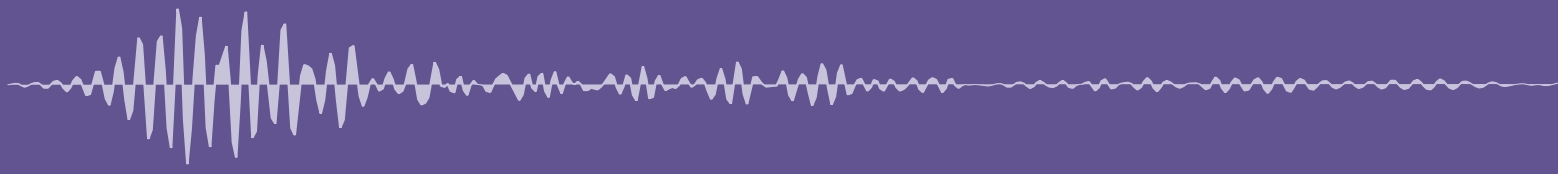
Costas E. Smaragdakis

PhD THESIS

Acoustic Signal Characterization using Hidden Markov Models with applications in Acoustical Oceanography



supervised by

Prof. Michael Taroudakis

Department of Mathematics and Applied Mathematics
University of Crete

University of Crete

Department of Mathematics and Applied Mathematics

Acoustic Signal Characterization using Hidden Markov Models with applications in Acoustical Oceanography

Thesis submitted for the degree of

Doctor of Philosophy

Author

Costas E. Smaragdakis

Supervisor

Prof. Michael Taroudakis

Heraklion, April 2019

Committee

PhD Advisory Committee

- Michael Taroudakis, Professor, Department of Mathematics and Applied Mathematics, University of Crete **(supervisor)**
- George Makrakis, Professor, Department of Mathematics and Applied Mathematics, University of Crete
- Chrysoula Tsogka, Professor, Department of Mathematics and Applied Mathematics, University of Crete

Dissertation Committee

- Stan Dosso, Professor and Director, Earth and Ocean Sciences, University of Victoria
- Vagelis Harmandaris, Associate Professor, Department of Mathematics and Applied Mathematics, University of Crete
- George Makrakis, Professor, Department of Mathematics and Applied Mathematics, University of Crete
- Emmanuel Skarsoulis, Research Director, IACM - Foundation for Research and Technology - Hellas
- Michael Taroudakis, Professor, Department of Mathematics and Applied Mathematics, University of Crete
- Chrysoula Tsogka, Professor, Department of Mathematics and Applied Mathematics, University of Crete
- George Tzagkarakis, Principal Researcher, ICS - Foundation for Research and Technology - Hellas

To my beloved wife Irene and our adorable kids Renos and Emmy, who are the reasons keep me alive.

To my beloved father, who did not get the opportunity for the education he was dreaming of.

Abstract

The scope of this PhD thesis is to develop a new probabilistic characterization scheme for acoustic signals recorded in the marine environment, with applications in acoustical oceanography. We will refer to the proposed scheme as Probabilistic Signal Characterization Scheme (PSCS). The schemes aims at the definition of a set of observables (signal features) that could characterize a signal to a unique way. To this end, a signal is decomposed into several levels using the stationary wavelet packet transform. This decomposition provides a time-frequency analysis of the characteristics of the signal. The stationary wavelet packet coefficients of the various levels are then modeled by a single left-to-right Hidden Markov Model (HMM) with Gaussian emission distributions. The concept behind the decision of using a sequential modeling of the signal's extracted coefficients, was the fact that a signal after propagation through a dispersive medium such as water column in the marine environment, exhibits evolving time-frequency characteristics. The association of a signal with a representative HMM is performed by means of the Expectation-Maximization (EM) algorithm. Eventually the signal is characterized by a set of parameters which describe the HMM.

The proposed signal characterization methods has been applied in inverse problems of acoustical oceanography. In particular, problems associated with the retrieval of the marine environmental parameters using measured features of the acoustic field due to a sound source have been considered. These problems being in nature non-linear are solved with optimization procedures requiring comparison of the characteristic of the measured acoustic signal with same of replica signals. In this work the Kullback-Leibler divergence is employed as the similarity measure of two signals, comparing their corresponding HMMs. To validate the performance of the proposed characterization scheme, the thesis presents few characteristic test cases in which simulated and real data have been considered. The measured signals are characterized by means of the proposed PSCS method and the model parameters of the marine environment have been estimated by employing a Genetic Algorithm (GA) over three sets of population of candidate model parameters. The GA leads to distributions of the model parameters of the

final population using Gaussian Mixture Model (GMM). This representation provides the solutions of the inverse problems in the form of the maximum of the marginal densities and a qualitative indication of the confidence intervals of the recoverable parameters. The results are compared with those obtained using the Statistical Signal Characterization Scheme (SSCS) proposed by Taroudakis et al. In addition, the results corresponding to the experimental data are compared to various approaches from the literature. The applications presented here confirmed the reliability and efficiency of the method when applied with typical signals used in acoustical oceanography.

