement. Deuxièmement, toutes les vraisemblances maximales des paramètres individuels ont des erreurs systématiques. En examinant les fonctions marginales individuelles de la densité de probabilité, nous pouvons voir que les valeurs individuelles les plus probables sont maintenant  $c_{st} = 0.9$ ,  $\log(a) = 1.9$ ,  $\sigma_\kappa = 10\%$  et  $\log(Q) = 3.0$ , alors que le modèle à vraisemblance maximale est  $c_{st} = 1.01$ ,  $\log(a) = 0.79$ ,  $\sigma_\kappa = 10.97\%$  et  $\log(Q) = 3.97$ . Cette différence est causée à nouveau par les corrélations entre les paramètres.

**Inversions supplémentaires** Les figures III.3.5 à III.3.8 montrent de très grandes différences. Pour distinguer entre les influences des divers paramètres, nous ferons encore 3 inversions sup-



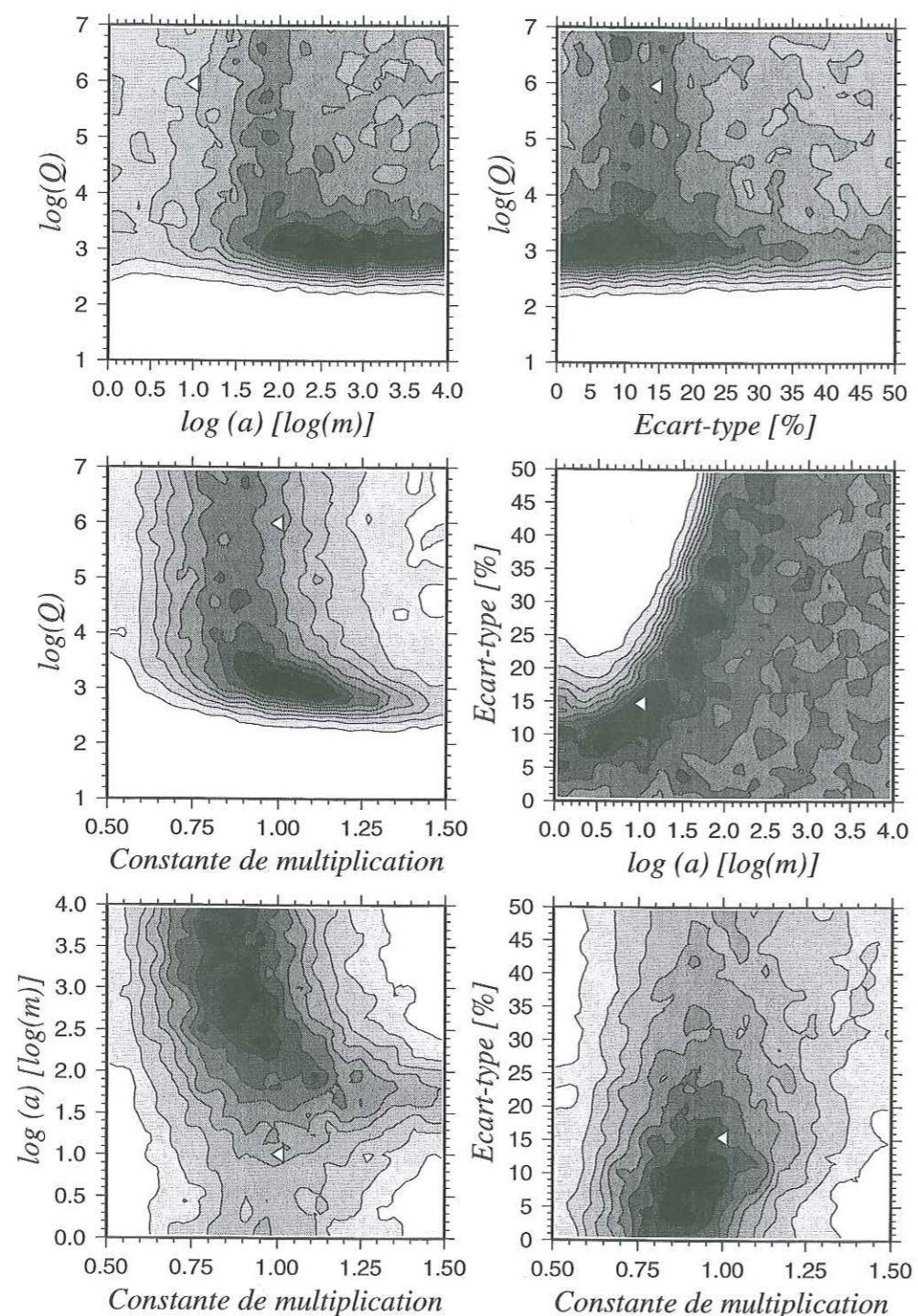


FIG. III.3.8 – Fonctions marginales combinées de la densité de probabilité pour  $\sigma_{\kappa} = 15\%$  et un réflecteur imparfait ( $\hat{r}_{cr} = 0.20$ , figure III.3.4.b). A comparer avec la figure III.3.6.

plémentaires.

Nous avons donc construit 3 mesures synthétiques du rapport spectral. Les deux premières courbes sont composées du rapport spectral pour  $\hat{r}_{cr} = 1.0$  (figure III.3.4.a) avec des écarts-types doublés et quadruplés. Une amplification d'un facteur de 4 correspond à peu près aux écarts-types pour  $\hat{r}_{cr} = 0.20$  (figure III.3.4.b), mais il manque la déviation systématique causée par l'énergie des réflexions multiples tardives (qui a tendance à diminuer l'atténuation apparente enregistrée). La dernière courbe est une combinaison du même rapport spectral ( $\hat{r}_{cr} = 1.0$ ), mais maintenant avec les erreurs de mesures observées pour  $\hat{r}_{cr} = 0.20$  (figure III.3.4.b). Ces trois modèles supplémentaires nous permettront de bien isoler l'influence des variations d'incertitudes sur les résultats obtenus.

Les 3 courbes ont été inversées en utilisant l'algorithme Vegas. Les résultats sont montrés sur les figures III.3.9 à III.3.11. La figure III.3.9 nous montre les fonctions marginales individuelles de la densité de probabilité comparées à celles des figures III.3.5 et III.3.7. Les figures III.3.10 et III.3.11 nous présentent les fonctions marginales combinées des courbes avec les erreurs doublées et quadruplées, respectivement. Les fonctions marginales combinées de la troisième courbe ne sont pas montrées, puisqu'elles ressemblent exactement à celles montrées sur la figure III.3.8.

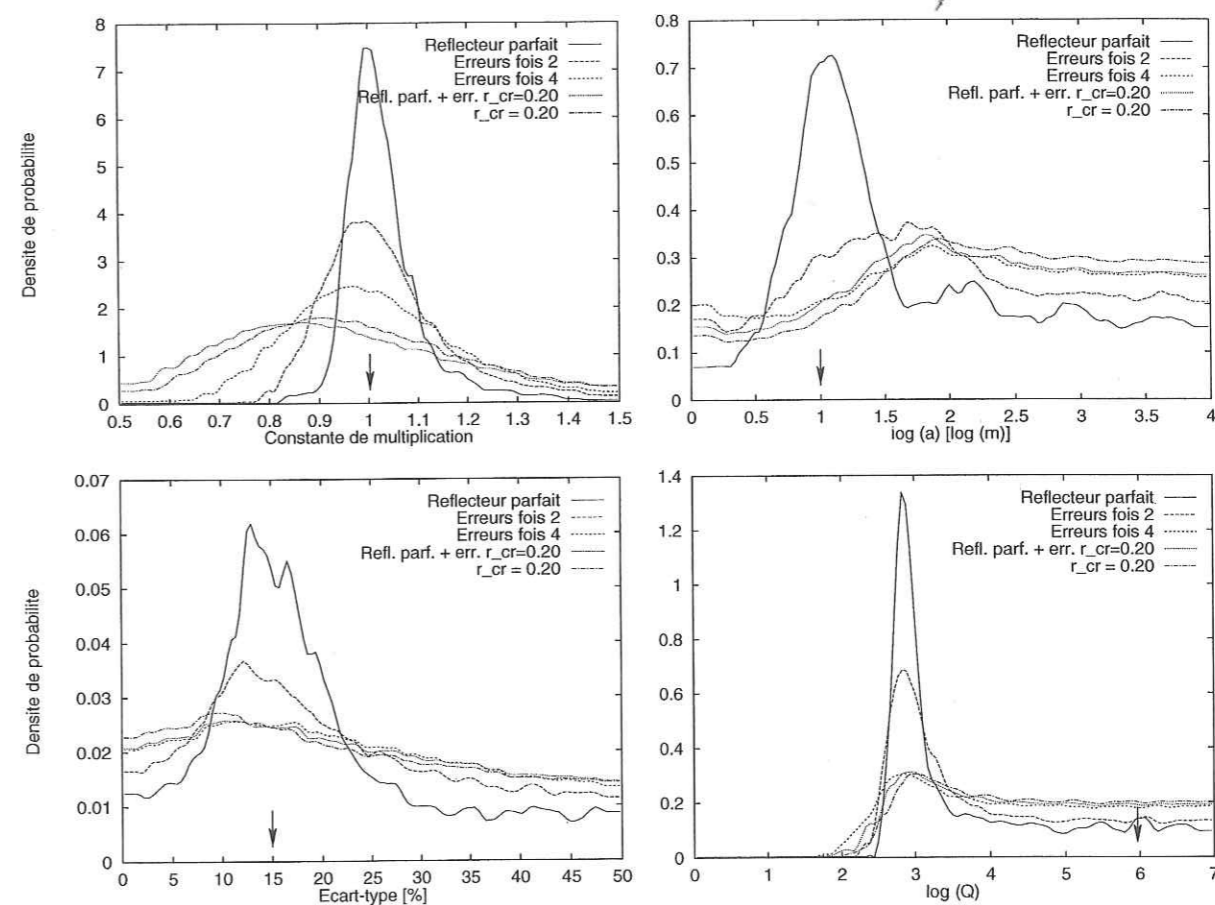


FIG. III.3.9 – Fonctions marginales individuelles de la densité de probabilité pour  $\sigma_{\kappa} = 15\%$  et respectivement un réflecteur parfait (reproduites de la figure III.3.5), un réflecteur parfait avec des erreurs doublées et quadruplées sur les mesures du rapport spectral d'un réflecteur parfait, puis le même rapport spectral mais avec les erreurs observées d'un réflecteur imparfait ( $\hat{r}_{cr} = 0.20$ , figure III.3.4.b) et enfin les résultats de  $\hat{r}_{cr} = 0.20$  (reproduits de la figure III.3.7).



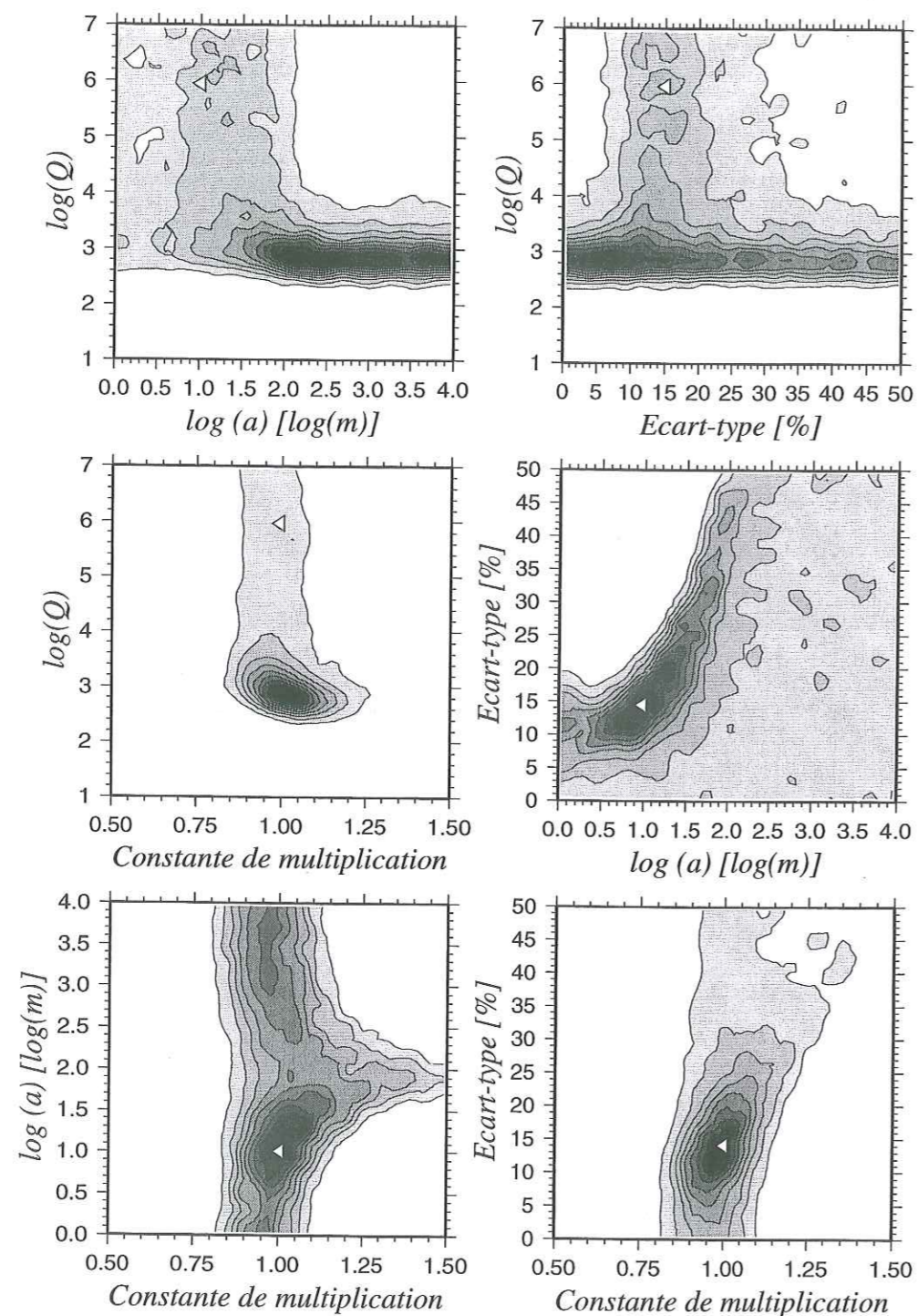


FIG. III.3.10 - Fonctions marginales combinées de la densité de probabilité pour  $\sigma_{\kappa} = 15\%$  et le rapport spectral d'un réflecteur parfait mais ayant des erreurs doublées. A comparer avec les figures III.3.6 et III.3.8.

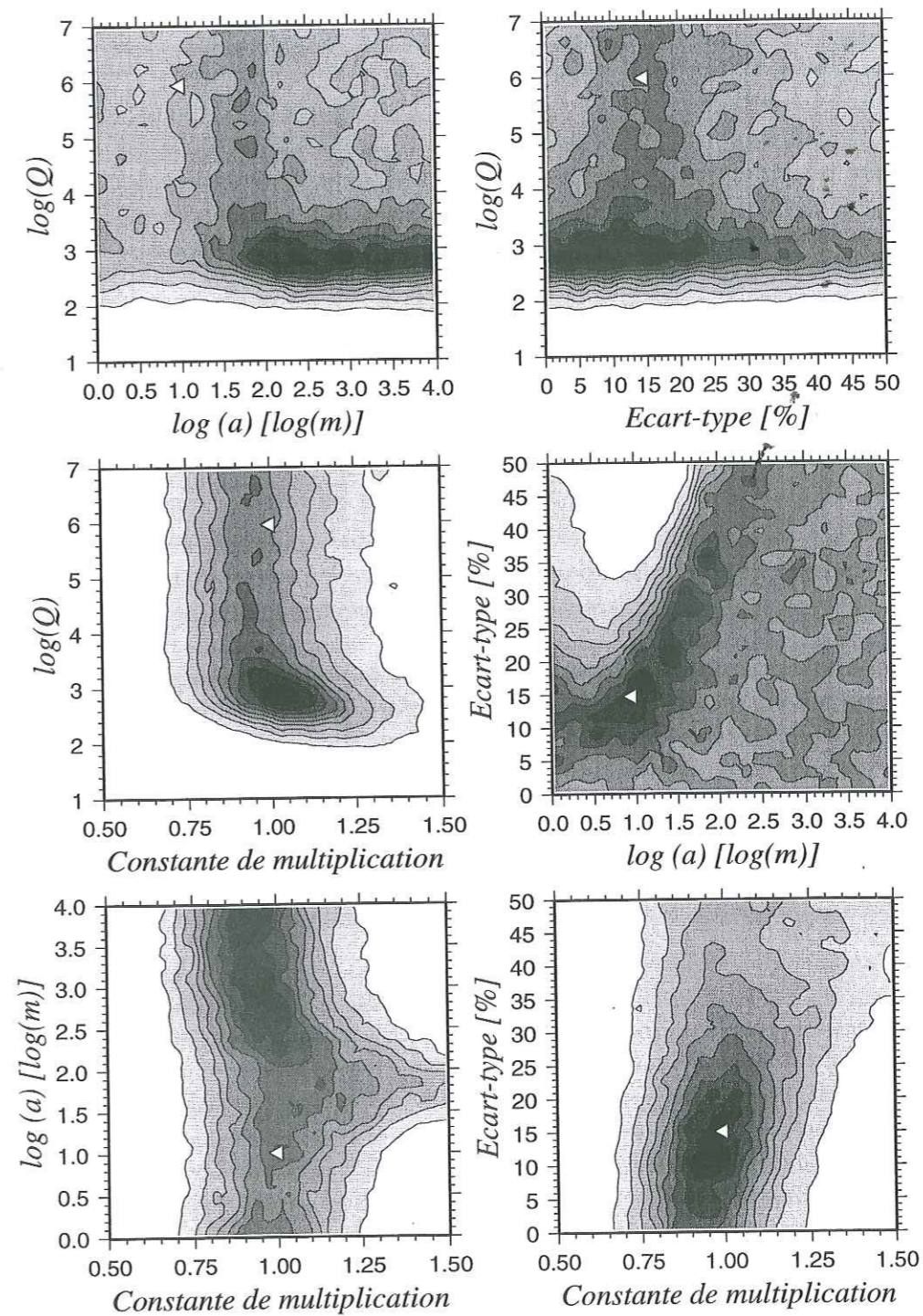


FIG. III.3.11 - Fonctions marginales combinées de la densité de probabilité pour  $\sigma_{\kappa} = 15\%$  et le rapport spectral d'un réflecteur parfait mais ayant des erreurs quadruplées. A comparer avec les figures III.3.6, III.3.8 et III.3.10.



Les figures III.3.10 et III.3.11 nous montrent clairement l'existence d'une transition entre les figures III.3.6 et III.3.8. Par exemple, la corrélation positive entre  $\log(a)$  et  $\sigma_\kappa$  (figure III.3.6) croît d'abord (figures III.3.10 et III.3.11) et devient finalement parabolique (figure III.3.8). Nous voyons aussi que la 'tâche' au-dessus de la corrélation positive entre  $c_{st}$  et  $\log(a)$  (figure III.3.6) grandit (figures III.3.10 et III.3.11) et qu'elle domine complètement à la fin (figure III.3.8). Finalement, il est clairement visible que la région contenant le maximum entre  $c_{st}$  et  $\sigma_\kappa$  grandit et qu'elle se déplace. Les figures nous indiquent clairement que cette transition est purement due à l'agrandissement des erreurs et non aux déviations systématiques.

La figure III.3.9 nous montre surtout que les fonctions marginales individuelles deviennent plus plates si les erreurs augmentent et que les maximums sont déplacés. Par contre, les différences entre les courbes pour  $\hat{r}_{cr} = 0.20$  et les courbes pour un réflecteur parfait ayant les erreurs d'un réflecteur avec  $\hat{r}_{cr} = 0.20$  sont très petites. Ceci nous indique à nouveau que les grandes incertitudes dans les vraisemblances individuelles maximales et leurs erreurs systématiques sont surtout dues aux larges incertitudes du rapport spectral et non aux déviations systématiques causées par les réflexions multiples tardives.

Mais d'un autre côté, les modèles aux vraisemblances maximales pour toutes les courbes utilisant le rapport spectral d'un réflecteur parfait se trouvent toujours autour de leurs vraies valeurs. Par contre, ceci n'est pas le cas si le rapport spectral subit des déviations systématiques. Par conséquent, nous avons bon espoir de retrouver le vrai modèle si nous sommes capables de corriger le rapport de toute déviation systématique causée par les réflexions multiples tardives.

Donc, les résultats de ces inversions indiquent clairement que les fonctions marginales individuelles et combinées peuvent donner des images trompeuses si les écarts-types du rapport spectral mesuré deviennent trop larges. Bref, il nous faut trouver des moyens pour les réduire ce qui correspond à réduire les effets aléatoires des réflexions multiples tardives. De plus, les résultats nous montrent aussi qu'il est impossible d'obtenir des estimations correctes des paramètres si nous ne corrigeons pas de leurs effets systématiques. Par conséquent, dans la suite, nous essayerons de parvenir à ces buts en examinant les méthodes théoriques proposées dans la section III.3.2.

### III.3.4.3 Examen des méthodes de correction proposées

Comme nous l'avons déjà dit plusieurs fois ci-dessus, les effets des réflexions multiples tardives ont un caractère statistique. Ceci a pour conséquence que les variations de  $|\hat{r}_{rmt}|$  sont à la fois aléatoires (pour une seule réalisation) et systématiques (en moyennant sur plusieurs réalisations). Le premier phénomène fait que les écarts-types des mesures du rapport spectral  $\hat{r}_{rsp,r}$  augmentent, alors que le second cause des déviations systématiques.

Il n'est presque pas possible de contourner les problèmes causés par leur caractère aléatoire. Heureusement, nous pouvons corriger au moins partiellement des déviations systématiques en utilisant les méthodes proposées dans la section III.3.2. Nous commencerons à examiner en détail le second problème et ensuite nous parlerons du premier.

**Déviations systématiques** Dans la section III.3.2 nous avons proposé trois méthodes pour prédire les déviations systématiques. La première technique consiste tout simplement à mesurer le spectre de bruit dans une partie d'une trace ne contenant que du 'bruit' (sans réflexion majeure). La seconde est due aux développements théoriques de Sheng et al. (1986b), White et al. (1987) et Sheng et al. (1990) sans surface libre (équations (III.3.12) et (III.3.16)). La troisième suppose que  $|\hat{r}_{rmt}| = \alpha|R|$  (équation (III.3.18)) où  $|R|$  dépend de  $|T|$  et donc de la longueur de localisation.

Sur la figure III.3.12 nous avons présenté le spectre de bruit mesuré sur des sismogrammes synthétiques après avoir corrigé de la divergence sphérique. Ce spectre est la moyenne des valeurs absolues des coefficients d'ondelette sur 15 points entre 0.4 et 0.6 s. Nous pouvons remarquer une divergence à basses fréquences du spectre de bruit. Cette divergence est due à la fois aux effets de bords de la Transformée Continue en Ondelette et à la présence de la réflexion majeure analysée. Les coefficients d'ondelette des plus basses fréquences sont influencés par cette réflexion, puisqu'une fenêtre d'analyse ne contient pas qu'une seule période mais plusieurs (4 à 5).

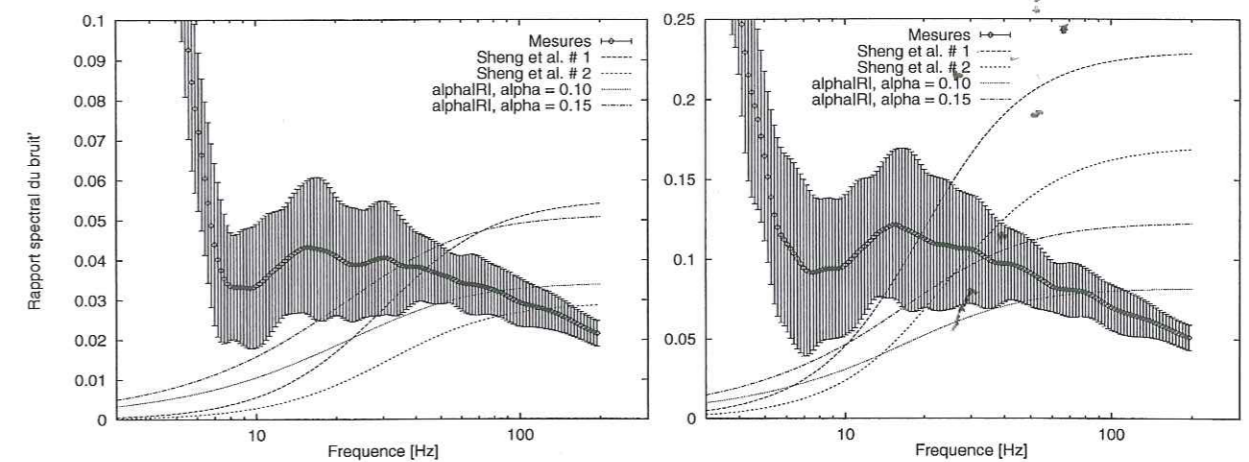


FIG. III.3.12 – Spectres de bruit mesurés et prédits par les théories pour un réflecteur parfait. a)  $\sigma_\kappa = 5\%$  et b)  $\sigma_\kappa = 15\%$ . Notez les échelles verticales différentes.

Sur la même figure nous avons montré les prédictions théoriques de Sheng et coauteurs et de  $\alpha|R|$  avec  $\alpha = 0.10$  et  $0.15$ . La différence entre les courbes marquées 'Sheng et al. # 1' et 'Sheng et al. # 2' est la vitesse utilisée pour calculer  $x$ . Dans le premier cas, nous utilisons  $c_0 = 4.0 \text{ km/s}$  (la vraie vitesse). Par contre, dans le second cas,  $c_0$  n'est que la moitié. La longueur de localisation n'a pas été adaptée. Nous avons utilisé ces deux vitesses pour distinguer entre 'temps double' ('Sheng et al. # 1') et 'temps simple' ('Sheng et al. # 2') (Papanicolaou, 1998).

Aucune des prédictions théoriques n'égale les valeurs mesurées, puisque par exemple les dernières montrent une valeur maximale à 13 Hz, alors que toutes les prédictions théoriques augmentent de façon monotone avec la fréquence. Néanmoins, en examinant les courbes de la figure III.3.12 de plus près, nous remarquons qu'au moins la méthode  $\alpha|R|$  donne des approximations consistantes, puisqu'elles sont pour  $\sigma_\kappa = 5\%$  et  $15\%$  du même ordre de grandeur que les mesures. Par contre, les prédictions de Sheng et al. '# 1' et '# 2' sont trop grandes pour  $\sigma_\kappa = 15\%$  (figure III.3.12.b), alors que ceci n'est pas le cas pour  $\sigma_\kappa = 5\%$  (figure III.3.12.a).

Sur la figure III.3.13 nous avons présenté le rapport spectral corrigé des effets des réflexions multiples tardives. Pour cela, nous avons calculé le rapport spectral mesuré moins le deuxième terme dans l'équation (III.3.21) en approximant  $|\hat{r}_{rmt}|$  par le spectre de bruit mesuré ou les prédictions théoriques. Si les corrections sont bonnes, les résultats doivent égaler les courbes théoriques de  $|\hat{r}_{rsp,r}|$ . Nous ne montrons que le cas où  $\hat{r}_{cr} = 0.20$ , puisque l'effet des réflexions multiples tardives est négligeable pour un réflecteur parfait (figures III.3.3.a et III.3.4.a).

D'abord, la figure III.3.13.a nous montre bien que les corrections effectuées pour  $\sigma_\kappa = 5\%$  n'influencent presque pas le rapport spectral définitif même si  $\hat{r}_{cr} = 0.20$ , puisque les effets des réflexions multiples tardives sont modestes. Ceci indique que nous pouvons probablement les négliger pour des milieux qui ne sont que légèrement hétérogènes ( $\sigma_\kappa = 5\%$  correspond à



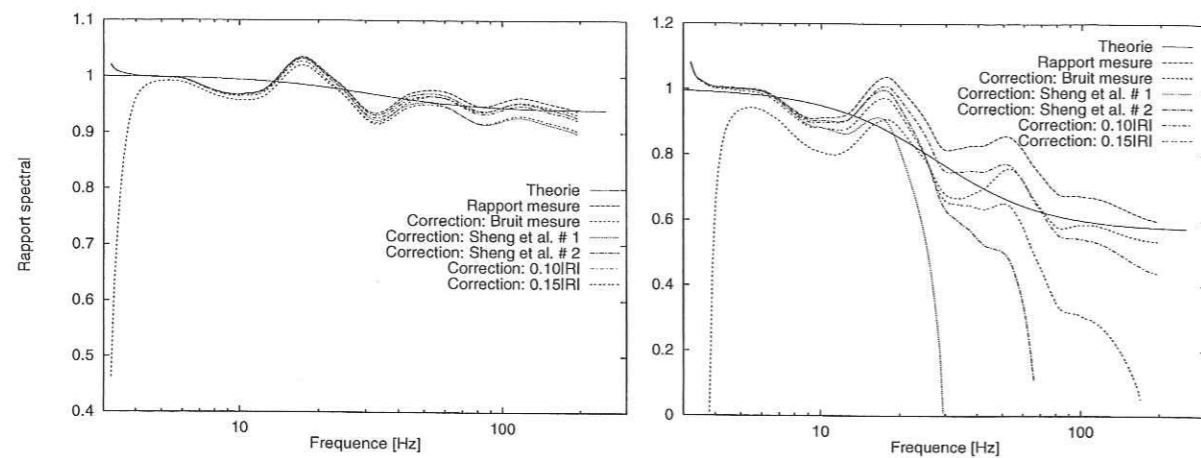


FIG. III.3.13 – Rapport spectral après corrections pour  $\hat{r}_{cr} = 0.20$ . a)  $\sigma_\kappa = 5\%$  et b)  $\sigma_\kappa = 15\%$ . Notez les échelles verticales différentes.

$\sigma_{c_0} = 2.5\%$  si la densité est constante). Les valeurs incorrectes à basses fréquences des corrections employant le bruit mesuré sont dues aux artefacts de la Transformée en Ondelette.

Par contre, si nous regardons les résultats des corrections pour  $\sigma_\kappa = 15\%$  et  $\hat{r}_{cr} = 0.20$  (figure III.3.13.b), nous voyons de grandes variations dans la qualité finale. Premièrement, nous remarquons que plusieurs courbes tendent vers zéro. Ceci est dû à des surestimations des effets des réflexions multiples tardives combinées avec la division par  $\hat{r}_{cr}^2$  dans le deuxième terme de l'équation (III.3.21). Donc, nous voyons de nouveau que les artefacts de la Transformée en Ondelette à basses fréquences ainsi que les prédictions de 'Sheng et al. # 1' et '# 2' et même  $0.15|R|$  à hautes fréquences résultent dans des valeurs aberrantes. De toutes les prédictions théoriques, il n'y a que la correction en  $0.10|R|$  qui donne d'assez bonnes approximations. Cependant, à partir de 15 Hz, les corrections effectuées avec le spectre de bruit mesuré sont les meilleures. Seulement, à basses fréquences, les courbes théorique et corrigée ne coïncident pas à cause des artefacts de la Transformée Continue en Ondelette (effets de bords et interférence avec la réflexion profonde). Par conséquent, si nous désirons utiliser cette technique avec des données réelles, il faudra tenir compte de ce problème.

**Déviations aléatoires** Quant à l'autre problème, nous avons vu que le caractère aléatoire de  $|\hat{r}_{rmt}|$  (dans une seule réalisation) fait que les écarts-types du rapport spectral augmentent de façon considérable. Malheureusement, il nous est pratiquement impossible de corriger de ce problème, puisqu'il faut non seulement connaître le spectre d'amplitudes, mais aussi le spectre de phase. Malheureusement, il n'existe aucune méthode fournissant une telle estimation. Nous ne pouvons même pas en obtenir une estimation en effectuant des mesures, puisque la phase est trop non-stationnaire.

La seule méthode capable de réduire légèrement les écarts-types du rapport spectral consiste à lisser toutes les mesures de l'atténuation apparente et du spectre de bruit pour chaque réalisation individuelle. Il nous est permis d'appliquer un tel lissage si les valeurs obtenues sur des fréquences voisines sont liées, c'est-à-dire s'il existe des corrélations en fréquence, ce qui est une approximation justifiée. Donc, nous avons lissé les mesures sur 20 points voisins (150 points sont montrés sur chaque figure).

Cependant, cette méthode produit surtout des courbes plus lisses et ne réduit pas les écarts-types (figures III.3.14 et III.3.15). Ceci est dû au fait qu'en lisant les courbes des réalisations particulières, nous ne nous approchons pas des valeurs théoriques (voir figure III.3.2), mais la

déviations autour des valeurs théoriques devient à peu près constante pour toutes les fréquences au lieu de montrer une dépendance aléatoire en fréquence. Pourtant, l'avantage est que les fluctuations de la courbe moyenne sont réduites considérablement (figure III.3.14). Les déplacements des courbes divergentes du spectre de bruit mesuré (figure III.3.15) sont des artefacts de lissage. En résumé, nous pouvons tirer plusieurs conclusions sur les effets des réflexions multiples tar-

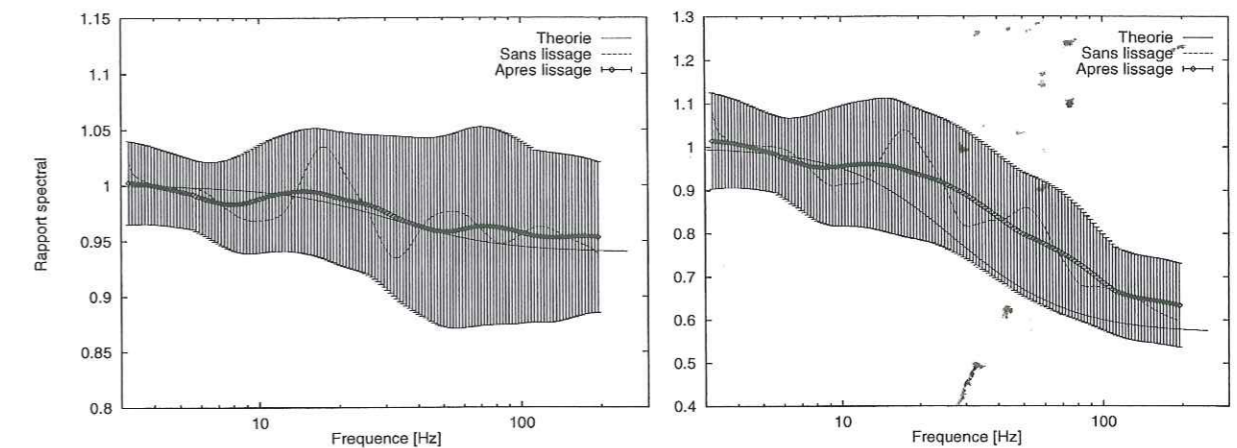


FIG. III.3.14 – Le rapport spectral mesuré (sans correction) après lissage pour  $\hat{r}_{cr} = 0.20$ . a)  $\sigma_\kappa = 5\%$  et b)  $\sigma_\kappa = 15\%$ . A comparer avec les figures III.3.3.b et III.3.4.b qui sont présentées à la même échelle. Les moyennes sans lissage sont aussi montrées.

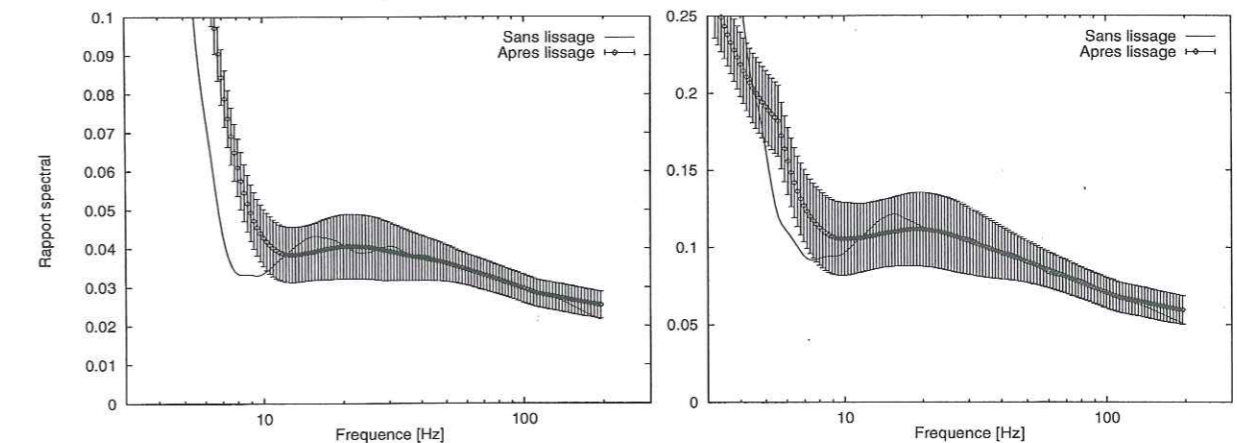


FIG. III.3.15 – Rapport spectral de bruit mesuré. a)  $\sigma_\kappa = 5\%$  et b)  $\sigma_\kappa = 15\%$ . A comparer avec la figure III.3.12 qui est montrée à la même échelle. Les moyennes sans lissage sont aussi montrées.

dives qui introduisent des erreurs systématiques dans les mesures du rapport spectral. D'abord, les déviations systématiques induites et leur erreurs relatives sont inversement proportionnelles au coefficient de réflexion du réflecteur profond. Elles dépendent aussi de la magnitude des hétérogénéités. Elles sont même négligeables si  $\sigma_\kappa$  est petit.

Ensuite, l'avantage d'un lissage est double: les erreurs de mesures finales et les variations de la courbe moyenne sont réduites. Finalement, les corrections des déviations systématiques effectuées avec le spectre de bruit mesuré donnent les meilleurs résultats, même s'il faut faire attention aux artefacts de la Transformée en Ondelette. De plus, les parts des mesures influencées



par ces artefacts peuvent être remplacées par des prédictions théoriques en utilisant de préférence celles de la méthode  $0.10|R|$ .

### III.3.4.4 Inversions finales

Dans la dernière partie de cette section, nous inversons les rapports spectraux corrigés afin d'estimer les paramètres statistiques décrivant le milieu hétérogène. D'abord nous montrons la qualité satisfaisante des courbes finales résultant des corrections combinées du spectre de bruit mesuré et la méthode  $0.10|R|$ . Puis, nous montrons les modèles obtenus par une méthode d'inversion locale.

**Rapport spectral corrigé** Comme nous avons vu dans la section précédente, les corrections du rapport spectral mesuré effectuées avec le spectre de bruit mesuré ont la meilleure qualité. De plus, nous pouvons utiliser les prédictions théoriques de la méthode  $0.10|R|$  pour remplacer ces parts du spectre de bruit contaminées par des artefacts de la Transformée en Ondelette.

Sur la figure III.3.16 sont présentées les rapports spectraux corrigés finaux, ainsi que ceux résultant du spectre de bruit et de la méthode théorique. Les corrections finales appliquées sont celles de la méthode  $0.10|R|$  pour  $f \leq 10$  Hz combinées avec celles du spectre de bruit pour  $f \geq 20$  Hz. Entre ces deux limites nous avons effectué une interpolation linéaire. Ces deux limites ont été choisies arbitrairement au minimum et maximum local du spectre de bruit (figure III.3.15). Nous voyons que cette combinaison donne d'assez bons résultats, puisque les courbes finales sont situées autour des rapports spectraux théoriques (figure III.3.16).

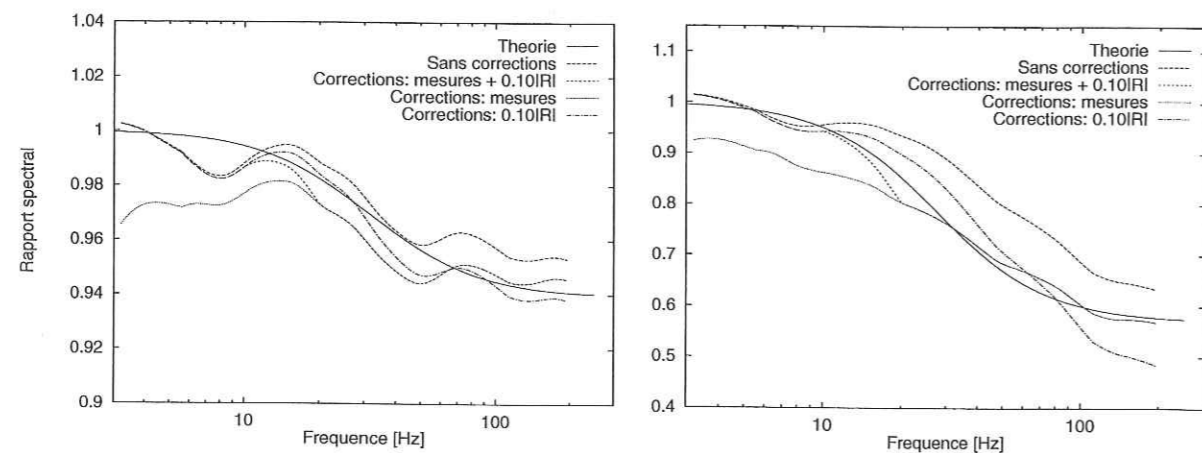


FIG. III.3.16 - Rapports spectraux corrigés pour  $\hat{r}_{cr} = 0.20$  après lissage. a)  $\sigma_{\kappa} = 5\%$  et b)  $\sigma_{\kappa} = 15\%$ . À comparer à la figure III.3.13.