Περίληψη

Αντικείμενο της παρούσας διδακτορικής διατριβής είναι η ανάπτυξη ενός νέου πιθανοθεωρητικού σχήματος για τον χαρακτηρισμό ακουστικών σημάτων που καταγράφονται στο θαλάσσιο περιβάλλον, με εφαρμογές στην ακουστική ωκεανογραφία. Το σχήμα στοχεύει στον ορισμό ενός συνόλου χαρακτηριστικών που θα χαρακτηρίζουν ένα σήμα με μοναδικό τρόπο. Για το σκοπό αυτό, ένα σήμα αρχικά αποσυντίθεται σε πολλαπλά επίπεδα χρησιμοποιώντας το στατικό μετασχηματισμό κυματιδιακών πακέτων (Stationary Wavelet Packet Transform). Η αποσύνθεση παρέχει μια χρονό-συχνοτική ανάλυση των χαρακτηριστικών του σήματος. Οι συντελεστές κυματιδιακών πακέτων του σήματος στα διάφορα επίπεδα περιγράφονται στη συνέχεια από ένα κρυπτό-Μαρκοβιανό μοντέλό (left-to-right Hidden Markov Model) με κανονικές κατανομές εκπομπών (Gaussian emission distributions). Η ιδέα πίσω τη χρήση μιας ακολουθιακής μοντελοποίησης των εξαγόμενων συντελεστών του σήματος βασίζεται στο γεγονός ότι ένα σήμα μετά τη διάδοση σε ένα μέσο διασποράς όπως είναι το θαλάσσιο περιβάλλον, παρουσιάζει εξελισσόμενα χρονο-συχνοτικά χαρακτηριστικά. Η αντιστοίχιση ενός σήματος με ένα αντιπροσωπευτικό κρυπτό-Μαρκοβιανό μοντέλο πραγματοποιείται με χρήση του αλγόριθμου Expectation-Maximization (EM). Τελικά το ακουστικό σήμα χαρακτηρίζεται από ένα σύνολο παραμέτρων που περιγράφουν το κρυπτό-Μαρκοβιανό μοντέλο.

Η προτεινόμενη μέθοδος χαρακτηρισμού ακουστικών σημάτων εφαρμόζεται στα πλαίσια της παρούσας διατριβής σε αντίστροφα προβλήματα ακουστικής ωκεανογραφίας. Συγκεκριμένα, σε προβλήματα που σχετίζονται με την ανάκτηση των θαλάσσιων περιβαλλοντικών παραμέτρων με τη χρήση μετρήσεων του ακουστικού πεδίου που παράγεται από μια υποβρύχια ακουστική πηγή. Αυτά τα προβλήματα είναι εν γένη μη γραμμικά και επιλύονται με μεθόδους βελτιστοποίησης συγκρίνοντας τα χαρακτηριστικά του μετρούμενου ακουστικού σήματος με τα αντίστοιχα συνθετικών σημάτων που παράγονται επιλύοντας το ευθύ πρόβλημα της ακουστικής διάδοσης μέσα από ένα εύρος πιθανών τιμών των προς ανάκτηση

παραμέτρων . Η σύγκριση πραγματοποιείται μέσω κατάλληλης μετρικής που αποτυπώνει το βαθμό ομοιότητας δυο σημάτων που χαρακτηρίζονται με την πιθανοθεωρήτικη μέθοδο που υιοθετήθηκε.

Σε αυτή την εργασία ως το μέτρο ομοιότητας δύο σημάτων χρησιμοποιείται η απόκλιση Kullback-Leibler (Kullback-Leibler Divergence) συγκρίνοντας τα αντίστοιχα κρυπτό-Μαρκοβιανά τους μοντέλα. Για να μελετήσουμε την απόδοση του προτεινόμενου σχήματος χαρακτηρισμού, παρουσιάζονται μερικές χαρακτηριστικές εφαρμογές αντιστροφής χρησιμοποιώντας τόσο συνθετικά όσο και πραγματικά δεδομένα.

Προκειμένου να γίνει αποδοτική αναζήτηση από το χώρο ερεύνης για τον υπολογισμό βέλτιστων λύσεων του αντιστρόφου προβλήματος χρησιμοποιήθηκε Γενετικός Αλγόριθμος. Οι παράμετροι που περιέχονται στο τελικό πληθυσμό του γενετικού αλγορίθμου περιγράφονται χρησιμοποιώντας ένα μοντέλο Gaussian Mixture (GMM). Αυτή η αναπαράσταση παρουσιάζει τις πιθανές λύσεις των αντίστροφων προβλημάτων με τη μορφή περιθωρίων κατανομών (marginal distributions) προσφέροντας μια ποιοτική ένδειξη των διαστημάτων εμπιστοσύνης των ανακτώμενων παραμέτρων. Τα αποτελέσματα συγκρίνονται με εκείνα που λαμβάνονται χρησιμοποιώντας τον Στατιστικό Χαρακτηρισμό Ακουστικών Σημάτων που έχει προταθεί από τον Ταρουδάκη και συνεργάτες. Επιπλέον, τα αποτελέσματα που αντιστοιχούν στα πειραματικά πραγματικά δεδομένα συγκρίνονται με διάφορες προσεγγίσεις άλλων ερευνητών από τη βιβλιογραφία. Οι εφαρμογές που παρουσιάστηκαν εδώ επιβεβαίωσαν την αξιοπιστία και την αποτελεσματικότητα της μεθόδου όταν εφαρμόστηκε σε τυπικά σήματα που χρησιμοποιούνται στην ακουστική ωκεανογραφία.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Professor Michael Taroudakis, for his generous and continuous support throughout my Master and Doctorate studies. His extensive knowledge of inverse problems in underwater acoustics was beneficial to me to work in this area. Without his supervision and his guidance, it would be impossible for me to improve my skills to conduct PhD level research. He is my mentor.

I would like to express my gratitude to the members of my advisory and dissertation committees for their comments and, suggestions on my work. I especially thank Professor Stan Dosso and Dr George Tzagkarakis for their insightful comments and suggestions on the text, which were crucial for me to produce the final version of my thesis.

My sincere acknowledgements go to Dr Emmanuel Skarsoulis for leading me into the field of Underwater Acoustics. Furthermore, his suggestions during the last year of my PhD studies, helped me to improve my work.

I especially thank Professor George Makrakis for his support and help in starting my PhD.

This work could not be achieved without the help and support of my beloved wife. She always believes in my abilities and makes me try hard to reach my goals. Without her understanding, I would not have been able to give full attention to this work and to complete it successfully. Thank you, my love!