**Détails sur la mise en œuvre des inversions locales** Afin d'inverser les courbes finales montrées sur la figure III.3.16, nous employons une technique d'inversion locale. La théorie correspondant à ce type de méthodes a été décrite dans l'annexe III.A. En résumé, nous essayons de minimiser le désaccord entre les observations et les prédictions théoriques en linéarisant l'influence d'une perturbation des paramètres du modèle sur ce désaccord.

La fonction à inverser ne sera ni le rapport spectral mesuré (équation (III.3.23)) ni le rapport spectral corrigé (le premier terme de cette équation), mais nous utilisons le logarithme naturel

des rapports corrigés, c'est-à-dire

$$f(\mathbf{m}) = \ln(c_{st}) - \gamma_{sc}z - \frac{1}{2}k_0zQ^{-1} \quad (\text{III.3.24})$$

où l'exposant de Lyapunov  $\gamma_{sc}$  est donné par la réciproque de l'équation (III.2.60) correspondant à la longueur de localisation d'un milieu exponentiel. Nous avons fait ce choix, car il entraîne une linéarisation de la courbe à inverser et donne ainsi de meilleurs résultats.

De nouveau, les paramètres du modèle sont  $c_{st}$ ,  $\log(a)$ ,  $\sigma_{\kappa}$  et  $\log(Q)$ . Ces paramètres ont été choisis pour qu'ils soient tous du même ordre de grandeur et que nous puissions limiter leurs variations sans artefact en introduisant un paramètre d'amortissement  $\beta$  unique (équations (III.A.6) et (III.A.7)). De plus, à cause de ce choix des paramètres et à cause du fait que nous connaissons les équations mathématiques du problème, nous pouvons calculer les dérivées de l'équation (III.3.24) par-rapport aux paramètres du modèle de façon analytique. Ces dérivées sont contenues dans la matrice  $\mathbf{A}$  (équation (III.A.2)).

Finalement, les nouvelles estimations des paramètres sont obtenues en résolvant l'équation (III.A.6). Nous n'avons pas employé l'équation (III.A.8), puisqu'elle donne des résultats sensiblement identiques pour les inversions des tests synthétiques. Ceci est dû à plusieurs raisons. Par exemple, les erreurs sur les données comme les paramètres du modèle sont toutes du même ordre de grandeur (figure III.3.14). Donc, nous n'avons pas besoin des matrices  $\mathbf{W}_M$  et  $\mathbf{W}_D$ . De plus, il est très difficile de corriger des corrélations non-linéaires entre les paramètres.

La solution de l'équation (III.A.5) a été trouvée par une méthode de gradients conjugués (VanDecar and Snieder, 1994). Les avantages de ce type de techniques ont déjà été décrits dans la section II.4.4.

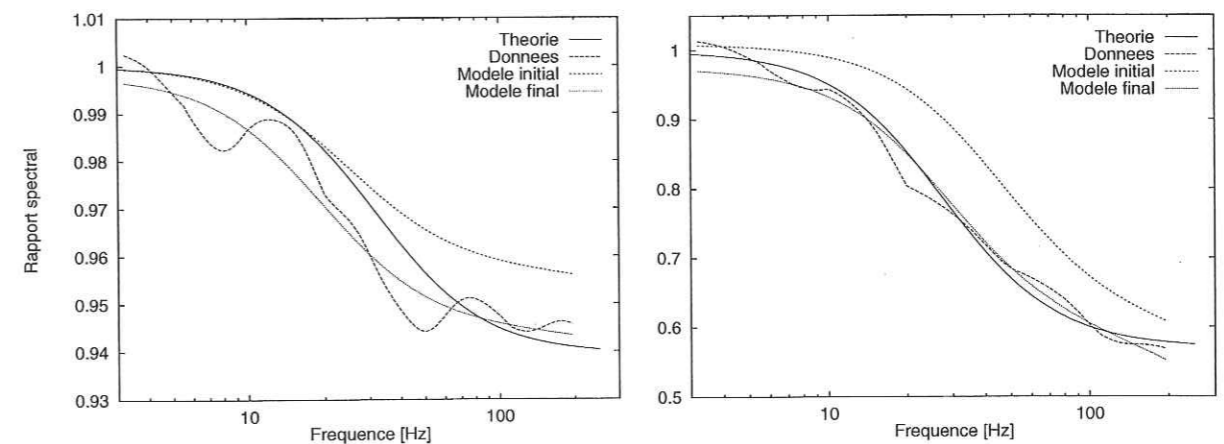


FIG. III.3.17 - Résultats de l'inversion du rapport spectral mesuré pour  $\hat{r}_{cr} = 0.20$ . a)  $\sigma_{\kappa} = 5\%$  et b)  $\sigma_{\kappa} = 15\%$ .

**Modèles finaux** La figure III.3.17 et le tableau III.3.3 nous montrent les résultats finaux des inversions locales ainsi que les modèles initiaux et théoriques. Le modèle initial est choisi comme le modèle à vraisemblance maximale issue des inversions globales en utilisant les rapports spectraux non-corrigés. C'est pour ce raison que les estimations initiales se trouvent au-dessus des courbes théoriques et corrigées du rapport spectral. Nous voyons que les prédictions théoriques après inversion coïncident bien avec les mesures corrigées. Cependant, le tableau III.3.3 nous apprend que les estimations obtenues des paramètres statistiques pour  $\sigma_{\kappa} = 5\%$  sont plutôt plus mauvaises que celles du modèle initial. Ceci est dû au fait que nous avons sur-corrigé le rapport



		$c_{st}$	$\log(a)$	$\sigma_\kappa$	$\log(Q)$
$\sigma_\kappa = 5\%$	Théorie	1.00	1.00	5.00	6.00
	Modèle initial	1.00	1.10	4.66	5.39
	Modèle final	1.00	1.21	5.91	5.39
$\sigma_\kappa = 15\%$	Théorie	1.00	1.00	15.00	6.00
	Modèle initial	1.01	0.79	10.97	3.97
	Modèle final	0.98	0.99	13.40	3.69

TAB. III.3.3 – Résultats des inversions des données synthétiques.

spectral pour ce milieu entre 5 et 70 Hz (figure III.3.17). D'un autre côté, les estimations pour le second milieu ont été améliorées.

Finalement, ceci indique que nous pouvons obtenir des estimations fiables des paramètres statistiques décrivant un milieu aléatoire en inversant le rapport spectral mesuré et corrigé. Mais pour des milieux qui ne sont que faiblement hétérogènes, nous pouvons probablement négliger les effets des réflexions multiples tardives et par conséquent les corrections prédites.

### III.3.5 Inversions de données réelles

#### III.3.5.1 Introduction: type de données

Après avoir testé l'aptitude de la théorie de localisation d'onde pour inverser de données en réflexion synthétiques dans la section précédente, nous allons essayer de l'appliquer sur des données réelles.

Comme dans le chapitre II, nous avons utilisé les données de l'expérience de sismique réflexion profonde Ecors 2.5-D enregistrées dans les Pyrénées Occidentales. La section II.2 nous montre bien que la géologie de cette partie des Pyrénées n'est pas caractérisée par des couches stratifiées. Par conséquent, les données Ecors 2.5-D ne sont pas les plus appropriées. Néanmoins, nous espérons que ce jeu de données nous permettra d'obtenir une idée du type de problèmes à résoudre.

Nous avons utilisé les 2 collections de traces à point de tir commun portant les numéros 8184 et 8194. Nous nous sommes déjà servis de la collection de traces 8184 dans les sections II.3, II.4 et II.5 afin de tester et de comparer de différentes techniques de reconnaissance de signaux. Le tir de la collection 8194 se trouve juste à côté et a été enregistré par les mêmes capteurs. Les localisations des 2 tirs sont montrées sur la figure II.2.4. Ils se trouvent dans la colonne 8 des sources juste à côté de la ligne 4 des capteurs.

Vue la faible distance entre ces 2 tirs (400 m), les résultats des inversions devraient être identiques. Par conséquent, ceci nous donne un bon test de l'inversion des données réelles. D'abord, nous décrivons la stratégie d'inversion suivie et ensuite les résultats obtenus.

#### III.3.5.2 Stratégie d'inversion

La stratégie appliquée ressemble beaucoup à celle utilisée dans la section précédente pour des données synthétiques. Elle est composée des étapes suivantes:

- D'abord nous pointons les signaux à analyser sur plusieurs traces, à savoir les premières arrivées, une réflexion dans la croûte supérieure et la réflexion sur le Moho. Ceci nous permet de caractériser non seulement la croûte entière, mais aussi les croûtes supérieure et inférieure.

- Ensuite, nous analysons le spectre de chaque pointé en effectuant une Transformée en Ondelette et nous calculons le spectre typique et les incertitudes de chaque type de signal. Contrairement au chapitre II et à la section précédente, nous n'utilisons pas le spectre moyen et les écarts-types, mais la valeur médiane et la déviation moyenne. La valeur médiane  $\bar{\mu}$  d'une série est celle qui est entourée d'autant de valeurs inférieures que de valeurs supérieures, tandis que la déviation moyenne  $\bar{\sigma}$  est définie comme la moyenne des différences absolues entre les valeurs d'une série et sa médiane, c'est-à-dire

$$\bar{\sigma} = I^{-1} \sum_{i=1}^I |x_i - \bar{\mu}|. \quad (\text{III.3.25})$$

Nous avons utilisé ces deux variables, car elles donnent des estimations plus stables que la moyenne arithmétique et l'écart-type (Press et al., 1992). C'est le cas, parce qu'elles sont beaucoup moins sensibles à des valeurs aberrantes, limitant ainsi l'influence des valeurs aberrantes sur les résultats des inversions.

- Maintenant que nous avons des estimations stables des spectres de 2 réflexions majeures et des premières arrivées, nous pouvons calculer leurs rapports spectraux. Comme nous l'avons déjà dit ci-dessus, ces 3 rapports spectraux nous permettent de caractériser et d'étudier la croûte entière (rapport Moho sur premières arrivées), la croûte supérieure et notamment les sédiments (rapport réflecteur majeur sous les sédiments sur premières arrivées) et finalement la croûte inférieure (rapport Moho sur réflecteur majeur).
- Finalement, les courbes obtenues sont inversées à l'aide de l'équation (III.3.23). Nous voulons obtenir 2 modèles extrêmes expliquant les rapports observés uniquement par une atténuation intrinsèque ou uniquement par une atténuation apparente causée par des diffractions multiples et un modèle où ces deux effets sont combinés.

#### III.3.5.3 Résultats

**Pointés** Sur les figures III.3.18 et III.3.19 sont montrées les collections de traces utilisées et les positions des pointés. La divergence sphérique a été corrigée par la méthode de Claerbout (1985) qui effectue une multiplication des amplitudes par un facteur proportionnel au temps au carré. Ensuite l'énergie des traces a été égalisée. Aucun autre traitement n'a été effectué (filtrage en fréquence, etc.).

Les figures III.3.18.b, III.3.18.c et III.3.18.d nous montrent avec plus de détails les positions des pointés sur les premières arrivées, la réflexion sur la rampe du Chevauchement de Sainte-Suzanne (réflecteur B sur la figure II.6.3) et la réflexion sur le Moho pour la collection de traces 8184. Les pointés correspondants pour la collection de traces 8194 sont montrés sur la figure III.3.19. Les figures III.3.18.c, III.3.18.d, III.3.19.c et III.3.19.d montrent clairement que les pointés ne sont pas faits sur des réflexions bien isolées. Par conséquent, même en utilisant une Transformée en Ondelette, nous ne pouvons éviter que les mesures soient influencées par l'énergie des signaux voisins. D'où l'intérêt d'utiliser des mesures plus stables comme la valeur médiane et la déviation moyenne.

**Mesures** Néanmoins, les spectres mesurés (figures III.3.20.a et III.3.20.c) correspondent à ce que nous attendons: le contenu en hautes fréquences diminue avec la profondeur. Par contre, le fait que tous les spectres ont à peu près la même valeur maximale indique que nous avons sur-corrigé de la divergence sphérique, puisque par exemple le spectre du Moho à basses fréquences ne



devrait pas dépasser celui des premières arrivées. Donc, la constante  $c_{st}$  dans l'équation (III.3.22) est supérieure à 1.0. D'un autre côté, la correction de divergence sphérique n'est pas effectuée sur le contenu fréquentiel, mais sur l'énergie totale présente à un certain temps d'arrivée, d'où la surestimation.

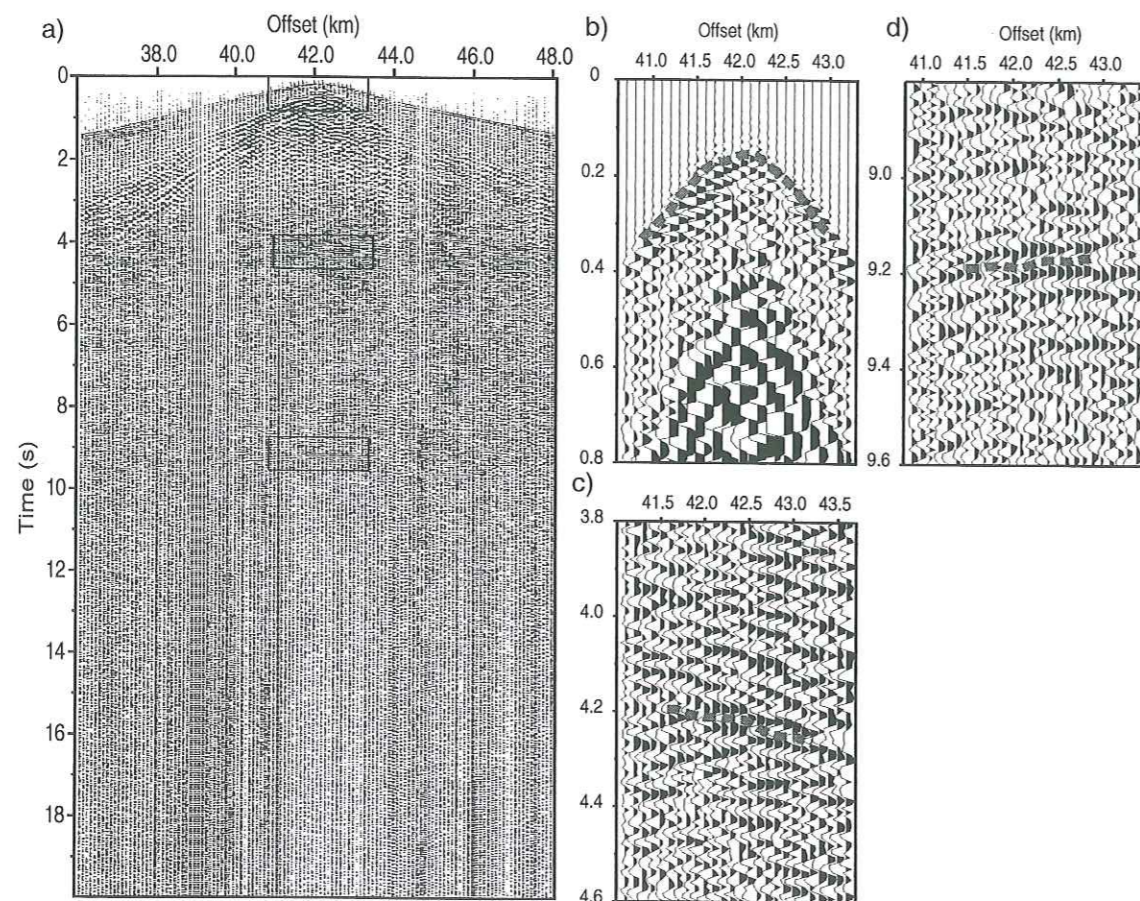


FIG. III.3.18 - a) Collection de traces à point à tir commun 8184 et positions des pointés sur b) les premières arrivées, c) une réflexion dans la croûte supérieure et d) le Moho.

Les figures III.3.20.a et III.3.20.c nous montrent aussi les spectres de bruit mesurés. Ces spectres sont obtenus en utilisant les traces où la réflexion sur le Moho a été pointée en calculant les valeurs médianes des coefficients d'ondelette sur 15 points entre 12.0 et 15.0 s. Puisque nous avons sur-corrigé de la divergence sphérique, nous avons probablement aussi surestimé le spectre de bruit. Par conséquent, nous avons décidé de ne pas corriger les rapports spectraux avec les spectres du bruit en espérant que l'erreur résultante sera inférieure aux incertitudes des mesures et donc négligeable.

Une inspection des figures III.3.20.a et III.3.20.c nous montre que les spectres mesurés des collections de traces 8184 et 8194 diffèrent. Ceci s'exprime à la fois par un déplacement des maxima et des différences relatives. Ces différences sont encore plus clairement visibles si nous regardons les rapports spectraux de toutes les phases (figures III.3.20.b et III.3.20.d). Pourtant, l'écart entre les tirs est seulement de 400 m. Donc, nous ne pensons pas que ces différences sont dues à la géologie, mais qu'elles sont purement causées par des contaminations dans les spectres estimés. Par conséquent, il est possible que les résultats des inversions soient fortement influencés par ces différences.

Deux autres aspects demandent notre attention. D'abord, contrairement aux prédictions de la théorie de l'homogénéisation, ce ne sont pas les basses fréquences qui sont les moins atténuées. Il existe des fréquences intermédiaires ayant des rapports spectraux supérieurs à ceux des basses fréquences indiquant que les basses fréquences sont plus atténuées.

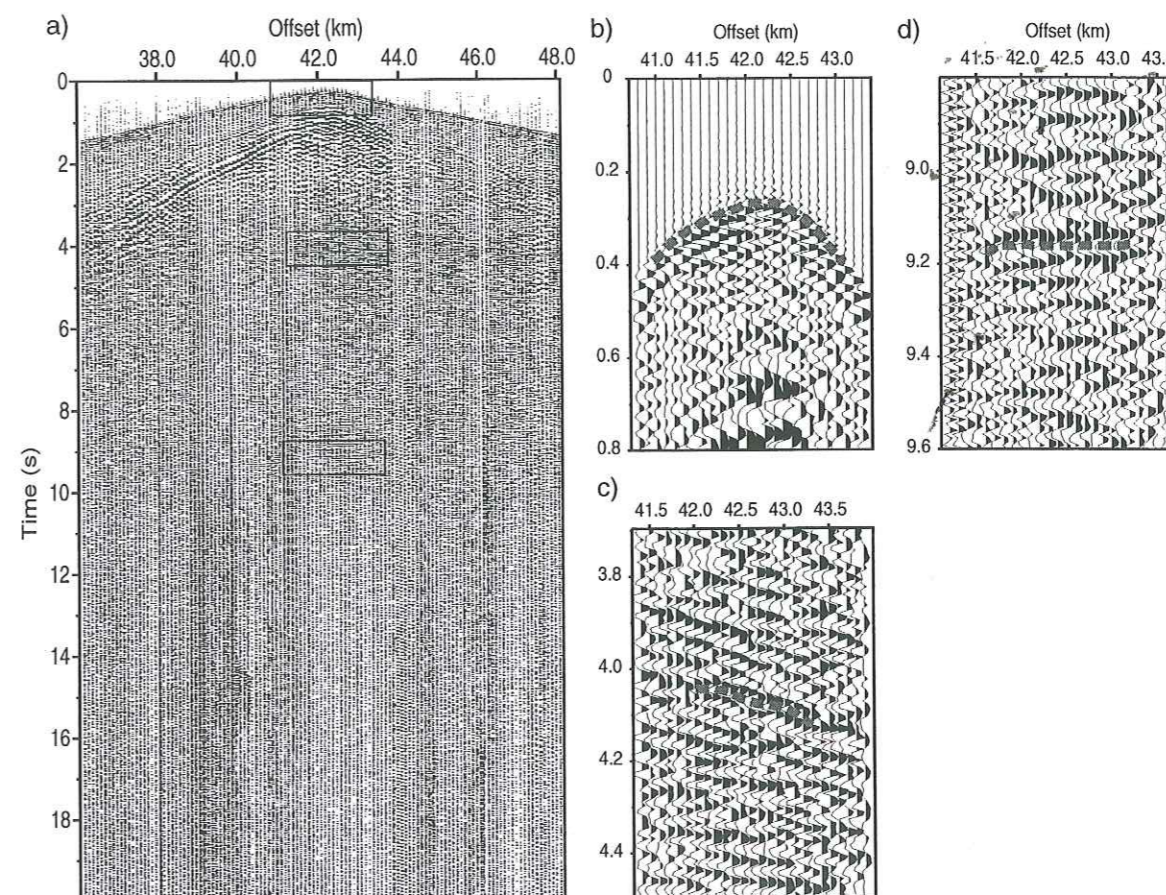


FIG. III.3.19 - Identique à la figure III.3.18, mais pour le groupe de traces 8194.

Nous pouvons citer plusieurs raisons expliquant ce phénomène. D'abord, peut-être s'agit-il simplement d'un artefact causé par du bruit. Il est aussi possible qu'il s'agisse d'un phénomène d'interférence (comme nous l'avons déjà vu avec les réflexions multiples tardives). Ensuite, il est possible que les basses fréquences se trouvent partiellement hors la gamme dynamique des capteurs. D'un autre côté, les ondes de surface sont enregistrées par ces mêmes capteurs et elles ont typiquement des fréquences entre 5 et 7 Hz. Enfin, un tel phénomène peut se produire si le milieu est lisse et composé de deux échelles d'hétérogénéités typiques. Il faut que le milieu soit lisse pour réduire la perte d'énergie des fréquences intermédiaires. En tous cas, ces courbes ne sont pas conformes à la théorie décrite. Donc, nous aurons du mal à les inverser proprement.

Le deuxième aspect curieux est le fait que les spectres mesurés ne convergent pas vers zéro autour de la fréquence de Nyquist (62.5 Hz). La raison n'en est pas claire, mais ce problème provient peut-être d'un sous-échantillonnage mal fait.

**Résultats des inversions** Puisque la théorie de la localisation d'onde ne peut expliquer les rapports spectraux mesurés (absence de basses fréquences), nous inversons chaque courbe deux



fois. La première fois nous utilisons la courbe entière (entre 5 et 58 Hz). Par contre, pour la deuxième inversion, nous n'employons qu'une partie sélectionnée. Cette partie est sélectionnée dans la gamme de fréquences où le spectre de bruit est inférieur aux autres spectres et où le rapport spectral ne passe pas par un maximum. Enfin, les plus hautes fréquences sont exclues à cause du problème de non-convergence vers zéro à 62.5 Hz. Par conséquent, la deuxième courbe inclue les fréquences 22 à 55 Hz pour la collection de traces 8184 et 15 à 55 Hz pour la collection 8194.

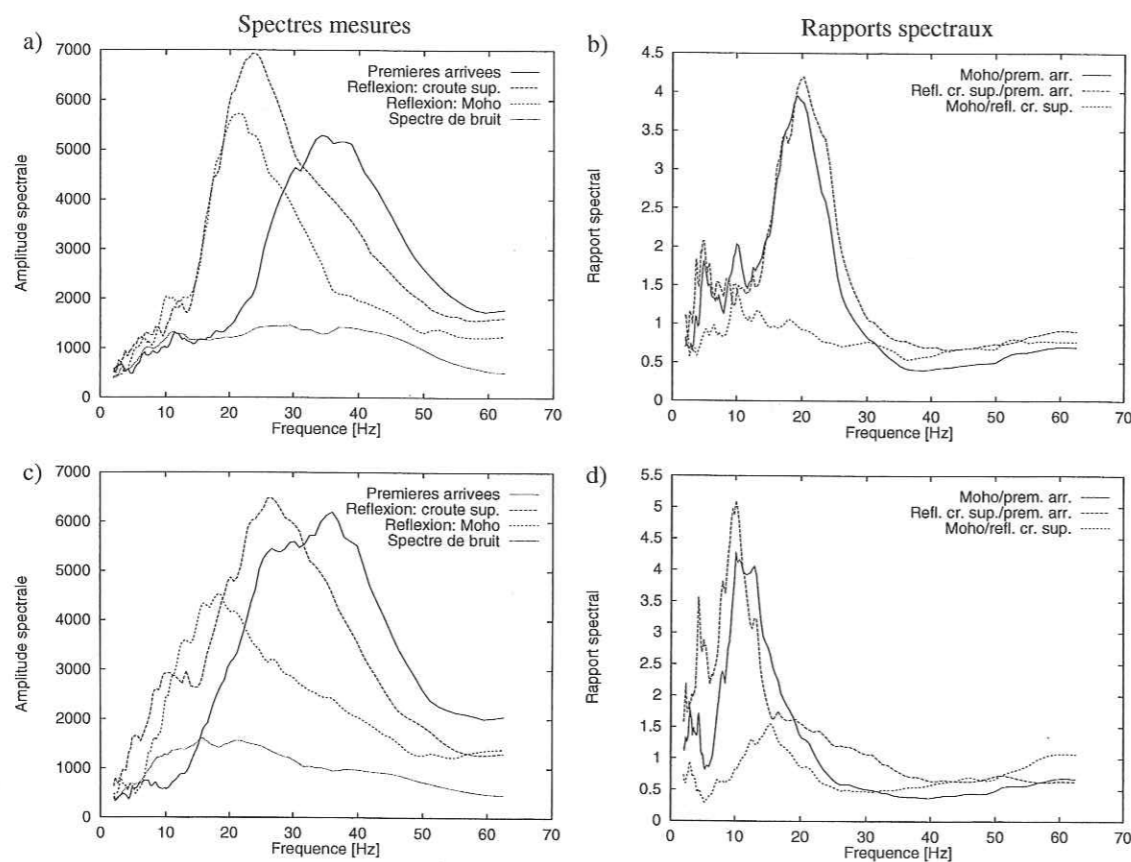


FIG. III.3.20 – Spectres mesurés (a et c) et rapports spectraux (b et d) pour les collections de traces 8184 (a et b) et 8194 (c et d).

Pour chaque courbe, nous désirons obtenir 5 modèles. Le premier modèle doit expliquer les observations par de l'atténuation intrinsèque pure, le second et le troisième purement par des diffractions multiples et les 2 derniers par ces deux effets combinés. Nous avons 4 modèles expliquant les données par des diffractions multiples, puisque les courbes nous font penser que les milieux sont lisses. Par conséquent, nous employons deux fonctions d'autocorrélation différentes, les fonctions exponentielle et Gaussienne. Puisque les rapports spectraux des hautes fréquences ne convergent pas vers zéro, nous n'utilisons pas de fonction d'autocorrélation pour décrire des milieux fractals.

Il nous reste deux remarques à faire. Afin de pouvoir calculer les rapports spectraux théoriques, nous utilisons les épaisseurs et les vitesses moyennes estimées par Siliqi (1994), c'est-à-dire 29 km et 6.4 km/s pour la croûte entière, 9 km et 4.6 km/s pour la croûte supérieure et enfin 20 km et 7.4 km/s pour la croûte inférieure. Finalement, les courbes sont inversées à l'aide de l'équation (III.A.8) où les éléments diagonaux de la matrice  $\mathbf{W}_D$  contiennent les réciproques des

déviation moyennes des rapports spectraux pour chaque fréquence. La matrice  $\mathbf{W}_M$  est égale à la matrice d'identité.

Sur les figures III.3.21 et III.3.22 et le tableau III.3.4 nous montrons les résultats des inversions. Nous voyons directement que les approximations des courbes entières laissent à désirer, puisque les résultats ne peuvent pas reproduire les maxima globaux.

Par contre, les courbes partielles ont beaucoup mieux été approximées. Il n'y a que le modèle  $Q_{in}$  (noté par  $Q$  sur les figures) qui ne réussit pas toujours bien à reproduire les observations. Sinon, les autres modèles donnent à peu près la même qualité d'approximation. On observe quand même que si les hautes fréquences divergent clairement (figure III.3.22.f), la fonction d'autocorrélation Gaussienne donne de meilleurs résultats qu'une fonction exponentielle.

Par contre, même si les courbes théoriques sont très similaires, les paramètres d'inversion sont caractérisés par une très grande diversité (tableau III.3.4). Ceci est partiellement dû au fait que la théorie décrite n'est pas adéquate pour expliquer les courbes observées. De plus, tous les paramètres sont corrélés ce qui cause encore plus de variations dans leurs estimations. Néanmoins, les moyennes des paramètres montrent des tendances systématiques. Ces moyennes sont contenues dans le tableau III.3.5. Elles sont calculées sans tenir compte des bandes de fréquences ou du type de fonction d'autocorrélation.

Les valeurs contenues dans le tableau III.3.5 nous montrent que les modèles extrêmes (expliquant les observations purement par de l'atténuation intrinsèque ( $Q$ ) ou par des diffractions multiples ( $D$ )) nous donnent bien évidemment des valeurs extrêmes pour l'atténuation intrinsèque (limite minimale) et de la magnitude caractéristique des perturbations des hétérogénéités (limite maximale). Les modèles combinés ( $D + Q$ ) ont un caractère plus modéré, mais ils tendent à faire augmenter la dimension typique des hétérogénéités. Puis, nous voyons que les paramètres estimés pour la croûte entière se trouvent entre ceux estimés pour les parties supérieure et inférieure, que la croûte supérieure est plus hétérogène et plus atténuante que la partie inférieure et enfin que la dimension typique des hétérogénéités augmente avec la profondeur. Par contre, il est étrange que la dimension des hétérogénéités ne dépasse pas l'échelle décimétrique. Malheureusement, les résultats ne donnent presque aucune contrainte sur la rugosité des hétérogénéités. Nous observons seulement qu'un milieu fractal est peu probable, puisque les rapports spectraux ne convergent pas vers zéro. Donc, des hétérogénéités lisses ou continues par morceau sont préférables.

### III.3.6 Discussion

Nous avons vu que nous pouvons appliquer la théorie de la localisation d'onde afin de déterminer par inversion les paramètres statistiques décrivant un milieu hétérogène. L'application à des données en transmission ne posera aucun problème. Par contre, l'utilisation de données en réflexion est plus compliquée, puisqu'il existe des réflexions multiples tardives qui interfèrent avec la réflexion analysée. Ces interférences ne sont pas décrites par la théorie. D'abord, nous traitons les résultats d'inversion des données synthétiques et ensuite nous comparons nos estimations sur des données réelles à celles d'autres études.

**Données synthétiques** Des tests synthétiques nous ont montré que nous pouvons obtenir de bonnes estimations des paramètres statistiques du milieu si nous savons corriger des effets des réflexions multiples tardives. Cependant, le problème principal n'est pas de déterminer leur spectre d'amplitude, mais leur vraie magnitude. Cette vraie magnitude s'exprime par le paramètre  $\mu$  dans l'équation (III.3.23) qui n'est ni connu ni facile à estimer. La cause principale de ce problème est le fait qu'il est difficile de corriger parfaitement de la divergence sphérique. Il est



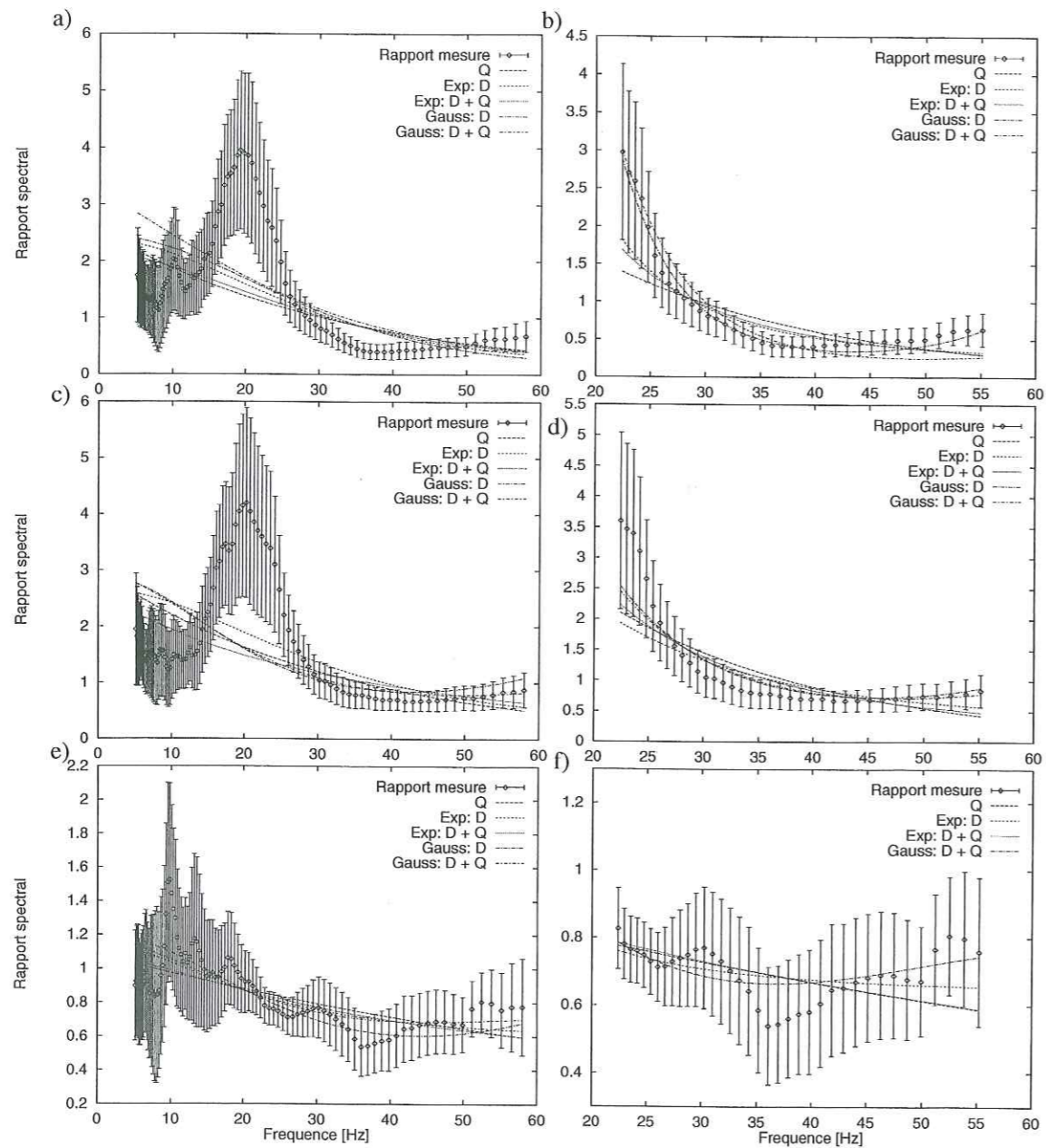


FIG. III.3.21 – Résultats des inversions pour la collection de trace 8184. a) et b) Rapports spectraux Moho sur premières arrivées, c) et d) rapports spectraux réflexion majeure sur premières arrivées et e) et f) rapports spectraux Moho sur réflexion majeure. Nous dénotons par  $Q$  le modèle n'ayant que de l'atténuation intrinsèque,  $D$  le modèle n'ayant que des diffractions multiples et  $D + Q$  les deux effets combinés.

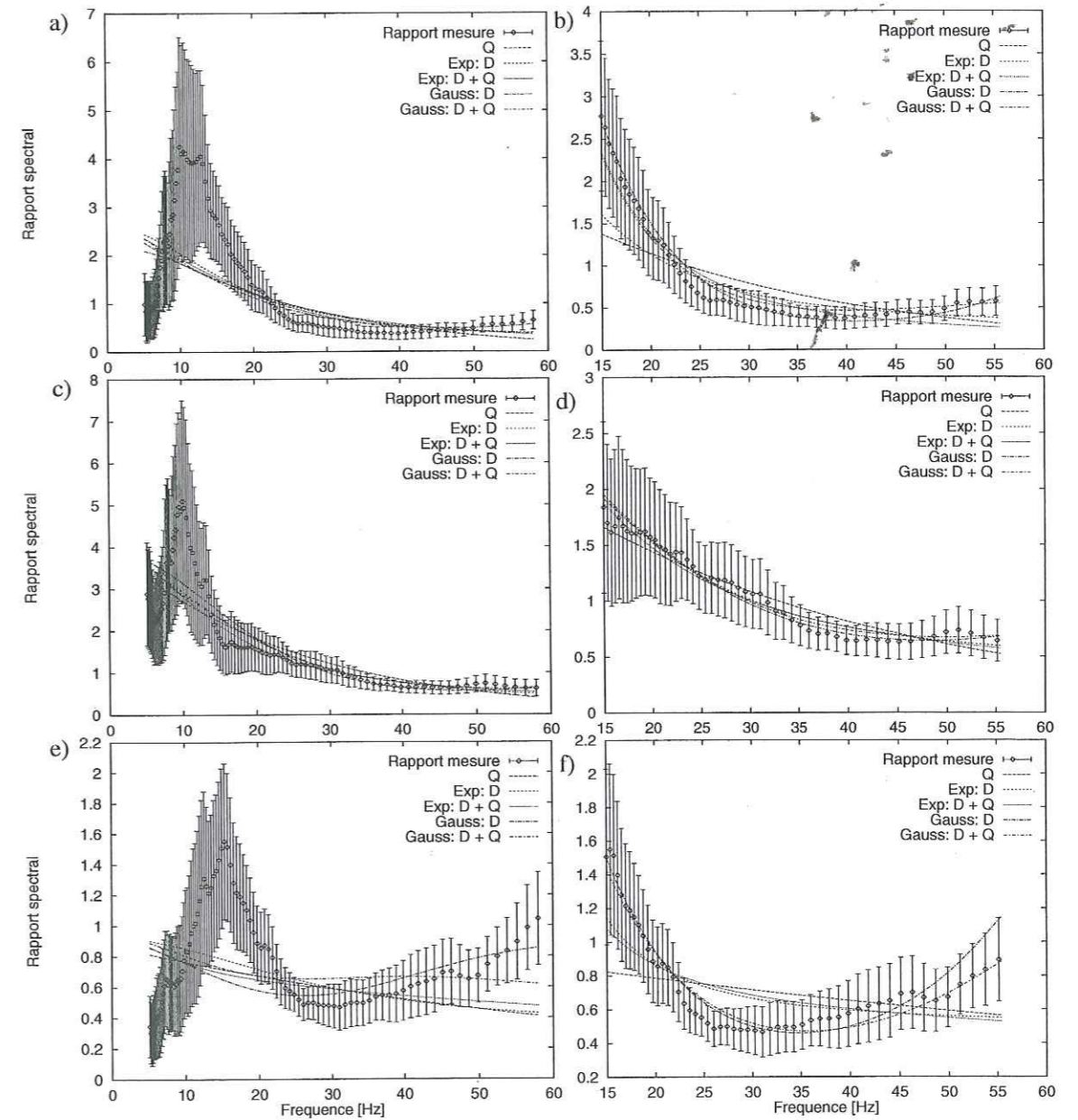


FIG. III.3.22 – Identique à la figure III.3.21, mais pour la collection de traces 8194.



Rapp. spectr.	Modèle	Tir	$f$ [Hz]	$c_{st}$	$a$ [m]	$\sigma_\kappa$ [%]	$Q_{in}$	Type de modèle	
Moho/src	$Q_{in}$ pure	8184	[5:58]	2.54	—	—	930	Q	
			[22:55]	4.02	—	—	600		
		8194	[5:58]	2.82	—	—	680		
			[15:55]	2.41	—	—	770		
	Exp.	8184	[5:58]	2.37	8.8	9.5	—	D	
				2.17	8.5	8.0	4600	D + Q	
			[22:55]	20.41	23.0	17.4	—	D	
				12.18	22.9	14.1	1590	D + Q	
		8194	[5:58]	2.62	17.2	10.7	—	D	
				2.69	18.7	8.1	1640	D + Q	
			[15:55]	3.53	22.2	12.7	—	D	
				9.36	25.4	15.6	2950	D + Q	
	Gauss.	8184	[5:58]	2.45	11.8	8.2	—	D	
				3.23	15.4	4.9	1240	D + Q	
			[22:55]	39.99	23.4	15.4	—	D	
				33.00	21.9	13.2	1220	D + Q	
		8194	[5:58]	2.20	21.6	8.3	—	D	
				2.80	27.0	5.8	1000	D + Q	
			[15:55]	7.20	24.8	12.7	—	D	
				8.00	31.0	10.1	790	D + Q	
	Refl/src	$Q_{in}$ pure	8184	[5:58]	2.99	—	—	410	Q
				[22:55]	6.28	—	—	250	
			8194	[5:58]	4.05	—	—	310	
				[15:55]	2.55	—	—	430	
Exp.		8184	[5:58]	2.65	6.7	13.1	—	D	
				2.50	20.8	6.6	630	D + Q	
			[22:55]	4.37	10.2	16.2	—	D	
				8.08	12.6	16.8	800	D + Q	
		8194	[5:58]	3.70	11.4	16.3	—	D	
				4.04	15.5	15.6	1420	D + Q	
			[15:55]	3.20	13.7	15.9	—	D	
				3.32	15.6	14.6	1520	D + Q	
Gauss.		8184	[5:58]	2.90	17.2	12.0	—	D	
				3.00	16.9	10.2	1260	D + Q	
			[22:55]	9.50	16.0	16.8	—	D	
				9.50	16.4	15.1	1000	D + Q	
		8194	[5:58]	3.90	15.1	14.1	—	D	
				3.80	15.9	10.6	790	D + Q	
			[15:55]	2.70	15.2	12.2	—	D	
				3.10	16.9	10.5	1000	D + Q	

TAB. III.3.4 - Résultats des inversions des données réelles des collections de traces 8184 et 8194. Les premières arrivées sont dénotées par l'étiquette 'src' et la réflexion majeure dans les sédiments par 'refl'.