My sincere thanks go to my dear parents, Manolis and Vaggelio Smaragdaki who grew me with love, and formed my personality. Moreover, my mother always tries to make my life easier by helping me with everyday things. Also, I would like to thank my dear sister Argiro and her husband George for being there for me whenever I need some help.

My sincere thanks to my dear parents in law, Nikos Kasotakis and Maria Skordili, as well as to my dear aunt, Despina Kasotaki, for their spiritual and financial support.

During my years of the doctoral dissertation, I worked with some talented young mathematicians and friends, Professor Victoria Taroudaki, John Mastrokalos and Ourania Siskoglou. John also supported me during my preparation for my PhD defense. Thank you all for your cooperation!

My sincere thanks to my dear friend and pediatrician of my children, Dr Lina Alexaki, for supporting my whole family. With her generosity, she helped us in difficult circumstances where we had problems with our medical insurance.

My sincere thanks and respect go to Dr Gerasimos Chouliaras. Our discussions, though by electronic means, helped and inspired me to a great extent.

Last but not least, I would like to thank many people that I cooperate with, and I get help and knowledge from them. Thus, my sincere thanks go to Professor N. Ross Chapman, Dr Nicos Melis, Dr Panagiotis Papadakis and Mr George Piperakis.

This work was supported by the following graduate fellowships which were carried out at IACM/FORTH:

- PEFYKA/KRIPIS - General Secreteriat for Research and Technology (2013/08 to 2015/08).
- SIEMENS - General Secreteriat for Reasearch and Technology (2016/01 to 2016/10).
- ARCHERS - Stavros Niarchos Foundation (2017/04 to 2018/04).
- PERAN/KRIPIS II - General Secreteriat for Research and Technology (2018/07 to 2019/07).

Contents

List of Acronyms

1D	One (**1**) **D**imensional
2D	Two (**2**) **D**imensional
3D	Three (**3**) **D**imensional
CFT	**C**ontinuous **F**ourier **T**ransform
CWT	**C**ontinuous **W**avelet **T**ransform
DFT	**D**iscrete **F**ourier **T**ransform
DWT	**D**iscrete **W**avelet **T**ransform
EM	**E**xpectation-**M**aximization Algorithm
E-step	**E**xpectation Step
FFT	**F**ast **F**ourier **T**ransform
FE	**F**eature **E**xtraction
FWT	**F**ast **W**avelet **T**ransform
FWPT	**F**ast **W**avelet **P**acket **T**ransform
FSWPT	**F**ast **S**tationary **W**avelet **P**acket **T**ransform
GA	**G**enetic **A**lgorithm
GMM	**G**aussian **M**ixture **M**odel
HLA	**H**orizontal **L**ine **A**rray of Hydrophones
HMM	**H**idden **M**arkov **M**odel
ICFT	**I**nverse **C**ontinuous **F**ourier **T**ransform
ICWT	**I**nverse **C**ontinuous **W**avelet **T**ransform
IDFT	**I**nverse **D**iscrete **F**ourier **T**ransform
IDWT	**I**nverse **D**iscrete **W**avelet **T**ransform
ISTA	**I**terative **S**hrinkage-**T**hresholding **A**lgorithm
KLD	**K**ullback-**L**eibler **D**ivergence
M-step	**M**aximization Step
MRA	**M**ulti-**R**esolution **A**nalysis
MLE	**M**aximum **L**ikelihood **E**stimator
NM	**N**ormal **M**ode
PPD	**P**osterior **P**robability **D**istribution
PSCS	**P**robabilistic **S**ignal **C**haracterization **S**cheme
RD	**R**ange **D**ependent
RI	**R**ange **I**ndependent
SDS	**S**parse **D**ensoising **S**cheme
SNR	**S**ignal-to-**N**oise **R**atio
SSCS	**S**tatistical **S**ignal **C**haracterization **S**cheme
SWD	**S**tationary **W**avelet **D**ecomposition
SWPT	**S**tationary **W**avelet **P**acket **T**ransform
SαS	**S**ymmetric **A**lpha-**S**table Distribution

VLA	**V**ertical **L**ine **A**rray of Hydrophones
WD	**W**avelet **D**ecomposition
WMA	**W**avelet **M**ultiresolution **A**nalysis
WPD	**W**avelet **P**acket **D**ecomposition
WPT	**W**avelet **P**acket **T**ransform

chapter 1

INTRODUCTION

As below, so above and beyond,
I imagine drawn beyond the lines of reason.
Spiral out. Keep going...

– Tool (Lateralus) –

Abstract

Underwater acoustics concerns the mathematical modelling of the propagation of the sound in the ocean environment. Underwater sounds have been used by marine animals for millions of years for navigation of the sea environment, sexual selection or to detect food. In addition, there are theories that Humpback whales sometimes sing just for enjoyment and satisfaction. Application of underwater acoustics "mimic" the capabilities of the marine animals to monitor the ocean environment by extracting proper features of the acoustic signals. Generally, underwater sounds span a wide range of frequencies, typically 10 Hz *to* 50 kHz. *In this work we shall deal with low frequency sounds in the range of which are pertinent to application of acoustical oceanography* 30 *to* 300 Hz. *In this introductory chapter we will present the motivation, the scope and a short outline of this PhD thesis.*

1.1 Motivation of the Research

The characteristics of an acoustic signal due to its propagation through a sea environment carry a lot of information related to the geometry and the properties of the environment.

Significant research has been carried out over the last decades in order to exploit these features to invert for the parameters of the marine environment. Acoustical oceanography to a great extend is associated with this type of problems.

Many of the methods that have been proposed for inverting for the recoverable model parameters exploit information relative to physical observables of the signals arrived at an array of hydrophones. The main drawback of using a multiple hydrophone setup, is of course the high expected deployment cost for in real world applications.