Rapp. spectr.	Modèle	Tir	$f$ [Hz]	$c_{st}$	$a$ [m]	$\sigma_\kappa$ [%]	$Q_{in}$	Type de modèle
Moho/refl	$Q_{in}$ pure	8184	[5:58]	1.08	—	—	1610	Q
			[22:55]	0.94	—	—	1930	
		8194	[5:58]	0.87	—	—	1280	
			[15:55]	0.95	—	—	1750	
	Exp.	8184	[5:58]	1.14	24.3	8.2	—	D
				1.08	19.8	5.1	2840	D + Q
			[22:55]	1.48	50.9	13.2	—	D
				1.15	82.0	8.5	2010	D + Q
		8194	[5:58]	0.92	17.0	8.5	—	D
				0.95	59.4	8.7	2500	D + Q
			[15:55]	5.13	54.5	22.5	—	D
				4.30	61.7	21.4	3070	D + Q
	Gauss.	8184	[5:58]	1.20	27.5	7.6	—	D
				1.19	33.9	5.5	2510	D + Q
			[22:55]	1.17	31.8	7.5	—	D
				1.30	37.2	7.1	2510	D + Q
		8194	[5:58]	0.90	46.0	8.4	—	D
				0.95	55.9	5.8	2240	D + Q
			[15:55]	3.50	34.7	14.7	—	D
				3.30	35.3	13.4	2000	D + Q

TAB. III.3.4 - Suite.

croûte	$a$ [m]	$\sigma_\kappa$ [%]	$Q_{in}$	Type de modèle
entière	—	—	$750 \pm 140$	Q
	$19 \pm 6$	$12 \pm 3$	—	D
	$21 \pm 7$	$10 \pm 4$	$1880 \pm 1280$	D + Q
supérieure	—	—	$350 \pm 80$	Q
	$13 \pm 4$	$15 \pm 2$	—	D
	$16 \pm 2$	$13 \pm 4$	$1050 \pm 320$	D + Q
inférieure	—	—	$1640 \pm 270$	Q
	$36 \pm 13$	$11 \pm 5$	—	D
	$48 \pm 20$	$9 \pm 6$	$2460 \pm 380$	D + Q

TAB. III.3.5 - Moyennes des résultats des inversions des données réelles.



possible d'incorporer ce paramètre dans les inversions. Toutefois, ceci ne mène probablement pas à de bons résultats, puisque cette stratégie tend à introduire encore plus d'ambiguïtés. De plus, il n'est pas certain que ce problème joue un grand rôle dans les inversions de données réelles. Il est bien possible que d'autres facteurs contaminant les spectres estimés dominant comme le bruit et l'énergie des signaux présents dans la même fenêtre de temps.

Si nous ne corrigeons pas des réflexions multiples tardives, leur effet principal est d'augmenter l'énergie totale enregistrée, diminuant ainsi l'atténuation apparente. Par conséquent, elles tendent à diminuer les paramètres  $\sigma_\kappa$  et  $Q_{in}^{-1}$ . Elles modifient probablement aussi l'estimation du paramètre  $a$ .

**Données réelles** Il n'est pas évident de comparer les résultats obtenus à ceux trouvés par d'autres études utilisant des données et des méthodes différentes.

D'un côté, la plupart des études existantes utilisent des données d'ondes télésismiques ou d'ondes de surface. Ces études obtiennent généralement des estimations à l'échelle kilométrique et de l'ordre de quelques pourcents pour respectivement la dimension  $a$  et le magnitude  $\sigma_c$  typiques des perturbations de la vitesse (Herraiz and Espinosa, 1987). Malheureusement, ces valeurs ne sont pas représentatives de la croûte, mais de la lithosphère entière.

D'un autre côté, les estimations obtenues par analyse de diagraphies<sup>3</sup> nous donnent des informations sur les sédiments notamment et ainsi seulement sur la partie superficielle de la croûte supérieure. Ces études trouvent des valeurs typiques pour  $a$  entre 3 et 10 m et montrent que les perturbations des vitesses ne dépassent que rarement 10 % (Banik et al., 1985; Herraiz and Espinosa, 1987; White et al., 1990).

Finalement, toutes les valeurs dépendent forcément de la géologie régionale. Donc, de grandes variations doivent exister.

Afin de faciliter la comparaison des résultats obtenus à ceux d'autres études nous les rappelons ici. Le tableau III.3.5 montre que la croûte entière est caractérisée par  $Q_{in} \geq 750$ ,  $\sigma_c \leq 6\%$  et  $a \sim 20$  m. Ensuite, les croûtes supérieure et inférieure ont des valeurs légèrement différentes, à savoir  $Q_{in} \geq 350$ ,  $\sigma_c \leq 7.5\%$  et  $a \sim 15$  m pour la croûte supérieure et  $Q_{in} \geq 1600$ ,  $\sigma_c \leq 5.5\%$  et  $a \sim 40$  à 50 m pour la croûte inférieure (pour un milieu de densité constante  $\sigma_\kappa = 2\sigma_c$ ).

Ces valeurs sont en concordance avec l'opinion générale que la dimension typique des hétérogénéités augmente avec la profondeur, tandis que leur magnitude caractéristique décroît. De plus, on trouve en générale que le facteur de qualité est inversement proportionnel à la quantité d'hétérogénéités présente. Par conséquent, il est logique que les valeurs fournies par l'analyse de diagraphies et des données télésismiques diffèrent.

Ainsi, nos estimations se trouvent entre celles pour les sédiments et la lithosphère, puisque  $a$  est d'ordre de 15 à 50 m et  $\sigma_c$  est inférieur à 10 %. Par contre, il est possible qu'il existe plusieurs dimensions typiques non-détectables par la gamme de fréquences utilisée. Par exemple, nous avons noté que nous pouvons expliquer la présence d'un maximum dans les rapports spectraux observés si le milieu est lisse et s'il existe une deuxième échelle typique. En supposant que le milieu est Gaussien et en cherchant le minimum du rapport spectral, nous pouvons même en donner une valeur approchée, puisque la longueur de localisation a un minimum pour  $k_0a = 1$  pour un tel milieu (section III.2 et figure III.2.1). Par conséquent, nous trouvons une deuxième échelle typique de l'ordre de 200 à 300 m. De plus, nous voyons que nos données ne peuvent probablement pas donner des informations sur des échelles supérieures.

Les études employant des données de sismique réflexion profonde afin d'établir les paramètres  $a$ ,  $\sigma_c$  ou  $Q_{in}$  de la croûte sont plus rares. Parmi elles nous pouvons citer Rappin (1992), Hobbs

3. 'Well logs' en anglais

(1990), Pullammanappallil et al. (1997) et Line et al. (1998). Dans la suite, nous les traitons une par une.

Rappin (1992) a étudié la décroissance des amplitudes des traces de l'expérience Arzacq dans les Pyrénées Occidentales en calculant  $Q_{coda}$ . Ce paramètre décrit la dépendance fréquentielle de la partie exponentielle de la décroissance mesurée. Cette expérience est décrite dans la section II.2. De plus, le profil passe juste à côté de l'endroit où se trouvent aussi les collections de traces 8184 et 8194. Pour le Bassin de Mauléon il a trouvé des valeurs de  $Q_{coda}$  de l'ordre de 150 pour la croûte supérieure et entre 400 et 600 pour la croûte inférieure.

Malheureusement, nous ne pouvons comparer ces estimations de  $Q_{coda}$  aux nôtres pour  $Q_{in}$ , puisqu'il n'existe aucune relation théorique ou même empirique entre ces deux variables. Généralement, on suppose que  $Q_{coda}$  et  $Q_s$  (facteur de qualité des ondes S) mesurés à partir d'enregistrements sismologiques sont proportionnels voire identiques (Herraiz and Espinosa, 1987). Par contre, les estimations de Rappin (1992) sont faites sur des enregistrements sismiques composés principalement d'ondes P et non sur la coda de données sismologiques composées principalement des ondes de surface. Ceci a pour conséquence que ses estimations de  $Q_{coda}$  ne sont pas liées au facteur de qualité des ondes de cisaillement  $Q_s$ , mais au celui des ondes de compression  $Q_p$ . Ceci fait que la relation exacte entre ses estimations et les nôtres est inconnue. Cependant, ses estimations montrent que  $Q$  augmente avec la profondeur avec un facteur supérieur à 3.

Hobbs (1990) a estimé  $Q_{in}$  pour la croûte inférieure de l'Angleterre par une méthode similaire à la nôtre. Il a calculé le rapport spectral du contenu fréquentiel de signaux entre le toit et la base de la croûte inférieure. Ainsi, il a trouvé une valeur de  $500 \pm 200$  pour  $Q_{in}$  pour la croûte inférieure de l'Angleterre. Cette valeur est largement inférieure à la nôtre et nous ne pouvons expliquer cette différence que par une géologie différente.

Une étude calculant les paramètres statistiques  $a$  et  $\sigma_c$  en employant des données de sismique réflexion profonde est celle de Line et al. (1998). Dans cette étude, les paramètres statistiques sont estimés à l'aide d'une adaptation de la théorie de Chernov (1960) où des fronts d'onde sont maintenant approximé par des paraboloides (Flatté and Wu, 1988). Par conséquent, cette approximation n'est valable que si  $k_0a > 1$  et s'il s'agit de diffusion faible et à angle faible (vers l'avant).

Cette étude trouve  $a = 150 \pm 50$  m et  $\sigma_c = 1.5 \pm 0.5\%$  pour la croûte supérieure du Golfe de Botnie (Scandinavie) en utilisant des réflecteurs à 3 à 4 s (ce qui correspond à une profondeur maximale de 15 km). Le paramètre  $\sigma_c$  augmente jusqu'à 3 à 6 % s'ils limitent la présence des hétérogénéités à une couche superficielle de 2 km d'épaisseur.

Naturellement les différences entre leurs estimations des paramètres statistiques et les nôtres peuvent être dues à la géologie. D'un autre côté, il existe aussi des différences entre les méthodes appliquées qui peuvent partiellement entraîner des différences entre les résultats. D'abord, la méthode de Flatté et Wu (1988) suppose que le milieu a des hétérogénéités à 3 dimensions avec une distribution isotrope, alors que nous supposons que le milieu est stratifié. Donc, la première méthode suppose que  $a_x = a_y = a_z = a$  et la nôtre que  $a_x = a_y = \infty$  et que  $a_z = a$ . Ensuite, leur méthode n'est appropriée que pour un milieu à diffusion faible et non à diffusion forte comme la nôtre. Enfin, la théorie de localisation cherche les dimensions typiques des hétérogénéités contribuant le plus à l'atténuation observée. Cette dimension typique se trouve généralement dans le domaine de diffusion de type 'Mie' ( $k_0a \approx 1$ ), alors que leur méthode est limitée au domaine de diffusion de type 'Rayleigh' ( $k_0a > 1$ ). Par conséquent, notre méthode a sans doute plus tendance à réduire l'échelle typique des hétérogénéités.

Nous ne croyons ni que les hétérogénéités dans la croûte sont caractérisées par une distribution isotrope ni qu'elles sont stratifiées. Il est beaucoup plus probable qu'elles aient une distribution anisotrope où  $a_x$ ,  $a_y$  et  $a_z$  diffèrent. Malheureusement, il n'existe aucune méthode



permettant d'estimer ces paramètres. La méthode de Pullammanappallil et al. (1998) peut nous fournir une estimation de  $a_x$  par une analyse  $k_x-t$  du spectre de puissance, mais elle ne donne aucune information sur  $a_z$ . De plus, cette méthode est limitée à la diffusion faible. Par conséquent, il faut rechercher une méthode pouvant déterminer la distribution anisotrope des hétérogénéités non seulement pour des milieux à diffusion faible, mais aussi à diffusion forte.

### III.3.7 Conclusions

La théorie de la localisation d'onde décrit bien l'atténuation apparente d'une impulsion traversant un milieu aléatoire. Par contre, il est plus difficile de l'appliquer sur des données en réflexion, puisqu'elle ne tient pas compte des réflexions multiples tardives arrivant au même temps que par exemple une réflexion profonde. Toutefois, nous l'avons utilisée afin d'estimer les paramètres statistiques décrivant un milieu hétérogène en utilisant des données en réflexion. Nous avons fait ce choix, parce ce type de données est plus facile à acquérir et parqu'elles nous permettent de sonder la Terre jusqu'à une plus grande profondeur.

Des tests nous ont montré que les réflexions multiples tardives ont un caractère statistique. Ainsi, dans une seule réalisation du milieu, leurs effets systématiques sont dominés par des fluctuations aléatoires. Les effets systématiques ne peuvent être montrés qu'en moyennant sur un certain nombre de réalisations. Toutefois, elles tendent à faire augmenter l'énergie totale enregistrée et donc à réduire l'atténuation apparente. Ceci a pour conséquence que la magnitude des fluctuations des hétérogénéités et la réciproque du facteur de qualité de l'atténuation intrinsèque sont sous-estimées.

Des tests et des considérations théoriques nous ont montré que leur influence directe est proportionnelle à la magnitude typique des hétérogénéités et inversement proportionnelle au coefficient du réflecteur. Ceci veut dire que probablement elles ne jouent aucun rôle d'importance dans des milieux à diffusion faible et que l'incertitude sur les estimations est plus importante pour des réflexions mineures que majeures.

Nous pouvons obtenir une approximation de leur influence en mesurant le spectre de bruit sur des traces ne contenant pas de réflexions majeures. Toutefois, même si nous pouvons obtenir une estimation de la dépendance fréquentielle des réflexions multiples tardives, il sera très difficile de quantifier leur magnitude exacte à cause des corrections imparfaites de la divergence sphérique. Donc, souvent nous serons obligés de les négliger. Par conséquent, il faut rechercher des méthodes théoriques les prédisant. De plus, pour des données réelles, il est bien possible que leur influence soit inférieure à celle d'autres facteurs contaminants comme par exemple le bruit de fond et l'énergie des signaux voisins.

Finalement, nous avons appliqué la théorie de la localisation d'onde pour déterminer les paramètres statistiques dans la croûte sous le Bassin de Mauléon dans les Pyrénées Occidentales. Nous avons trouvé les valeurs suivantes pour la magnitude  $\sigma_c$  et la dimension  $a$  typiques des fluctuations des vitesses et l'atténuation intrinsèque  $Q_{in}$ :

- la croûte entière:  $a \sim 20$  m,  $\sigma_c \leq 6$  % et  $Q_{in} \geq 750$ ;
- la croûte supérieure:  $a \sim 15$  m,  $\sigma_c \leq 7.5$  % et  $Q_{in} \geq 350$ ;
- la croûte inférieure:  $a \sim 40$  à  $50$  m,  $\sigma_c \leq 5.5$  % et  $Q_{in} \geq 1600$ .

Même à l'incidence normale, les approximations faites (densité constante, milieu acoustique et stratifié et réflexions multiples tardives négligeables) ne sont pas complètement justifiées pour les Pyrénées. Toutefois, elles ne changeront probablement pas les résultats principaux montrant

que les fluctuations des vitesses diminuent et que leur dimension typique et l'atténuation intrinsèque augmentent avec la profondeur et ainsi que la croûte supérieure y est plus atténuante et plus hétérogène que la croûte inférieure. Ces résultats ne sont pas illogiques, puisqu'on trouve en générale que le facteur de qualité est inversement proportionnel à la quantité d'hétérogénéités présente.



### III.4 Conclusions

Le phénomène de la diffusion multiple dans des milieux hétérogène est très compliqué. Une approche statistique permet de simplifier la rigueur des mathématiques impliquées de façon considérable, mais elle ne peut pas garantir que les quantités prédites et observées coïncident. Par conséquent, il est difficile d'étudier ce phénomène par des simulations numériques à moins que nous ne considérons des quantités auto-moyennantes, puisque de telles quantités convergent vers leurs moyennes avec une vraisemblance de 1.

La théorie de la localisation d'onde est une théorie de diffusion forte qui utilise de telles quantités auto-moyennantes. Par conséquent, elle nous permet d'étudier l'influence de la présence d'hétérogénéités sur l'atténuation apparente et la dispersion d'une impulsion traversant un milieu aléatoire et de comparer les prédictions théoriques avec des simulations numériques.

**Prédictions théoriques** Dans un milieu acoustique et stratifié où la densité est constante, les prédictions théoriques sont confirmées par les simulations numériques sur presque toute la gamme de fréquences lorsque l'angle d'incidence reste inférieur à  $30^\circ$  et l'écart-type des fluctuations des vitesses ne dépasse pas 15 %. De plus, la théorie montre que, à l'exception des milieux fractals, l'atténuation apparente est maximale dans le domaine de la diffusion de type 'Mie' où l'échelle typique des hétérogénéités et la longueur d'onde ont le même ordre de grandeur. Par contre, dans des milieux fractals, aucune échelle typique n'existe et le facteur de qualité devient constant sur une large gamme de fréquences. Enfin, les vitesses de groupe et de phase sont en concordance avec la théorie d'homogénéisation et les relations de Kramers-Krönig. Donc, elles convergent vers les vitesses effective et géométrique pour respectivement les basses et les hautes fréquences. Toutefois, pour des fréquences intermédiaires, leur comportement dépend fortement de la rugosité du milieu et il existe plusieurs extrema et des phases d'Airy.

**Inférence statistique** Bien que la théorie de la localisation d'onde décrive bien l'atténuation et la dispersion d'une impulsion traversant un milieu aléatoire, elle est moins appropriée pour l'inférence statistique utilisant des données en réflexion. Ceci est dû au fait qu'il s'agit d'une théorie en transmission. Par conséquent, elle ne tient pas compte de réflexions multiples tardives arrivant au même temps que par exemple une réflexion profonde.

Probablement, l'influence des réflexions multiples tardives n'est que négligeable pour la diffusion faible. Elles tendent à augmenter l'énergie totale enregistrée et ainsi à réduire l'atténuation apparente. Ceci a pour conséquence que la magnitude caractéristique des hétérogénéités sera sous-estimée et le facteur de qualité de l'atténuation intrinsèque surestimé. Malheureusement, il est très dur de corriger les estimations de leurs effets.

Enfin, nous avons appliqué la théorie de la localisation d'onde sur des données Ecors 2.5-D afin de déterminer les paramètres statistiques de la croûte sous le Bassin de Mauléon. Ainsi nous avons pu estimer la magnitude  $\sigma_c$  et la dimension  $a$  caractéristiques des fluctuations des vitesses et le facteur de qualité de l'atténuation intrinsèque  $Q_{in}$ :

- la croûte entière:  $a \sim 20$  m,  $\sigma_c \leq 6$  % et  $Q_{in} \geq 750$ ;
- la croûte supérieure:  $a \sim 15$  m,  $\sigma_c \leq 7.5$  % et  $Q_{in} \geq 350$ ;
- la croûte inférieure:  $a \sim 40$  à  $50$  m,  $\sigma_c \leq 5.5$  % et  $Q_{in} \geq 1600$ .

Même à l'incidence verticale, les approximations faites (densité constante, milieu acoustique et stratifié et réflexions multiples tardives négligeables) ne sont pas complètement justifiées pour les Pyrénées. Toutefois, elles ne changeront probablement pas les résultats principaux montrant que



les fluctuations des vitesses diminuent et que leur dimension typique et l'atténuation intrinsèque augmentent avec la profondeur et ainsi que la croûte supérieure est plus atténuante et plus hétérogène que la croûte inférieure.

### III.A Méthode d'inversion locale

Cette annexe contient une description détaillée de la méthode appliquée pour obtenir une meilleure estimation des paramètres du modèle. Cette nouvelle estimation est obtenue en linéarisant l'influence d'une perturbation des paramètres du modèle sur le désaccord entre les observations et les prédictions théoriques. Ensuite, nous pouvons calculer quelle adaptation des paramètres produit un désaccord minimal.

Généralement, nous pouvons supposer que les quantités observées sont dues à un processus physique contaminé par du bruit. Ceci est exprimé par

$$\mathbf{d} = \mathbf{f}(\mathbf{m}) + \mathbf{e} \quad (\text{III.A.1})$$

où le vecteur  $\mathbf{d}$  représente les  $D$  données et  $\mathbf{e}$  le bruit. Le processus physique est décrit par la fonction  $\mathbf{f}$  qui dépend de façon linéaire ou non-linéaire des  $M$  paramètres du modèle  $\mathbf{m}$ . De plus, nous supposons qu'il existe une approximation mathématique connue de la fonction  $\mathbf{f}$  décrivant le processus physique et que le bruit ne dépend pas des paramètres du modèle.

Le but d'une inversion est d'obtenir une estimation des paramètres du modèle expliquant au mieux les valeurs observées. Afin de parvenir à ce but, nous faisons appel à une expansion de Taylor de la fonction  $\mathbf{f}$  autour d'une estimation initiale  $\mathbf{m}_0$  des paramètres du modèle ce qui donne

$$\mathbf{f}(\mathbf{m}_1) = \mathbf{f}(\mathbf{m}_0) + \nabla_{\mathbf{m}} \mathbf{f}(\mathbf{m}_0) \cdot \delta \mathbf{m} + \dots = \mathbf{f}_0 + \mathbf{A} \cdot \delta \mathbf{m} + \dots \quad (\text{III.A.2})$$

où  $\mathbf{m}_1$  sont les nouveaux paramètres du modèle et le symbole  $\nabla_{\mathbf{m}}$  représente les dérivées premières par-rapport à  $\mathbf{m}$ . Ces dérivées sont contenues dans la matrice  $\mathbf{A}$  qui est connue sous le nom de 'Jacobien'.

Il est évident que les dérivées premières sont calculées autour de l'estimation initiale  $\mathbf{m}_0$ . Par conséquent, la nouvelle estimation dépend directement de cette estimation initiale. De plus, nous supposons que cette estimation initiale est proche du minimum global afin que nous puissions négliger les dérivées secondes.

La nouvelle estimation peut être calculée en minimisant la différence entre les observations et les prédictions théoriques, à savoir

$$S(\mathbf{m}_1) = \sum_{d=1}^D (d_d - f_d(\mathbf{m}_1))^2. \quad (\text{III.A.3})$$

Ce minimum est défini par le point où

$$\frac{\partial S}{\partial m_m} = 0, \quad m = 1, \dots, M, \quad (\text{III.A.4})$$

ce qui résulte en

$$\delta \mathbf{m} = (\mathbf{A}^t \mathbf{A})^{-1} \mathbf{A}^t \delta \mathbf{d} \quad (\text{III.A.5})$$

$$\text{avec } \delta \mathbf{m} = \mathbf{m}_1 - \mathbf{m}_0 \\ \delta \mathbf{d} = \mathbf{d} - \mathbf{f}(\mathbf{m}_0)$$

Les équations (III.A.5) sont connues sous le nom des 'équations normales'.

Malheureusement, l'approche décrite ci-dessus a plusieurs inconvénients. Premièrement, des singularités présentes dans la matrice  $\mathbf{A}^t \mathbf{A}$  amplifieront la norme Euclidienne  $|\delta \mathbf{m}|$ , puisque cette norme est proportionnelle à l'inverse de cette matrice. Deuxièmement, cette approche ne



permet pas d'inclure des incertitudes sur les données et les paramètres du modèle (par exemple leurs écarts-types) et/ou des corrélations éventuelles.

Le premier inconvénient peut être évité en remplaçant  $(\mathbf{A}^t\mathbf{A})^{-1}$  par  $(\mathbf{A}^t\mathbf{A} + \beta\mathbf{I})^{-1}$  dans l'équation (III.A.5) où  $\beta$  une constante et  $\mathbf{I}$  la matrice identité. Ceci résulte en

$$\delta\mathbf{m} = (\mathbf{A}^t\mathbf{A} + \beta\mathbf{I})^{-1}\mathbf{A}^t\delta\mathbf{d}. \quad (\text{III.A.6})$$

Cette approche est connue sous le nom de régularisation du type 'Levenberg-Marquardt' ou 'Tikhonov' (Lines and Treitel, 1984; Van der Sluis and Van der Vorst, 1987).

Nous pouvons aussi inclure des incertitudes sur les données et les paramètres du modèle en pondérant l'équation (III.A.3) et en ajoutant un deuxième terme lié aux paramètres du modèle, résolvant ainsi le deuxième inconvénient. Ceci aboutit à la relation

$$S(\mathbf{m}_1) = \sum_{d_1=1}^M \left( \sum_{d_2=1}^D W_{D,d_1d_2} [d_{d_2} - f_{d_2}(\mathbf{m}_1)] \right)^2 + \beta \sum_{m_1=1}^M \left( \sum_{m_2=1}^D W_{M,m_1m_2} \delta m_{m_2} \right)^2. \quad (\text{III.A.7})$$

Les matrices  $\mathbf{W}_D$  et  $\mathbf{W}_M$  contiennent les poids appliqués respectivement sur les données et les paramètres du modèle.

La minimisation de cette équation nous donne

$$\delta\mathbf{m} = (\mathbf{A}^t\mathbf{W}_D^t\mathbf{W}_D\mathbf{A} + \beta\mathbf{W}_M^t\mathbf{W}_M)^{-1}\mathbf{A}^t\mathbf{W}_D^t\mathbf{W}_D\delta\mathbf{d} \quad (\text{III.A.8})$$

en utilisant à nouveau l'équation (III.A.4).

En remplaçant  $\mathbf{W}_D$  et  $\mathbf{W}_M$  par  $\mathbf{I}$  dans l'équation (III.A.8), nous obtenons à nouveau l'équation (III.A.6). Donc, nous voyons que l'approche Levenberg-Marquardt correspond à une minimisation simultanée du désaccord entre les observations et les prédictions théoriques et des modifications des paramètres du modèle. De plus, le paramètre  $\beta$  fixe l'influence relative entre ces deux termes.

Maintenant, nous pouvons inclure par exemple les incertitudes sur les données en remplaçant  $\mathbf{W}_D$  par une matrice diagonale contenant les réciproques des écarts-types des observations. De la même façon, la matrice diagonale  $\mathbf{W}_M$  peut contenir par exemple les réciproques des incertitudes sur les paramètres du modèle ou les réciproques des magnitudes de leurs estimations initiales. Cette dernière option est très utile si nous inversons pour des paramètres ayant des magnitudes très différentes afin d'éviter de ne limiter que les variations des plus grands paramètres ce qui peut introduire des ambiguïtés dans les estimations.

Finalement, nous pouvons aussi remplacer les matrices  $\mathbf{W}_D^t\mathbf{W}_D$  et  $\beta\mathbf{W}_M^t\mathbf{W}_M$  par les inverses des matrices de covariance des données  $\mathbf{C}_D^{-1}$  et des paramètres du modèle  $\mathbf{C}_M^{-1}$ , respectivement. Ceci aboutit à une approche Bayésienne. Donc, dans une telle approche, nous résolvons

$$\delta\mathbf{m} = (\mathbf{A}^t\mathbf{C}_D^{-1}\mathbf{A} + \mathbf{C}_M^{-1})^{-1}\mathbf{A}^t\mathbf{C}_D^{-1}\delta\mathbf{d}. \quad (\text{III.A.9})$$

Cette dernière approche nous permet d'inclure toutes les corrélations et (co-)variances existantes entre les données et les paramètres du modèle. Tarantola (1987) nous donne plus de détails sur cette approche.

En dernière remarque, il faut savoir qu'il est possible d'utiliser une approche adaptative. Ceci signifie que nous pouvons utiliser la nouvelle estimation  $\mathbf{m}_1$  des paramètres du modèle pour linéariser à nouveau l'influence d'une perturbation des paramètres du modèle sur le désaccord entre les observations et les prédictions. Une telle approche adaptative peut surtout nous aider si le problème considéré est fortement non-linéaire. Par contre, Dorren et Snieder (1995) ont démontré que cette approche mène aussi à une amplification non-linéaire du bruit. De plus, à

cause de cette amplification du bruit, l'inversion peut même devenir instable. Dans ce cas, la contamination à cause de l'amplification du bruit dépasse l'amélioration des résultats par les linéarisations successives.



### III.B Méthode d'inversion globale

L'algorithme Vegas (Lepage, 1978; Press et al., 1992) que nous allons décrire ci-dessus est originellement une technique d'intégration concentrant l'échantillonnage sur les régions contribuant au maximum à l'intégrale. Par conséquent, cet algorithme ne cherche pas les minima mais justement les maxima. Donc, si nous désirons employer cette technique comme une méthode d'inversion, nous ne pouvons pas utiliser le désaccord entre les observations et les prédictions théoriques comme norme (comme nous avons fait dans l'annexe III.A pour les inversions locales). Il nous faut chercher une autre norme. Par exemple, nous pouvons appliquer une approche statistique en cherchant le modèle ayant la vraisemblance maximale. Par conséquent, dans la suite, nous décrivons d'abord cette approche statistique et ensuite nous traitons l'algorithme Vegas utilisé dans les inversions globales.

#### III.B.1 Inversions Bayésiennes

Afin d'inverser pour les rapports spectraux par une méthode globale, nous faisons à nouveau appel à une approche Bayésienne. Une telle approche a déjà été utilisée dans la section II.3 afin de détecter et d'extraire des signaux particuliers. Par conséquent, nous renvoyons le lecteur à la section II.3.2.1.

Alors, nous pouvons calculer la vraisemblance d'un modèle particulier en appliquant la loi de Bayes (Bayes, 1763), équation (II.3.2), maintenant formulée comme

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d})}. \quad (\text{III.B.1})$$

Nous voyons que le vecteur d'attributs  $\mathbf{x}$  contient ici les données  $\mathbf{d}$  (c'est-à-dire, les rapports spectraux mesurés) au lieu des coefficients de Fourier. De plus, nous ne désirons pas savoir quelle est la vraisemblance qu'un certain rapport spectral appartienne à une classe particulière, mais quelle est la vraisemblance que le rapport spectral mesuré soit dû au modèle  $\mathbf{m}$ .

Néanmoins, tous les termes ont la même signification qu'avant. Donc,  $p(\mathbf{m}|\mathbf{d})$  représente la densité de probabilité conditionnelle *a posteriori* des paramètres du modèle étant données les observations  $\mathbf{d}$ . Puis, les fonctions  $p(\mathbf{m})$  et  $p(\mathbf{d})$  sont les densités de probabilité *a priori* des paramètres du modèle et des données, respectivement. Finalement,  $p(\mathbf{d}|\mathbf{m})$  représente la densité de probabilité des observations étant donnés les paramètres du modèle, c'est-à-dire la vraisemblance que ce modèle-là aboutisse aux mesures.

De nouveau, nous supposons que  $p(\mathbf{d}|\mathbf{m})$  est caractérisé par une distribution Gaussienne, qu'il n'existe pas de corrélations entre les observations et que nous pouvons traiter  $p^{-1}(\mathbf{d})$  comme une constante de normalisation. Par conséquent, l'équation (III.B.1) est transformée en

$$p(\mathbf{m}|\mathbf{d}) = cp(\mathbf{m})(2\pi)^{-D/2} \prod_{d=1}^D \sigma_d^{-1} \exp \left[ - (d_d - f_d(\mathbf{m}))^2 / 2\sigma_d^2 \right], \quad (\text{III.B.2})$$

où  $c$  représente le coefficient de normalisation et  $\sigma_d$  est l'écart-type de l'observation  $d_d$ .

En appliquant une méthode d'inversion globale, nous voulons résoudre l'équation (III.B.2). Malheureusement, la résolution exacte de cette équation demande une évaluation complète de l'espace des modèles. Par conséquent, ceci nécessite un très grand temps de calcul, surtout si l'espace des modèles est très grand ou si chaque évaluation de  $\mathbf{f}(\mathbf{m})$  est très compliquée ou lente. De plus, la fonction  $p(\mathbf{m}|\mathbf{d})$  peut être difficile à visualiser à cause de sa grande dimension.

Par conséquent, généralement, nous n'utilisons que le modèle moyenné  $\langle \mathbf{m} \rangle$ :

$$\langle \mathbf{m} \rangle = \int \mathbf{m} p(\mathbf{m}|\mathbf{d}) d\mathbf{m} \quad (\text{III.B.3})$$



et la matrice de covariances *a posteriori*  $C_M^{post}$ :

$$C_M^{post} = \int (\mathbf{m} - \langle \mathbf{m} \rangle)(\mathbf{m} - \langle \mathbf{m} \rangle)^t p(\mathbf{m}|\mathbf{d}) d\mathbf{m}. \quad (\text{III.B.4})$$

Pourtant, une telle approche est seulement recommandable si la fonction  $p(\mathbf{m}|\mathbf{d})$  est relativement simple (c'est-à-dire, non asymétrique et/ou multimodale), car sinon ces quantités perdent toute utilité. Pour des fonctions plus complexes, il vaut mieux calculer les fonctions marginales individuelles et combinées de la densité de probabilité. Les fonctions marginales individuelles sont définies par

$$p(m_i|\mathbf{d}) = \int p(\mathbf{m}|\mathbf{d}) dm_1 \dots dm_{i-1} dm_{i+1} \dots dm_M \quad (\text{III.B.5})$$

et les fonctions marginales combinées par

$$p(m_i, m_j|\mathbf{d}) = \int p(\mathbf{m}|\mathbf{d}) dm_1 \dots dm_{i-1} dm_{i+1} \dots dm_{j-1} dm_{j+1} \dots dm_M \quad (\text{III.B.6})$$

Les fonctions marginales de la densité de probabilité nous donnent la possibilité d'évaluer les incertitudes des paramètres du modèle, de montrer les corrélations existantes entre les divers paramètres et finalement d'éprouver l'existence de minima locaux. Par conséquent, nous pouvons les voir comme des matrices de covariances généralisées (Muyzert, 1998).

### III.B.2 L'algorithme Vegas: Echantillonnage par importance

Dans cette partie de l'annexe nous suivons surtout Lepage (1978) et Press et al. (1992) pour décrire la théorie d'intégration sur laquelle Vegas est basée et Van der Baan (1996) et Muyzert (1998) pour les adaptations nécessaires afin de le transformer en une méthode d'inversion globale.

**Méthode d'intégration** Le théorème fondamental employé par Vegas est que nous pouvons approximer l'intégrale d'une fonction  $f(\mathbf{m})$  sur le volume  $\Omega$  en additionnant les  $N$  échantillons  $\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_N$  si la distribution d'échantillonnage  $\rho(\mathbf{m})$  est connue. Donc, nous pouvons approximer l'intégrale

$$I \equiv \int_{\Omega} f(\mathbf{m}) d\mathbf{m} \quad (\text{III.B.7})$$

par

$$I = \int_{\Omega} \frac{f(\mathbf{m})}{\rho(\mathbf{m})} \rho(\mathbf{m}) d\mathbf{m} \approx \left[ \langle f/\rho \rangle \pm N^{-1/2} \sqrt{\langle f^2/\rho^2 \rangle - \langle f/\rho \rangle^2} \right] \int_{\Omega} \rho(\mathbf{m}) d\mathbf{m}. \quad (\text{III.B.8})$$

Dans cette équation l'opérateur  $\langle \cdot \rangle$  signifie la moyenne arithmétique sur les  $N$  échantillons. Par définition  $\int_{\Omega} \rho(\mathbf{m}) d\mathbf{m}$  est égale à 1. Donc, l'expression (III.B.8) se simplifie en

$$I \approx \langle f/\rho \rangle \pm N^{-1/2} \sqrt{\langle f^2/\rho^2 \rangle - \langle f/\rho \rangle^2} \quad (\text{III.B.9})$$

où le premier terme est l'estimation de l'intégrale, alors que le second nous fournit son écart-type.

En minimisant la variance de cette estimation, nous pouvons obtenir la distribution d'échantillonnage optimale  $\rho_{opt}(\mathbf{m})$ . Lepage (1978) et Press et al. (1992) montrent que cette distribution optimale est donnée par l'expression

$$\rho_{opt}(\mathbf{m}) = |f(\mathbf{m})| / \int_{\Omega} |f(\mathbf{m})| d\mathbf{m}. \quad (\text{III.B.10})$$

Donc, la distribution optimale concentre les échantillons dans les régions de l'espace total contribuant au maximum à l'intégrale, sans qu'elle néglige les autres régions. On appelle une telle distribution un échantillonnage par importance.

Malheureusement, afin de bien pouvoir employer un échantillonnage par importance parfait, il nous faut des connaissances détaillées de la fonction à inverser (ou intégrer). Mais, en général, nous ne disposons pas de telles connaissances.

Vegas évite ce problème en utilisant une approche itérative. Ainsi, après chaque itération, la distribution d'échantillonnage est adaptée selon la contribution relative de chaque région. De plus, l'algorithme utilise l'approximation que la distribution d'échantillonnage est séparable, c'est-à-dire

$$\rho(\mathbf{m}) = \rho_1(m_1)\rho_2(m_2) \dots \rho_M(m_M), \quad (\text{III.B.11})$$

où  $\rho_m(m_m)$  est la distribution d'échantillonnage à une dimension du paramètre  $m$  du modèle  $\mathbf{m}$ , c'est-à-dire l'axe  $m_m$ .

Dans ce cas particulier, l'équation (III.B.10) se transforme en

$$\rho_1(m_1) = \bar{f}(m_1) / \int_{m_1} \bar{f}(m_1) dm_1 \quad (\text{III.B.12})$$

avec

$$\bar{f}(m_1) = \left[ \int_{m_2} \int_{m_3} \dots \int_{m_M} \frac{f^2(\mathbf{m})}{\rho_2 \rho_3 \dots \rho_M} dm_2 dm_3 \dots dm_M \right]^{1/2} \quad (\text{III.B.13})$$

et des expressions similaires pour  $\rho_2, \rho_3, \dots, \rho_M$ . Ici nous avons abrégé  $\rho_m(m_m)$  comme  $\rho_m$ .

Afin que Vegas puisse traiter des fonctions continues, il divise chaque axe en  $H$  pavés. Par conséquent,  $\bar{f}(m_1)$  dans le pavé  $h$  est approximée par  $\langle |\bar{f}_{1,h}| \rangle \Delta m_{1,h}$  où est  $\Delta m_{1,h}$  la taille du pavé  $h$  de l'axe  $m_1$  et  $\langle \cdot \rangle$  la moyenne de la fonction  $|f(\mathbf{m})|$  dans le même pavé de cet axe.

Comme nous l'avons déjà dit ci-dessus, Vegas est un algorithme adaptatif. Donc, après chaque itération, il utilise les équations (III.B.12) et (III.B.13) pour mieux approximer la distribution d'échantillonnage optimale. La première itération utilise généralement une distribution uniforme.

La méthode a comme avantage qu'aucune connaissance sur la fonction à inverser n'est nécessaire. Par contre, l'algorithme devient beaucoup moins efficace si l'hypothèse III.B.11 est violée. Cette hypothèse signifie que les paramètres du modèle ne sont que faiblement corrélés. Pourtant, même si de fortes corrélations existent entre les paramètres, l'algorithme donne de bons résultats. Seulement, il lui faut plus d'itérations et d'échantillons.

**Application aux inversions** L'adaptation de l'algorithme à une méthode d'inversion globale est très simple. Au lieu de résoudre directement la fonction  $f(\mathbf{m})$ , nous appliquons la méthode sur la densité de probabilité conditionnelle *a posteriori*  $p(\mathbf{m}|\mathbf{d})$ , équation (III.B.2). De cette façon, nous utilisons l'efficacité de Vegas pour traiter des fonctions complexes et nous obtenons ainsi des résultats beaucoup plus rapidement que par exemple avec des algorithmes du type Monte Carlo pure (qui correspondent à un échantillonnage uniforme de l'espace des modèles).

De plus, dans ce cas, la fonction  $\bar{f}(m_1)$  dans l'équation (III.B.12) correspond à la fonction marginale individuelle de la densité de probabilité du paramètre  $m_1$  (équation (III.B.5)). Par conséquent, Vegas nous donne directement cette fonction désirée. Finalement, une modification mineure dans l'algorithme nous donne aussi la fonction

$$\rho_{i,j}(m_i, m_j) = \bar{f}(m_i, m_j) / \int_{m_i, m_j} \bar{f}(m_i, m_j) dm_i dm_j, \quad (\text{III.B.14})$$



ce qui correspond à la fonction marginale combinée, équation (III.B.6) (Van der Baan, 1996).

Finalement, afin de stabiliser les fonctions marginales obtenues, nous appliquons toujours un lissage des courbes avec un filtre de 5 points.

## Conclusions finales



## Chapitre IV

### Conclusions finales

Nous avons montré que les méthodes d'inférence statistique forment des alternatives intéressantes aux méthodes déterministes. Même si elles ne peuvent les remplacer, elles ont plusieurs avantages.