On the other hand, many researchers have focused their efforts on the development of inversion schemes that make use of the recordings at a single receivers. In order to have an adequate amount of information of the recoverable parameters, broadband sources have been used.

The motivation to develop and propose a new inversion scheme initiated by the fact that new techniques based on machine learning seem to be good candidates for the extraction of the acoustic signal characteristics that could be employed in inversion procedures. We wished to develop and test this novel techniques in problems of acoustical oceanography by introducing stochastic features of the signals which to our knowledge is unusual in this area of research. The new technique is called Probabilistic Signal Characterization Scheme (PSCS). Inspired by the Statistical Signal Characterization Scheme (SSCS) proposed by Taroudakis et al. [1], we wish to characterize acoustic signals using features (observables) that would be associated with a stochastic analysis of them. In the SSCS scheme, an acoustic signal is decomposed in various levels using appropriate filters through a wavelet transform and the wavelet subband coefficients were described by means of the alpha stable distributions. The research originally started with goal to improve the already existing statistical characterization scheme. However, looking at typical recordings of acoustic signals, we observe that the assumption that the wavelet coefficients of a signal are described as stationary process (one statistical distribution per coefficient vector), could be replaced by a non-stationary generative model. Thus, a probabilistic representation of an acoustic signal that utilizes a non-stationary sequential model could be a more appropriate decision. An appropriate reason choosing the probabilistic approach was that the SSCS scheme very sensitive to the presence of noise. Thus, we decided to study the probabilistic approach as a potential improvement over the SSCS in relation to the noise content of the signal. It came out that the new scheme is superior to the SSCS with respect to the signal characterization in presence of noise, still denoising algorithm should also be applied to obtain optimum results. Thus, a signal denoising scheme has been developed and applied in the applications presented in this thesis.

Eventually, the main goal of the research presented here was to study and develop tools for recognizing hidden patterns of a single recording due to a known source in order to be used for the recovery of the environmental parameters of an ocean environment when a typical experiment of ocean acoustic tomography or sea-bed classification is considered. We wish this study to conclude to a signal characterization scheme which will combine machine learning

approaches and the physics based modelling of the propagation problem in a sea environment and to applicable in inverse problems of underwater acoustics.

Following studies relative to speech recognition and various non-stationary processes such as time-series forecasting, we decided to use Hidden Markov Models (HMM) in order to describe mathematically the time-frequency characteristics of the signals.

The study presented in this thesis can be summarized in the following subsection.

1.2 Work Summary

The main contribution of the work presented in this thesis in the field of acoustical oceanography and machine learning applications is the introduction of a novel time-invariant probabilistic characterization scheme for acoustic signals. The proposed approach assigns the stationary wavelet packet coefficients of a signal to a Hidden Markov Model (HMM). Providing a novel approach for signal feature extraction. This method has not been applied so far, at least to our knowledge, to acoustic signals as the idea of associating wavelet packet coefficients with HMMs is also a new concept.

In addition, in order that the characterization of the signal is reliable, a new denoising scheme using dictionary learning and sparse decomposition of the raw signals was introduced. This denoising scheme also improves the quality of the signal characterizations obtained by the Statistical Signal Characterization Scheme (SSCS) introduced by Taroudakis et al. [1].

The new characterization scheme was applied with success in problems of acoustical oceanography using both simulated and real data, providing an alternative approach for estimating the unknown parameters of oceanic or shallow-water environment using a single recording of the acoustic field.

1.3 Thesis Outline

In this section a brief outline of the rest of this thesis will be presented:

CHAPTER 2 : The second chapter presents an overview of the modelling of the sound propagation in an oceanic or a shallow-water environment. The analysis is based on Normal Mode (NM) theory due to the fact that this method is to be used for solving the forward propagation problems which are associated with the inverse problems that we consider in this thesis. Also, some technical details about propagation using broadband sources are presented .

CHAPTER 3 : This chapter presents the formulation of the inverse problems in underwater acoustics accompanied with the state of the art of these problems. In this respect, various inversion techniques which uses alternative observables of the signals are briefly presented. Substantially, this chapter highlights the need for alternative inversion schemes that could take advantage of stochastic characteristics of a single recording of the acoustic field.

CHAPTER 4 : This chapter presents an overview of the mathematical and machine learning theory regarding the Statistical Signal Characterization Scheme (SSCS) as well as the proposed Probabilistic Signal Characterization Scheme (PSCS). This theory mainly concerns the Fourier and t:he Wavelet analysis, probabilistic models with hidden variables, such as Gaussian Mixture Models (GMMs) and Hidden Markov Models (HMMs), sparse dictionary learning and non-linear optimization using Genetic Algorithms (GAs).

CHAPTER 5 : In this chapter we present the Statistical Signal Characterization Scheme (SSCS) and we introduce the Probabilistic Signal Characterization Scheme (PSCS). Moreover, we describe the Sparse Denoising Scheme (SDS). The last two schemes consist the novel parts of the dissertation.

CHAPTER 6 : This chapter validates the robustness of the proposed schemes, for problems associated with the estimation of unknown model parameters of an underwater environment using both synthetic and experimental data.

CHAPTER 7 : The last chapter provides a summary of the work and its main results. Furthermore, possible improvements of the proposed framework as well as future applications are also provided.

chapter 2

Forward Propagation Problems in Underwater Acoustics

Abstract

The propagation of the sound in an oceanic or shallow-water environment is treated as a boundary value partial differential equation (PDE) problem in three dimensional space consisting of the linear acoustic equation and appropriate boundary conditions. This chapter will present the basic elements of the normal-mode theory of acoustic propagation which will be considered here as the main tool for handling forward propagation problems associated with the inverse problems considered in the thesis.

2.1 Modeling of the Sound Propagation in the Marine Environment

Sound propagation in a marine environment obeys the wave equation, which is derived by combining the fluid mechanics equations in connection with a set of boundary conditions.