- D'abord, elles peuvent nous fournir une appréciation de la vraisemblance. Ceci peut nous aider dans le problème de la détection de signaux.
- Ensuite, une approche statistique mène souvent à une simplification des mathématiques impliquées, puisqu'elle n'est sensible qu'aux caractéristiques générales de l'ensemble des détails et il ne nous faut plus décrire chaque détail particulier du milieu et son influence sur le processus physique. Ceci peut nous aider à mieux étudier des phénomènes physiques caractérisés par une certaine rigueur mathématique.
- Enfin, puisque les méthodes statistiques ne dépendent que des caractéristiques typiques de l'ensemble des détails, elles peuvent sonder la Terre sur une échelle inférieure à celle atteinte par les méthodes déterministes. Par conséquent, elles peuvent augmenter la résolution.

Dans un premier temps, nous avons appliqué des méthodes d'inférence statistique sur le problème de la détection de signaux et ensuite sur la propagation d'ondes en milieux aléatoires.

**Détection de signaux** Nous avons montré qu'il est possible d'améliorer la qualité des données de sismique réflexion profonde à l'aide de techniques automatiques de détection et de reconstruction de signaux. Cette qualité améliorée facilitera ensuite l'interprétation des données.

**Détection** Nous avons testé deux méthodes de détection, la méthode de Classification par Vraisemblance Maximale et les Réseaux de Neurones. Ces deux techniques ont des avantages et inconvénients opposés.

- La Classification par Vraisemblance Maximale est facile à mettre en œuvre. Par contre, elle donne de mauvais résultats si l'on ne connaît pas la distribution statistique sous-jacente des différentes catégories de signaux. Ceci pose surtout des problèmes dans le cas où la vraie distribution sous-jacente est multi-modale ou asymétrique, puisqu'elle doit être estimée par des approximations successives.
- Les Réseaux de Neurones sont capables d'estimer cette distribution sous-jacente de façon automatique, mais souvent il est plus difficile de les mettre en œuvre.



**Reconstruction** L'inconvénient principal des techniques décrites de détection est qu'elles ne peuvent fournir aucune estimation de toute énergie 'manquante'. Ceci requiert une technique de reconstruction. Dans cette étude, nous avons appliqué des sommations obliques pondérées par la semblance afin de reconstruire cette énergie 'manquante'. Cette méthode donne de bons résultats après l'extraction d'un signal particulier. Cependant, dans le cas des données Ecors 2.5-D, il serait très utile d'appliquer une extension de cette technique à 3 dimensions.

**Localisation d'onde** Afin d'étudier le problème de la propagation d'ondes en milieux aléatoires, nous avons fait appel à la théorie de la localisation d'onde. Ceci est une théorie de diffusion forte qui utilise des quantités auto-moyennantes. Par conséquent, elle nous permet de quantifier l'influence de la présence d'hétérogénéités sur l'atténuation apparente et la dispersion d'une impulsion traversant un milieu aléatoire et de comparer les prédictions théoriques avec des simulations numériques.

**Prédictions théoriques** Dans un milieu acoustique et stratifié où la densité est constante, les prédictions théoriques sont confirmées par les simulations numériques sur presque toute la gamme de fréquences lorsque l'angle d'incidence reste inférieur à  $30^\circ$  et l'écart-type des fluctuations des vitesses ne dépasse pas 15 %. De plus, la théorie montre que l'atténuation apparente est maximale dans le domaine de la diffusion de type 'Mie' pour des milieux non-fractals. Finalement, les vitesses de groupe et de phase sont en concordance avec la théorie de l'homogénéisation et les relations de Kramers-Krönig. Toutefois, pour des fréquences intermédiaires, leur comportement dépend fortement de la rugosité du milieu.

**Inférence statistique** Bien que la théorie de la localisation d'onde décrive bien l'atténuation et la dispersion d'une impulsion traversant un milieu aléatoire, elle est moins appropriée pour l'inférence statistique utilisant des données en réflexion. Ceci est dû au fait qu'elle ne tient pas compte de réflexions multiples tardives qui tendent à augmenter l'énergie totale enregistrée et ainsi à réduire l'atténuation apparente. Par conséquent, la magnitude caractéristique des hétérogénéités sera sous-estimée et le facteur de qualité de l'atténuation intrinsèque surestimé. Malheureusement, il est très dur de corriger les estimations de leurs effets.

Enfin, nous avons appliqué la théorie de la localisation d'onde sur des données Ecors 2.5-D afin d'estimer la magnitude  $\sigma_c$  et la dimension  $a$  caractéristiques des fluctuations des vitesses et le facteur de qualité de l'atténuation intrinsèque  $Q_{in}$  de la croûte des Pyrénées Occidentales. Elle est caractérisée par

- la croûte supérieure:  $a \sim 15$  m,  $\sigma_c \leq 7.5$  % et  $Q_{in} \geq 350$ ;
- la croûte inférieure:  $a \sim 40$  à  $50$  m,  $\sigma_c \leq 5.5$  % et  $Q_{in} \geq 1600$ .

Même si probablement les approximations faites ne sont pas complètement justifiées pour les Pyrénées, les résultats nous montrent que, dans cette région, les fluctuations des vitesses diminuent et que leur dimension typique et l'atténuation intrinsèque augmentent avec la profondeur et ainsi que la croûte supérieure y est plus atténuante et plus hétérogène que la croûte inférieure.

## Références



## Références

- Abma, R., and Claerbout, J., 1995, Lateral prediction for noise attenuation by t-x and f-x techniques: *Geophysics*, **60**, 1887-1896.
- Abrahams, E., Anderson, P. W., Licciardello, D. C., and Ramakrishnan, T. V., 1979, Scaling theory of localization: Absence of quantum diffusion in two dimensions: *Phys. Rev. Lett.*, **42**, 673-676.
- Abramowitz, M., and Stegun, I. A., 1972, *Handbook of mathematical functions*: Dover Publications.
- Akaike, H., 1970, Statistical predictor identification: *Ann. Inst. Statist. Math.*, **22**, 203-217.
- Almeida, L. B., 1994, Neural preprocessing methods, in Cherkassy, V., Frieman, J. H., and Wechsler, H., Eds., *From statistics to neural networks: Theory and pattern recognition applications*: Springer-Verlag, Nato ASI Series F: Computer and systems sciences, 213-225.
- Alsdorf, D., 1997, A simple method for migrating narrow aperture, noisy seismic reflection data and application to project INDEPTH (INTERNATIONAL DEEP PROFILING OF TIBET AND THE HIMALAYA) deep seismic profiles: *J. Geophys. Res.*, **102**, 17807-17811.
- Anderson, P. W., 1958, Absence of diffusion in certain random lattices: *Phys. Rev.*, **109**, 1492-1505.
- Asch, M., Kohler, W., Papanicolaou, G., Postel, M., and White, B., 1991, Frequency content of randomly scattered signals: *SIAM Review*, **33**, 526-629.
- Backus, G. E., 1962, Long-wave elastic anisotropy produced by horizontal layering: *J. Geophys. Res.*, **67**, 4427-4440.
- Bahorich, M. S., and Farmer, S. L., 1995, 3-D seismic discontinuity for faults and stratigraphic features: The coherence cube: *The Leading Edge*, **14**, no. 10, 1053-1058.
- Banik, N. C., Lerche, I., Resnick, J. R., and Shuey, R. T., 1985, Stratigraphic filtering, Part II: Model spectra: *Geophysics*, **50**, 2775-2783.
- Bano, M., 1989, *Extraction automatique des réflexions, modélisation des diffractons et migration des données de sismique profonde Ecors*: Ph.D. thesis, Université Louis Pasteur, Strasbourg.
- Battiti, R., 1992, First and second order methods for learning between steepest descent and Newton's methods: *Neural Computation*, **4**, 141-166.
- Baum, E. B., and Haussler, D., 1989, What size net gives valid generalization?: *Neural Computation*, **1**, 151-160.



- Bayes, T., 1763, Essay towards solving a problem in the doctrine of chances, *in* *Biometrika*, 1958, 45, 293-315.
- Bear, L. K., and Pavlis, G. L., 1997, Estimation of slowness vectors and their uncertainties using multi-wavelet seismic array processing: *Bull. Seism. Soc. Am.*, **87**, 755-769.
- Beltzer, A., 1989, Dispersion of seismic waves by a causal approach: *Pure Appl. Geophys.*, **128**, 147-156.
- Berthold, M. R., and Diamond, J., 1995, Boosting the performance of RBF networks with Dynamic Decay Adjustment, *in* Tesauro, G., Touretzky, D. S., and Leen, T. K., Eds., *Advances in neural processing information systems 7*: MIT Press, 521-528.
- Bishop, C. M., 1995, Training with noise is equivalent to Tikhonov regularization: *Neural Computation*, **7**, 108-116.
- Bouchon, M., 1981, A simple method to calculate Green's functions for elastic layered media: *Bull. Seism. Soc. Am.*, **71**, 959-971.
- Burridge, R., Papanicolaou, G., and White, B. S., 1988, One-dimensional wave propagation in a highly discontinuous medium: *Wave Motion*, **10**, 19-44.
- Burridge, R., Lewicki, P., and Papanicolaou, G., 1994, Pulse stabilization in a strongly heterogeneous medium: *Wave Motion*, **20**, 177-195.
- Calderón-Macías, C., Sen, M. K., and Stoffa, P. L., 1997, Hopfield neural networks, and mean field annealing for seismic deconvolution and multiple attenuation: *Geophysics*, **62**, 992-1002.
- Calderón-Macías, C., Sen, M. K., and Stoffa, P. L., 1998, Automatic NMO correction and velocity estimation by a feedforward neural network: *Geophysics*, **63**, 1696-1707.
- Carpenter, G. A., and Grossberg, S., 1987, Art2: Self-organization of stable category recognition codes for analog input patterns: *Applied Optics*, pages 4919-4930.
- Chandrasekhar, S., 1960, *Radiative transfer*: Dover Publications.
- Chen, S., Cowan, C. F. N., and Grant, P. M., 1991, Orthogonal Least Squares learning algorithm for Radial Basis Function networks: *IEEE Trans. Neural Networks*, **2**, 302-309.
- Chentouf, R., and Jutten, C., 1996a, Combining sigmoids and Radial Basis Functions in evolutive neural architectures: *Eur. Symp. Artificial Neural Networks*, D Facto Publications, 129-134.
- 1996b, DWINA: Depth and Width Incremental Neural Algorithm: *Proc. IEEE Int. Conf. Neural Networks*, 153-158.
- Chentouf, R., 1997, Construction de réseaux de neurones multicouches pour l'approximation: Ph.D. thesis, Institut National Polytechnique, Grenoble.
- Chernov, L. A., 1960, *Wave propagation in a random medium*: McGraw-Hill.
- Choukroune, P., and Ecors-Team, 1989, The Ecors Pyrenean deep seismic profile reflection data and the overall structure of an orogenic belt: *Tectonics*, **8**, 23-39.
- Claerbout, J. F., 1985, *Imaging the Earth's interior*: Blackwell Scientific Publications.

- Clouet, J. F., and Fouque, J. P., 1994, Spreading of a pulse traveling in random media: *Annals of Applied Probability*, **4**, 1083-1097.
- Cottrell, M., Girard, B., Girard, Y., Mangeas, M., and Muller, C., 1995, Neural modeling for time series: A statistical stepwise method for weight elimination: *IEEE Trans. Neural Networks*, **6**, 1355-1364.
- Cybenko, G., 1989, Approximation by superpositions of a sigmoidal function: *Mathematics of Control, Signals and Systems*, **2**, 303-314.
- Dahl, E. D., 1987, Accelerated learning using the generalized delta rule: *Proc. IEEE Int. Conf. Neural Networks*, **2**, 523-530.
- Dai, H., and MacBeth, C., 1994, Split shear-wave analysis using an artificial neural network?: *First Break*, **12**, 605-613.
- Daignières, M., Séguret, M., Specht, M., and Ecors-Team, 1994, The Arzacq-Western Pyrenees Ecors deep seismic profile: Special publication E.A.P.G., **4**, Hydrocarbon and petroleum geology of France, 199-209.
- Delyon, F., Kunz, H., and Souillard, B., 1983, 1D wave equations in disordered media: *J. Phys. A*, **16**, 25.
- Dencœur, T., and Lengellé, R., 1993, Initializing back-propagation networks with prototypes: *Neural Networks*, **6**, 351-363.
- Dietrich, M., 1988, Modeling of marine seismic profiles in the  $t-x$  and  $\tau-p$  domains: *Geophysics*, **53**, 453-465.
- Dorren, H. J. S., and Snieder, R. K., 1995, The stability of finite-dimensional inverse problems: *Inverse Problems*, **11**, 889-911.
- Dowla, F. U., Taylor, S. R., and Anderson, R. W., 1990, Seismic discrimination with artificial neural networks: Preliminary results with regional spectral data: *Bull. Seism. Soc. Am.*, **80**, 1346-1373.
- Dumay, J., and Fournier, F., 1988, Multivariate statistical analyses applied to seismic facies recognition: *Geophysics*, **53**, 1151-1159.
- Efron, B., and Tibshirani, R. J., 1993, *An introduction to the bootstrap*: Monographs on statistics and applied probability, Chapman and Hall.
- Efron, B., 1979, Bootstrap methods: Another look at the Jackknife: *Ann. Statist.*, **7**, 1-26.
- Fahlman, S. E., and Lebiere, C., 1990, The cascade-correlation learning architecture, *in* Touretzky, D. S., Ed., *Advances in neural information processing systems 2*: Morgan Kaufmann, 524-532.
- Fedi, M., and Quarta, T., 1998, Wavelet analysis for the regional-residual and local separation of potential field anomalies: *Geophys. Prosp.*, **46**, 507-525.
- Flatté, S. M., and Wu, R.-S., 1988, Small-scale structure in the lithosphere and asthenosphere deduced from arrival-time and amplitude fluctuations at Norsar: *J. Geophys. Res.*, **93**, 6601-6614.



- Frankel, A., and Clayton, R., 1986, Finite difference simulations of seismic scattering: Implications for the propagation of short-period seismic waves in the crust and models of crustal heterogeneity: *J. Geophys. Res.*, **91**, 6465-6489.
- Freire, S. L. M., and Ulrych, T. J., 1988, Application of singular value decomposition to vertical seismic profiling: *Geophysics*, **53**, 778-785.
- Frisch, U., 1968, Wave propagation in random media, *in* Bharucha-Reid, A. T., Ed., *Probabilistic methods in applied mathematics*: Academic press, 1, 75-198.
- Fürstenberg, H., 1963, Noncommuting random products: *Trans. Amer. Math. Soc.*, **108**, 377.
- Gallinari, P., Thiria, S., Badran, F., and Fogelman-Soulie, F., 1991, On the relations between discriminant analysis and multilayer perceptrons: *Neural Networks*, **4**, 349-360.
- Gazdag, J., 1978, Wave equation migration by phase shift: *Geophysics*, **43**, 1342-1351.
- Geman, S., Bienenstock, E., and Doursat, R., 1992, Neural networks and the bias/variance dilemma: *Neural Computation*, **4**, 1-58.
- Glangeaud, F., and Mari, J. L., 1994, *Wave separation*: Editions Technip.
- Goldberg, D. E., 1989, *Genetic algorithms in search, optimization and machine learning*: Addison Wesley.
- Gradshteyn, I. S., and Ryzhik, I. M., 1980, *Table of integrals, series and products*: Academic Press.
- Grandvalet, Y., and Canu, S., 1995, Comments on "Noise injection into inputs in back propagation learning": *IEEE Trans. Systems, Man, and Cybernetics*, **25**, 678-681.
- Gredeskul, S. A., and Freilikher, V. D., 1990, Localization and wave propagation in randomly layered media: *Soviet Physics Uspekhi*, **33**, 134-146.
- Hansen, K. M., Roy-Chowdhury, K., and Phinney, R. A., 1988, The sign filter for seismic event detection: *Geophysics*, **53**, 1024-1033.
- Hanson, S. J., and Pratt, L. Y., 1989, Comparing biases for minimal network construction with backpropagation, *in* Touretzky, D. S., Ed., *Advances in neural information processing systems 1*: Morgan Kaufmann, 177-185.
- Harlan, W. S., Claerbout, J. F., and Rocca, F., 1984, Signal/noise separation and velocity estimation: *Geophysics*, **49**, 1869-1880.
- Hassibi, B., and Stork, D. G., 1993, Second order derivatives for network pruning: Optimal Brain Surgeon, *in* Hanson, S. J., Cowan, J. D., and Giles, C. L., Eds., *Advances in neural information processing systems 5*: Morgan Kaufmann, 164-171.
- Hemon, C., and Mace, D., 1978, Essai d'une application de la transformation de Karhunen-Loève au traitement sismique: *Geophys. Prosp.*, **26**, 600-626.
- Hérault, J., and Jutten, C., 1994, *Réseaux neuronaux et traitement de signal*: Traitement du signal, Hermès édition.

- Herraiz, M., and Espinosa, A. F., 1987, Coda waves: A review: *Pure Appl. Geophys.*, **125**, 499-577.
- Hertz, J., Krogh, A., and Palmer, R. G., 1991, *Introduction to the theory of neural computation*: Addison-Wesley.
- Hobbs, R. W., 1990, Effective Q determination using frequency methods on BIRPS data: *Tectonophysics*, **173**, 25-30.
- Hopfield, J. J., 1984, Neurons with graded response have collective computational properties like those of two-state neurons: *Proc. Natl. Acad. Sci. USA*, **81**, 3088-3092.
- Hornbostel, S., 1991, Spatial prediction filtering in the t-x and f-x domains: *Geophysics*, **56**, 2019-2026.
- Hosseini, S., and Jutten, C., 1999, Simultaneous estimation of signal and noise in constructive neural networks: *Neural Computation*, submitted.
- Huang, S. C., and Huang, Y. F., 1991, Bounds on the number of hidden neurons in multilayer perceptrons: *IEEE Trans. Neural Networks*, **2**, 47-55.
- Huang, Z., Shimeld, J., Williamson, M., and Katsube, J., 1996, Permeability prediction with artificial neural network modeling in the Ventura gas field, offshore Eastern Canada: *Geophysics*, **61**, 422-436.
- Hush, D. R., and Horne, B. G., 1993, Progress in supervised neural networks - What's new since Lippmann?: *IEEE Sign. Process. Mag.*, **10(1)**, 8-39.
- Hush, D., Horne, B., and Salas, J. M., 1992, Error surfaces for multi-layer perceptrons: *IEEE Trans. Systems, Man and Cybernetics*, **22**, 1152-1161.
- Hwang, J. N., Lat, S. R., Maechler, M., Martin, D., and Schimert, J., 1994, Regression modeling in back-propagation and Projection Pursuit Learning: *IEEE Trans. Neural Networks*, **5**, 342-353.
- Jacobs, R. A., 1988, Increased rates of convergence through learning rate adaptation: *Neural Networks*, **1**, 295-308.
- Jakubowicz, H., and Levin, S., 1983, A simple exact method of 3D migration: *Geophys. Prosp.*, **31**, 34-35.
- Judge, G. G., Griffiths, W. E., Hill, R. C., and Lee, T., 1980, *The theory and practice of econometrics*: John Wiley and Sons.
- Jutten, C., and Chentouf, R., 1995, A new scheme for incremental learning: *Neural Processing Letters*, **2**, 1-4.
- Karal, F. C., and Keller, J. B., 1964, Elastic, electromagnetic and other waves in a random medium: *J. Math. Phys.*, **5**, 537-547.
- Karnin, E., 1990, A simple procedure for pruning backpropagation trained neural networks: *IEEE Trans. Neural Networks*, **1**, 239-242.



- Karouia, M., Lengellé, R., and Dencoux, T., 1994, Weight initialization in BP networks using discriminant analysis techniques: Proc. Neural Networks and their Applications; 15-16 December 1994; Marseille, pages 171-180.
- Kennett, B. L. N., 1974, Reflections, rays and reverberations: Bull. Seism. Soc. Am., **64**, 1685-1696.
- Kirkpatrick, T. R., 1985, Localization of acoustic waves: Phys. Rev. B, **31**, 5746-5755.
- Kohler, W., Papanicolaou, G., and White, B., 1996, Localization and mode conversion for elastic waves in randomly layered media: Wave Motion, **23**, 1-22 and 181-201.
- Kohler, W., Papanicolaou, G., and White, B., 1999, Reflection and transmission of acoustic waves by a locally-layered slab, in Fouque, J. P., Ed., Diffuse waves in complex media: Kluwer Academic, NATO Sciences Series - Series C: Mathematical and Physical Sciences, 531, 347-381.
- Kohonen, T., 1989, Self-organization and associative memory, 3rd ed.: Springer-Verlag.
- Kong, S. M., Phinney, R. A., and Roy-Chowdhury, K., 1985, A nonlinear signal detector for enhancement of noisy seismic record sections: Geophysics, **50**, 539-550.
- Kumar, P., and Fofoula-Georgiou, E., 1997, Wavelet analysis for geophysical applications: Rev. of Geophysics, **35**, 385-412.
- Kwok, T.-Y., and Yeung, D.-Y., 1997, Constructive algorithms for structure learning in feedforward neural networks for regression problems: IEEE Trans. Neural Networks, **8**, 630-645.
- Lanczos, C., 1961, Linear differential operators: D. Van Nostrand Co.
- Langer, H., Nunnari, G., and Occhipinti, L., 1996, Estimation of seismic waveform governing parameters with neural networks: J. Geophys. Res., **101**, 20109-20118.
- Le, L. H., and Nyland, E., 1990, Pattern analysis of seismic records: Geophysics, **55**, 20-28.
- Le Cun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D., 1989, Backpropagation applied to handwritten zip code recognition: Neural Computation, **1**, 541-551.
- Le Cun, Y., Denker, J. S., and Solla, S. A., 1990, Optimal Brain Damage, in Touretzky, D., Ed., Advances in neural information processing systems 2: Morgan Kaufmann, 598-605.
- Le Cun, Y., Kanter, I., and Solla, S., 1991, Eigenvalues of covariance matrices: Application to neural network learning: Phys. Rev. Lett., **66**, 2396-2399.
- Lecerf, D., 1996, Apports des données sismiques E.C.O.R.S. 2,5D à la connaissance de la structure superficielle de la croûte dans la Zone Nord Pyrénéenne: Ph.D. thesis, Université Louis Pasteur, Strasbourg.
- Lepage, G. P., 1978, A new algorithm for adaptive multidimensional integration: J. Comp. Phys., **27**, 192-203.
- Lewicki, P., and Burrige, R., 1996, Reflection from a deep interface in a strongly heterogeneous layered medium: Geophys. Prosp., **44**, 571-581.

- Li, Q., Vasudevan, K., and Cook, F. A., 1997, Seismic skeletonization: A new approach to interpretation of seismic reflection data: J. Geophys. Res., **102**, 8427-8445.
- Lifshits, L., Gredeskul, S. A., and Pastur, L. A., 1988, Introduction to the theory of disordered systems: John Wiley.
- Lilly, J. M., and Park, J., 1995, Multiwavelet spectral and polarization analyses of seismic records: Geophys. J. Int., **122**, 1001-1021.
- Line, C. E. R., Hobbs, R. W., Hudson, J. A., and Snyder, D. B., 1998, Statistical inversion of controlled-source seismic data using parabolic wave scattering theory: Geophys. J. Int., **132**, 61-78.
- Lines, L. R., and Treitel, S., 1984, Tutorial: A review of least-squares inversion and its application to the geophysical domain: Geophys. Prosp., **32**, 159-186.
- Lippmann, R. P., 1987, An introduction to computing with neural nets: IEEE ASSP Mag., **4**(2), 4-22.
- Lu, S.-Y., and Cheng, Y.-C., 1990, An iterative approach to seismic skeletonization: Geophysics, **55**, 1312-1320.
- MacKay, D. J. C., 1992, Bayesian interpolation: Neural Computation, **4**, 415-447.
- Marfurt, K., Kirilin, R., Farmer, S., and Bahorich, M., 1998, 3-D seismic attributes using a semblance-based coherency algorithm: Geophysics, **63**, 1150-1165.
- Margerin, L., Campillo, M., and van Tiggelen, B., 1998, Radiative transfer and diffusion of waves in a layered medium: New insight into coda Q: Geophys. J. Int., **134**, 596-612.
- Marthelot, J.-M., Siliqi, R., Bitri, A., Paul, A., Hirn, A., Daignières, M., Damotte, B., Specht, M., de Bazelaire, E., Lortscher, A., and Rappin, D., 1994, Three-dimensional imaging of the crust using a sparse land acquisition grid: the Ecors 2.5-D experiment: Tectonophysics, **232**, 365-377.
- Masters, T., 1995, Advanced algorithms for neural networks - A C++ sourcebook: John Wiley.
- Matsuoka, K., 1992, Noise injection into inputs in back-propagation learning: IEEE Trans. Systems, Man, and Cybernetics, **22**, 436-440.
- Mattauer, M., 1990, Une autre interprétation du profil Ecors Pyrénées: Bull. Soc. géol. Fr., **6**, 307-311.
- McCormack, M. D., Zaucha, D. E., and Dushek, D. W., 1993, First-break refraction event picking and seismic data trace editing using neural networks: Geophysics, **58**, 67-78.
- McCulloch, W. S., and Pitts, W., 1943, A logical calculus of the ideas immanent in nervous activity: Bull. Math. Biophys., **5**, 115-133.
- Møller, M. F., 1993, A scaled conjugate gradient algorithm for fast supervised learning: Neural Networks, **6**, 525-533.



- Moody, J. E., 1992, The effective number of parameters: An analysis of generalization and regularization in nonlinear learning systems, *in* Moody, J. E., Hanson, S. J., and Lippmann, R. P., Eds., *Advances in neural information processing systems 4*: Morgan Kaufmann, 847-854.
- Moody, J., 1994, Prediction risk and architecture selection for neural networks, *in* Cherkassy, V., Frieman, J. H., and Wechsler, H., Eds., *From statistics to neural networks: Theory and pattern recognition applications*: Springer-Verlag, Nato ASI Series F: Computer and systems sciences, 213-225.
- Morlet, J., Arens, G., E. Fourgeau, E., and Giardt, D., 1982, Wave propagation and sampling theory - Part II: Sampling theory and complex waves: *Geophysics*, **47**, 222-236.
- Morse, P. M., and Feshbach, H., 1953, *Methods of theoretical physics*: McGraw-Hill.
- Mozer, M. C., and Smolensky, P., 1989, Skeletonization: A technique for trimming the fat from a network via relevance assessment, *in* Touretzky, D. S., Ed., *Advances in neural information processing systems 1*: Morgan Kaufmann, 107-115.
- Murat, M. E., and Rudman, A. J., 1992, Automated first arrival picking: A neural network approach: *Geophys. Prosp.*, **40**, 587-604.
- Murata, N., Yoshizawa, S., and Amari, S., 1994, Network Information Criterion - Determining the number of hidden units for an artificial neural network model: *IEEE Trans. Neural Networks*, **5**, 865-872.
- Muyzert, E. J., 1998, Monte Carlo waveform inversion and deep continental structure: Ph.D. thesis, University of Utrecht.
- Newman, P., 1973, Divergence effects in a layered Earth: *Geophysics*, **38**, 481-488.
- Nguyen, D., and Widrow, B., 1990, Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights: *Proc. Int. Joint Conf. Neural Networks*, **III**, 2063-2068.
- Nolet, G., 1987, Seismic wave propagation and seismic tomography, *in* Nolet, G., Ed., *Seismic tomography*: D. Reidel Publishing Company, 1-24.
- Nowlan, S. J., and Hinton, G. E., 1992, Simplifying neural networks using soft weight-sharing: *Neural Computation*, **4**, 473-493.
- O'Doherty, R. F., and Anstey, N. A., 1971, Reflections on amplitudes: *Geophys. Prospect.*, **19**, 430-458.
- Oseledec, V., 1968, A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems: *Trans. Moscow Math. Soc.*, **19**, 197-231.
- Pain, H. J., 1983, *The physics of vibrations and waves*: 3rd edition, Wiley.
- Papanicolaou, G., 1998, Mathematical problems in geophysical wave propagation: *Proceedings of the International Congress of Mathematicians, Extra Volume ICM 98 I, Documenta Mathematica*, 241-265.

- Park, J., and Sandberg, I. W., 1991, Universal approximation using Radial-Basis-Function networks: *Neural Computation*, **3**, 246-257.
- Park, J., Lindberg, C. R., and III, F. L. V., 1987, Multitaper spectral analysis of high-frequency seismograms: *J. Geophys. Res.*, **92**, 12675-12684.
- Pellilo, M., and Fanelli, A. M., 1993, A method of pruning layered feed forward neural networks, *in* Prieto, A., Ed., *International workshop on artificial neural networks*: Springer Verlag, Lecture notes in computer science, 278-283.
- Platt, J., 1991, A resource-allocating network for function interpolation: *Neural Computation*, **3**, 213-225.
- Poulton, M. M., Sternberg, B. K., and Glass, C. E., 1992, Location of subsurface targets in geophysical data using neural networks: *Geophysics*, **57**, 1534-1544.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P., 1992, *Numerical recipes in Fortran: The art of scientific computing*: 2nd edition, Cambridge University Press.
- Pullammanappallil, S., Levander, A., and Larkin, S. P., 1997, Estimation of crustal stochastic parameters from seismic exploration data: *J. Geophys. Res.*, **102**, 15269-15286.
- Ramakrishnan, T. V., 1987, Electron localization, *in* Souletie, J., Vannimenus, J., and Stora, R., Eds., *Chance and Matter - Les Houches, Session XLVI, 1986*: North-Holland, 213-304.
- Rappin, D., Marthelot, J.-M., and de Bazelaire, E., 1991, Amplitude and time-frequency analysis of Ecores deep reflection seismic data: 53rd Mtg. Eur. Assoc. Expl Geophys., Abstracts, 384-385.
- Rappin, D., 1992, Apport des analyses d'amplitude et temps-fréquence à l'exploitation des données de sismique profonde: Ph.D. thesis, Université Louis Pasteur, Strasbourg.
- Reed, R., 1993, Pruning algorithms - A survey: *IEEE Trans. Neural Networks*, **4**, 740-747.
- Richards, J. A., 1993, *Remote sensing digital image analysis, an introduction*: Springer-Verlag.
- Riedmiller, M., and Braun, H., 1993, A direct adaptive method for faster backpropagation learning: the RPROP algorithm: *Proc. IEEE Int. Conf. Neural Networks*, **1**, 586-591.
- Rioul, O., and Vetterli, M., 1991, Wavelets and signal processing: *IEEE Sign. Process. Mag.*, **8(4)**, 14-38.
- Robinson, E. A., 1996, Seismic source signature estimation and measurement: *Soc. Expl. Geophys.*
- Romeo, G., 1994, Seismic signals detection and classification using artificial neural networks: *Annali di Geofisica*, **37**, 343-353.
- Röth, G., and Tarantola, A., 1994, Neural networks and inversion of seismic data: *J. Geophys. Res.*, **99**, 6753-6768.
- Roure, F., Choukroune, P., Berastegui, X., Munoz, J. A., Villien, A., Matheron, P., Bareyt, M., Séguret, M., Camara, P., and Deramond, J., 1989, Ecores deep seismic data and balanced cross sections: Geometric constraints on the evolution of the Pyrenees: *Tectonics*, **8**, 41-50.



- Rumelhart, D. E., Hinton, G. E., and Williams, R. J., 1986, Learning internal representation by backpropagating errors: *Nature*, **332**, 533–536.
- Saito, N., 1994, Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion, *in* Foufoula-Georgiou, E., and Kumar, P., Eds., *Wavelets in geophysics*: Academic Press, Inc., 299–324.
- Sato, H., and Fehler, M., 1998, *Seismic wave propagation and scattering in the heterogeneous Earth*: AIP Press.
- Sato, H., 1982, Amplitude attenuation of impulsive waves in random media based on traveltime corrected mean wave formalism: *J. Acoust. Soc. Am.*, **71**, 559–564.
- Scales, J. A., and Snieder, R., 1998, What is noise?: *Geophysics*, **63**, 1122–1124.
- Scales, J. A., 1993, On the use of localization theory to characterize elastic wave propagation in randomly stratified 1-D media: *Geophysics*, **58**, 177–179.
- Sethi, I. K., 1990, Entropy nets: From decision trees to neural networks: *Proc. IEEE*, **78**, 1605–1613.
- Sethi, I. K., 1995, Neural implementation of tree classifiers: *IEEE Trans. Systems, Man and Cybernetics*, **25**, 1243–1249.
- Shapiro, S. A., and Hubral, P., 1999, Elastic waves in random media – Fundamentals of seismic stratigraphic filtering: volume 80 of **Lecture Notes in Earth Sciences**, Springer Verlag.
- Shapiro, S. A., and Zien, H., 1993, The O'Doherty-Anstey formula and the localization of seismic waves: *Geophysics*, **58**, 736–740.
- Shapiro, S. A., Zien, H., and Hubral, P., 1994, A generalized O'Doherty-Anstey formula for waves in finely layered media: *Geophysics*, **59**, 1750–1762.
- Shapiro, S. A., Hubral, P., and Ursin, B., 1996, Reflectivity/transmissivity for one-dimensional inhomogeneous elastic media: Dynamic-equivalent-medium approach: *Geophys. J. Int.*, **126**, 184–196.
- Sheng, P., and Zhang, Z.-Q., 1995, Is a layered medium one dimensional?: *Phys. Rev. Lett.*, **74**, 1343–1346.
- Sheng, P., White, B., Zhang, Z.-Q., and Papanicolaou, G., 1986a, Minimum wave-localization length in a one-dimensional random medium: *Phys. Rev. B*, **34**, 4757–4761.
- Sheng, P., Zhang, Z.-Q., White, B., and Papanicolaou, G., 1986b, Multiple-scattering noise in one dimension: Universality through localization-length scaling: *Phys. Rev. Lett.*, **57**, 1000–1003.
- Sheng, P., White, B., Zhang, Z.-Q., and Papanicolaou, G., 1990, Wave localization and multiple scattering in randomly-layered media, *in* Sheng, P., Ed., *Scattering and localization of classical waves in random media*: World Scientific, 563–619.
- Sheng, P., 1995, *Introduction to wave scattering, localization and mesoscopic phenomena*: World Scientific.
- Sietsma, J., and Dow, R. D. F., 1991, Creating artificial neural networks that generalize: *Neural Networks*, **4**, 67–79.

- Siliqi, R., 1994, Elaboration du traitement et interprétation d'une expérience de sismique réflexion 3D à l'échelle de la croûte dans les Pyrénées (Ecors 2.5D): Ph.D. thesis, Université Louis Pasteur, Strasbourg.
- Slepian, D., 1983, Some comments on fourier analysis, uncertainty and modeling: *SIAM Review*, **24**, 379–393.
- Snieder, R., 1990, Linearized inversion of seismic waveforms: Non-published manuscript.
- SNNS, User manual, 1996, Version 4.1: Available by anonymous ftp: ftp.informatik.uni-stuttgart.de (129.69.211.2).
- Souillard, B., 1987, Waves and electrons in inhomogeneous media, *in* Souletie, J., Vannimenus, J., and Stora, R., Eds., *Chance and Matter – Les Houches, Session XLVI, 1986*: North-Holland, 305–381.
- Specht, D., 1990, Probabilistic Neural Network: *Neural Networks*, **3**, 109–118.
- Stark, H., 1987, *Image recovery, theory and application*: Academic Press.
- Stone, M., 1974, Cross-validatory choice and assessment of statistical predictions: *J. Roy. Statist. Soc.*, **36**, 111–147.
- Tarantola, A., 1987, *Inverse problem theory – Methods for data fitting and model parameter estimation*: Elsevier.
- Tatarskii, V. I., 1961, *Wave propagation in a turbulent medium*: McGraw-Hill.
- Thomson, D. J., 1982, Spectral estimation and harmonic analysis: *IEEE Proc.*, **70**, 1055–1096.
- Trorey, A. W., 1970, A simple theory for seismic diffractions: *Geophysics*, **35**, 762–784.
- Van der Baan, M., and Paul, A., 1999, Recognition and reconstruction of coherent energy with application to deep seismic reflection data: *Geophysics*, in press.
- Van der Baan, M., 1996, Partitioned waveform inversion using probabilistic, adaptive Monte Carlo importance sampling: Master's thesis, University of Utrecht.
- Van der Sluis, A., and Van der Vorst, H. A., 1987, Numerical solutions of large, sparse linear algebraic systems arising from tomographic problems, *in* Nolet, G., Ed., *Seismic tomography*: D. Reidel publishing Co., 49–83.
- Van Tiggelen, B. A., 1999, Localization of waves, *in* Fouque, J. P., Ed., *Diffuse waves in complex media*: Kluwer Academic, NATO Sciences Series – Series C: Mathematical and Physical Sciences, 531, 1–60.
- VanDecar, J. C., and Snieder, R., 1994, Obtaining smooth solutions to large, linear, inverse problems: *Geophysics*, **59**, 818–829.
- Virster, A. D., 1979, On the products of random matrices and operators: *Theor. Prob. Appl.*, **24**, 367.
- Wahba, G., and Wold, S., 1975, A completely automatic French curve: Fitting spline functions by cross-validation: *Communications in Statistics*, **4**, 1–17.



- Wang, L.-X., and Mendel, J. M., 1992, Adaptive minimum prediction-error deconvolution and source wavelet estimation using Hopfield neural networks: *Geophysics*, **57**, 670-679.
- Warner, H., 1987, Migration - Why doesn't it work for deep continental data?: *Geophys. J. R. Astron. Soc.*, **89**, 21-26.
- Weigend, A. S., Rumelhart, D. E., and Huberman, B. A., 1991, Generalization by weight-elimination with application to forecasting, *in* Lippmann, R. P., Moody, J. E., and Touretzky, D. S., Eds., *Advances in neural information processing systems 3*: Morgan Kaufmann, 875-882.
- White, B., Sheng, P., Zhang, Z.-Q., and Papanicolaou, G., 1987, Wave localization characteristics in the time domain: *Phys. Rev. Lett.*, **59**, 1918-1921.
- White, B., Sheng, P., and Nair, B., 1990, Localization and backscattering spectrum of seismic waves in stratified lithology: *Geophysics*, **55**, 1158-1165.
- Wiersma, D. S., Bartolini, P., Lagendijk, A., and Righini, R., 1997, Localization of light in a disordered medium: *Nature*, **390**, 671-673.
- Williams, P. M., 1995, Bayesian regularization and pruning using a Laplace prior: *Neural Computation*, **7**, 117-143.
- Wu, R.-S., 1982, Attenuation of short period seismic waves due to scattering: *Geophys. Res. Lett.*, **9**, 9-12.
- Wu, R.-S., 1985, Multiple scattering and energy transfer of seismic waves - Separation of scattering effect from intrinsic attenuation - I. Theoretical modeling: *Geophys. J. R. astr. Soc.*, **82**, 57-80.
- Yilmaz, O., 1987, Seismic data processing: volume 2 of *Investigations in geophysics*, Soc. Expl. Geophys.
- Zhang, Y., and Paulson, K. V., 1997, Magnetotelluric inversion using regularized Hopfield neural networks: *Geophys. Prosp.*, **45**, 725-743.



**Thèse de Doctorat  
de l'Université Joseph Fourier Grenoble I**

---

*Titre de l'ouvrage:*

**Deux méthodes d'inférence statistique appliquées  
aux données de sismique réflexion profonde:  
Détection de signaux et localisation d'onde**

*Auteur:* **Mirko VAN DER BAAN**

*Établissement:*

**Observatoire de Grenoble  
Laboratoire de Géophysique Interne et  
Tectonophysique**

*Résumé*

L'objectif de toute acquisition de données géophysiques est d'obtenir une image plus précise de l'intérieur de la Terre. Afin de préciser cette image, on utilise principalement des méthodes d'inférence déterministe ayant comme but de fournir une description exacte de chaque détail particulier.

Contraire aux méthodes déterministes, les approches statistiques n'essaient de saisir que les caractéristiques principales de l'ensemble des détails. Ceci réduit la complexité mathématique, nous permettant ainsi de mieux étudier certains phénomènes physiques. De plus, elles peuvent nous aider à sonder la Terre à une échelle plus petite que la résolution des méthodes déterministes. Finalement, elles peuvent nous fournir une appréciation de vraisemblance, nous aidant ainsi à prendre des décisions ou à reconnaître des signaux.

Dans ce mémoire, nous avons utilisé ces 3 avantages sur 2 problèmes différents. D'abord, nous les avons employés afin d'augmenter le rapport signal sur bruit en détectant des réflexions dans des données de sismique réflexion profonde d'une faible qualité. Ceci nous a permis de faciliter l'interprétation du profil considéré. Deux méthodes différentes ont été considérées, celle dite Classification par Vraisemblance Maximale et celle des Réseaux de Neurones.

Ensuite, nous les avons utilisés afin d'étudier le processus de la diffusion forte dans un milieu acoustique, stratifié et aléatoire. Afin d'étudier ce problème, nous avons appliqué la théorie de la localisation d'onde prédisant l'atténuation apparente et la dispersion d'une impulsion traversant un tel milieu. Cette théorie nous a aussi permis d'estimer la magnitude et la dimension typiques des fluctuations des vitesses et le facteur de qualité de l'atténuation intrinsèque sous le Bassin de Mauléon dans les Pyrénées Occidentales. Ceci nous a montré que la croûte supérieure y est plus hétérogène et plus atténuante que la croûte inférieure et que les dimensions typiques y sont plus petites.

*Mots clés*

Sismologie - sismique réflexion profonde - détection de signaux - reconnaissance - réseaux neuronaux - localisation d'onde - atténuation - milieux aléatoires.