Let us consider the acoustic pressure $P(\mathbf{x}, t)$ due to an underwater source that it is mathematically expressed by a function $f(\mathbf{x}, t)$. Figure 2.1 presents a typical marine environment with a point sound source.

A linear wave equation for the acoustic pressure in an ideal fluid medium can be derived after performing a linearization process of the hydrodynamic equations in connection with the adiabatic relation between the pressure and density [2].

Linearized Wave Equation

$$\nabla^2 P(\mathbf{x}, t) - \frac{1}{c^2(\mathbf{x})} \frac{\partial^2 P(\mathbf{x}, t)}{\partial t^2} = f(\mathbf{x}, t), \tag{2.1}$$

with $c(\mathbf{x})$ denoting the sound speed.

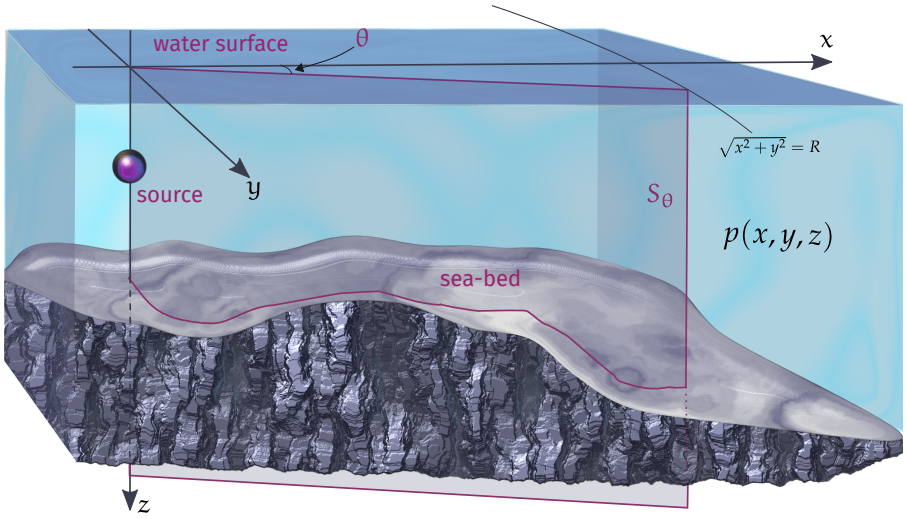


Figure 2.1: Schematic of a typical 3D ocean environment. The Nx2D solution is evaluated by considering radial 2D problems at specific angles θ and 3D synthesis using an interpolation technique.

If the inhomogeneity of the wave equation 2.1 is due to a time-harmonic point source, the right hand term takes the steady-state form:

HARMONIC POINT SOURCE

$$f(\mathbf{x},t) = -\delta(\mathbf{x}-\mathbf{x}_s)e^{-i\omega t}, \tag{2.2}$$

where δ is the *Dirac function* and $\mathbf{x}_s$ is the position vector of the source and ω is the angular frequency ($\omega = 2\pi f$), f is the source frequency.

Then the general solution of the wave equation is written in the form

$$P(\mathbf{x},t) = p(\mathbf{x})e^{-i\omega t}, \tag{2.3}$$

where $p(\mathbf{x})$ is the spatial component of the acoustic pressure.

If we substitute the expression 2.3 into 2.1, the spatial component of the acoustic pressure obey the following equation

HELMHOLTZ EQUATION

$$\nabla^2 p(\mathbf{x}) + k^2(\mathbf{x})p(\mathbf{x}) = -\delta(\mathbf{x}-\mathbf{x}_s), \tag{2.4}$$

where k is the wavenumber given as $k(\mathbf{x}) = \frac{\omega}{c(\mathbf{x})}$.

Several types of models to formulate the sound propagation in a marine environment have been proposed. The most popular among them are:

1. Ray models [3, 4, 5, 6]
2. Wavenumber integration techniques [7, 8],

3. Normal mode theory [9]

4. Parabolic approximations [10, 11, 12].

A detailed analysis of these methods can be found in the book by Jensen et al. [2].

The vast majority of them initially solve the problem in the frequency domain (Helmholtz equation) and then they transform the solution in the time domain. However, there were attempts to apply finite difference techniques directly in the wave equation providing approximations of the acoustic pressure in the time domain [13, 14].

Here, we will restrict our analysis to the normal mode theory as this is the method that has been used for the calculation of the acoustic field associated with the inverse problems that we have considered in this thesis.

Solving the three-dimensional propagation problem using normal mode theory is computationally intensive. There exist several attempts to treat this problem where special 3D geometry is considered [15, 16, 17]. A treatment that often is proposed for obtaining weakly three-dimensional solutions, is a Nx2D modelling approach [18] that combines a number of two-dimensional solutions along radial slices to produce an estimation of the three-dimensional solution. In Figure 2.1 a slice S_θ in an arbitrary angle θ is shown.

When the properties of the environment do not vary significantly with range, modelling of a range independent environment can give accurate results. On the other hand, if the environment presents characteristics with a strong dependency with range, a range dependent model consisting of a number of segments, each one without dependency on range, can be a compromise between yielding a reliable approximation of the acoustic field and not exceeding certain levels of computational complexity [19, 20].

2.2 Nx2D Approach

Perkins et al. [18] proposed a Nx2D approach of approximating the 3D solution of the underwater wave propagation in a specific domain, by solving N 2D propagation problems, each one corresponding to a vertical plane that passes through the point source.

When 3D effects are insignificant, the Nx2D approach provides a good approximation of the 3D model. The significance of these effects depends on the form of the anomaly as well as the parameters of the environment. In the environments that contain a local anomaly A, belonging

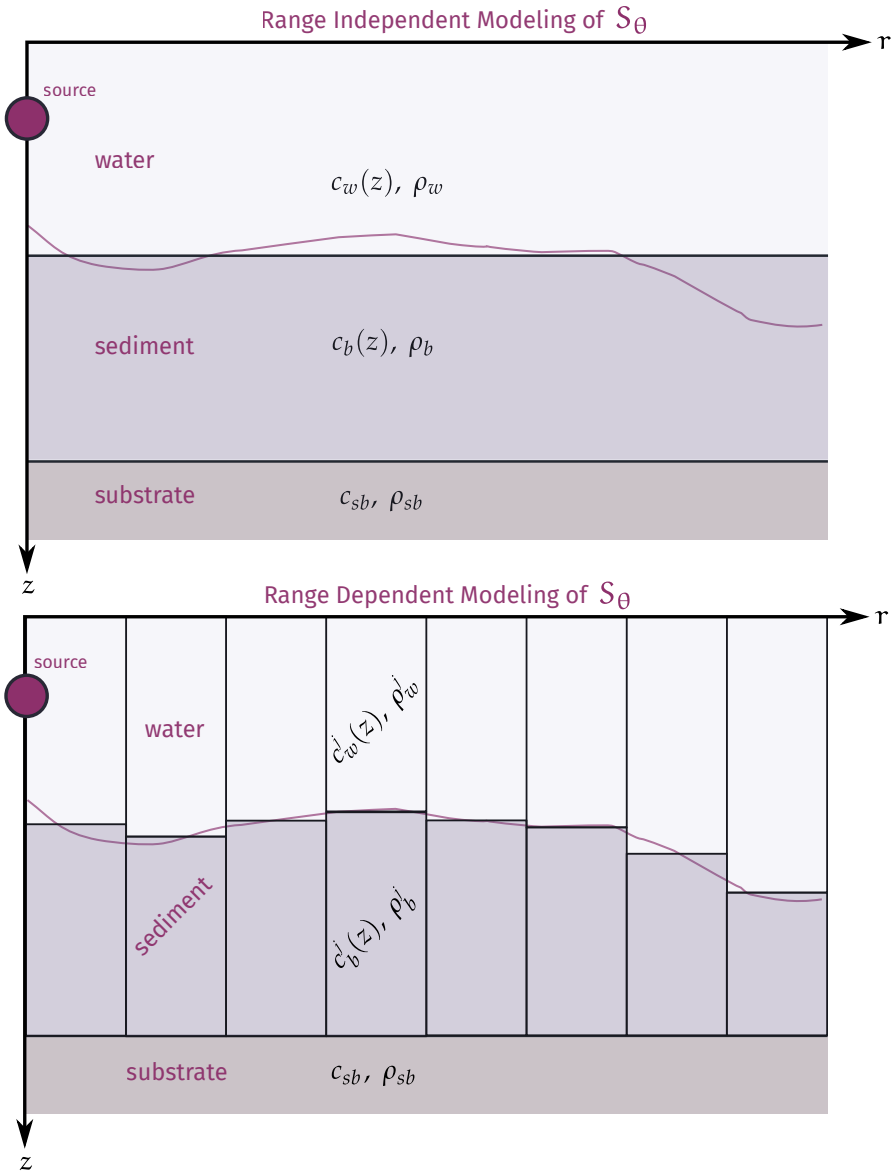


Figure 2.2: Two typical alternative models of a complex range-dependent ocean environment, corresponding to the vertical plane (slice) S_θ of the 3D environment of Figure 2.1.

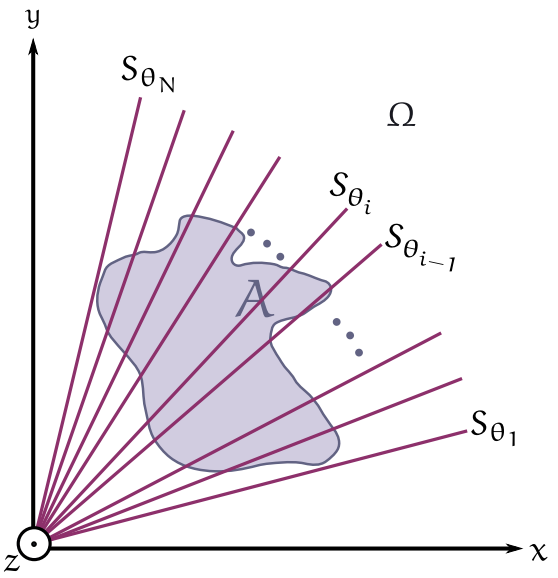


Figure 2.3: Top-view representation of an application of the Nx2D approach to a environment with a restricted anomaly.

to a domain under consideration Ω. The scattering due to the anomaly is restricted only in a sub-region of Ω, and typically a Nx2D consideration can give accurate results.

Figure 2.3 shows the geometry of an Nx2D approximation of an environment, with a restricted anomaly. In this geometry, the domain contains the anomaly is determined by the boundary planes $S_{\theta_1}, S_{\theta_N}$. The vertical slices S_{Θ_i} determine an axisymmetric problem in which case the pressure is expressed as

$$p_i(\mathbf{x}) = p_i(r,z), \quad \mathbf{x} \in S_{\theta_i},\ r \in [0,\infty),\ z \in [0,+\infty). \tag{2.5}$$

Then an appropriate interpolation procedure can combine the partial approximations of the acoustic field to give a united approximation of the acoustic pressure in Ω.

Similarly, an Nx2D approach can been considered for inverting the structure of the anomaly using acoustic means. Such a setup is presented in Figure 2.4.

It should be noted that the 2D environments intercepting the anomaly are generally treated as range dependent while the ones presenting (almost) constant characteristics are addressed by using a range-independent formulation.

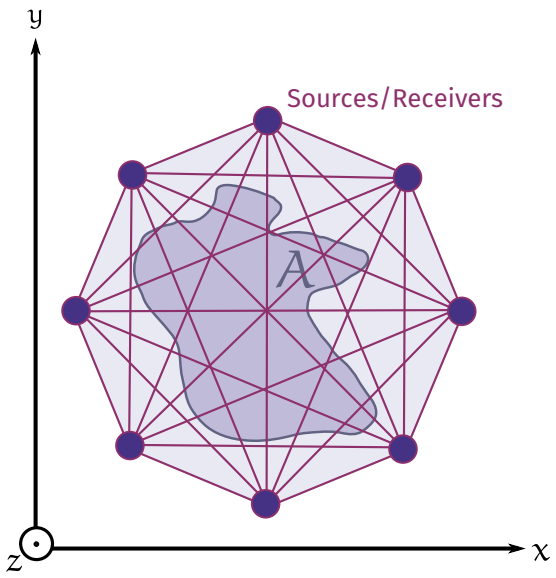


Figure 2.4: Top-view representation of an application of the Nx2D approach to retrieve information from a restricted anomaly.

2.3 Range Independent (RI) Environments

In this section we assume that the oceanic environment is horizontally stratified. In particular, the environmental parameters such as the sound speed profile and density are are invariant with range. In this work the stratification, is restricted to a $(L+2)$-layered fluid structure as described in Figure 2.5. The figure depicts an oceanic environment with L sediment layers over a semi-infinite sub-bottom. For most practical applications this is an acceptable representation of the marine environment.

Each layer is characterized by constant density and a sound speed that varies with, with the exception of the sound speed in substrate (sub-bottom) which is considered constant. The relations 2.7, 2.6 summarize these assumptions

$$c(z) = \begin{cases} c_w(z), & 0 \le z < h_0 \\ c_{b,1}(z), & h_0 \le z < h_1 \\ \quad \dots & \\ c_{b,L}(z), & h_{L-1} \le z < h_L \\ c_{sb}, & z \ge h_L \end{cases} \tag{2.6}$$

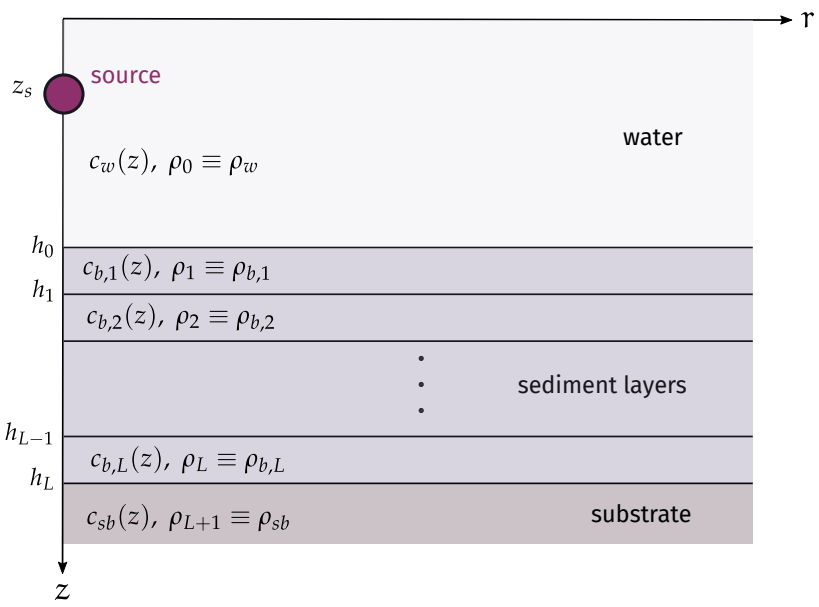


Figure 2.5: The range independent environment under consideration, consisting of $L+2$ fluid layers. The water column, L sediment layers and a semi-infinite substrate. The properties of the substrate are considered as constant.

$$\rho(z) = \begin{cases} \rho_0 \equiv \rho_w, & 0 \le z < h_0 \\ \rho_1 \equiv \rho_{b,1}, & h_0 \le z < h_1 \\ \quad \dots & \\ \rho_L \equiv \rho_{b,L}, & h_{L-1} \le z < h_L \\ \rho_{L+1} \equiv \rho_{sb}, & z \ge h_L \end{cases} \tag{2.7}$$

By considering of a harmonic, monochromatic point source in coordinates $(0, z_s)$ the spatial part of the wave equation is given by:

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial p(r,z)}{\partial r}\right) + \frac{\partial^2 p(r,z)}{\partial z^2} - \frac{1}{\rho(z)}\frac{\partial \rho(z)}{\partial z}\frac{\partial p(r,z)}{\partial z} + \frac{\omega^2}{c^2(z)}p(r,z) = -\frac{\delta(r)\delta(z-z_s)}{2\pi r} \tag{2.8}$$

Note that the term including the derivative of density with respect to depth is omitted when equation 2.8 is applied separately to each layer.

In order to find a unique solution for the acoustic pressure at each point of the environment, we have to complete the definition of the problem by providing a proper set of boundary conditions.

- Zero acoustic pressure at the water surface $(z = 0)$.

$$p(r, 0) = 0 \tag{2.9}$$

- Continuity of the pressure at each interface.

$$\lim_{z\to h_\ell^-} p(r,z) = \lim_{z\to h_\ell^+} p(r,z), \quad \ell = 0,\dots,L \tag{2.10}$$

- Continuity of the vertical component of particle velocity between each interface

$$\lim_{z\to h_\ell^-} \frac{1}{\rho_\ell}\frac{\partial p(r,z)}{\partial z} = \lim_{z\to h_\ell^+} \frac{1}{\rho_{\ell+1}}\frac{\partial p(r,z)}{\partial z}, \quad \ell = 0,\dots,L \tag{2.11}$$

- Zero acoustic pressure at infinite depth $(z \to \infty)$

$$\lim_{z\to\infty} p(r,z) = 0 \tag{2.12}$$

- The Sommerfeld radiation condition $(r \to \infty)$

$$\lim_{r\to\infty} \sqrt{r}\left(\frac{\partial}{\partial r} - i\frac{\omega}{c(z)}\right) p(r,z) = 0 \tag{2.13}$$

  Physically, this condition ensures that there is no incoming energy from infinity.

2.3.1 The depth problem

The following *Sturm-Liouvile* problem is called *depth problem*:

$$\rho(z)\frac{d}{dz}\left(\frac{1}{\rho(z)}u(z)\right) + \Big[\kappa^2(z) - \lambda\Big]u(z) = 0 \tag{2.14}$$

$$u(0) = 0 \tag{2.15}$$

$$\lim_{z\to h_\ell^-} u(z) = \lim_{z\to h_\ell^+} u(z), \quad \ell = 0,\dots,L \tag{2.16}$$

$$\lim_{z\to h_\ell^-} \frac{1}{\rho_\ell}\frac{du(z)}{dz} = \lim_{z\to h_\ell^+} \frac{1}{\rho_{\ell+1}}\frac{du(z)}{dz}, \quad \ell = 0,\dots,L \tag{2.17}$$

$$\lim_{z\to+\infty} u(z) = 0 \tag{2.18}$$

where $\kappa(z) = \omega/c(z)$.

The admissible values of λ_n consist the eigenvalues of the S-L problem and the associated functions $u_n(z)$ are the eigenfunctions.

Since the above problem is defined in a semi-infinite domain, it is singular [21]. For singular problems we know that there exist a set of discrete eigenvalues and a set of continues ones, consisting the so called "continuous spectrum" of the singular Sturm-Liouvile problem.

For singular Sturm-Liouvile problems the following representation theorem can be proved [22]

Theorem 2.3.1 (Representation theorem) *Let* $\{(u_n(z), \lambda_n),\ n = 1, \ldots, N\}$ *and* $\{(u(z;\lambda), \lambda),\ \lambda \in S \subset \mathbb{R}\}$ *the discrete and continuous spectra of a singular Sturm-Liouvile problem. Then the set of eigenvectors comprise a basis over the family of the continuous functions with piecewise continuous first derivatives, obeying the same boundary conditions as the S-L problem.*

By applying the above representation theorem to the Helmholtz equation, the spatial acoustic pressure $p(r,z)$ can be expressed in terms of the eigenfunctions of the problem as follows:

$$p(r,z) = \sum_{n=1}^{N} A_n(r)u_n(z) + \int_S b(r,\lambda)u(z;\lambda)d\lambda \tag{2.19}$$

where $A_n(r),\ b(r,\lambda)$ are coefficients to be determined.

Furthermore, far from the source the integral over the continuous spectrum of the eigenvalues does not contribute significantly to the acoustic field and consequently it can be omitted [2]. Therefore, we get the following approximation

$$p(r,z) \approx \sum_{n=1}^{N} A_n(r)u_n(z), \text{ for } r >> \kappa^{-1}(z), \tag{2.20}$$

where $\kappa^{-1}(z)$ is the reciprocal of the wavenumber.

At this point, let us state useful terminology

NORMAL MODES : The family of the eigenfunctions $\{u_n(z)\}_{n=1}^{N}$ correspond to the discrete spectrum of the eigenvalues of the depth problem. The normal modes are subject to the orthogonal relation

$$\int_0^{+\infty} \frac{1}{\rho(z)} u_m^*(z)u_n(z)dz = \delta_{mn}, \tag{2.21}$$

where $*$ denotes complex conjugate.

EIGENFORMS : The family of the coefficients $A_n(r)$, expressing the intensity of each normal mode in the expression of the acoustic field in any specific range r.

The solution of the *depth problem* is divided into two parts.

1. For $0 \leq z \leq h_L$, since the sound speed has been modelled to vary with depth in the water and in the sediment, there is no an analytical expression for the solution of the depth problem in these regions. An numerical approximation is to be considered.

2. On the contrary, for $z \geq h_L$ there is a closed form solution because of the constant sound speed profile.

— Solution of the depth problem in the substrate —

Lets recall that in the substrate the sound speed is assumed constant; therefore we can express the *wavenumber* in that region by a single variable κ_{sb}.

The general solution of equation 2.14 in the halfspace $[h_L, +\infty)$ after exploiting condition 2.18 is given by

$$u_n(z) = C_n \exp\left\{-\sqrt{k_n^2 - \kappa_{sb}^2}\, z\right\}, \text{ for all } z \geq h_L, \tag{2.22}$$

where $\lambda_n = k_n^2$ the eigenvalue corresponding to the n-th normal mode. Note that the constants C_n are to be computed after performing the normalization condition to the *normal modes*, once we have got the numerical approximation of the solution for $0 \leq z \leq h_L$.

— Discussion for the solution of the depth problem in the upper layers —

Since the sound velocity has been modeled to be varying with depth in the water and in the sediment, there is no an analytical expression for the solution of the depth problem in these regions.

An approximation of the solution of this problem is obtained by considering a discretization of the problem with respect to a vertical grid and applying a finite difference scheme, taking into account the normalization condition 2.21.

The numerical schemes for the estimation of the eigenvalues and the eigenfunctions u_n of the depth problem in water and sediment, is out of the scope of this work and they are not described here. For the implementation of such a scheme, refer to [2].

2.3.2 Asymptotic approximation of the acoustic pressure

By substituting the expansion of the pressure $\sum_n A_n(r)u_n(z)$ in the spatial wave equation 2.8 and taking into account that the modal functions $\{u_n(z)\}_{n=1}^N$ satisfy the depth problem, the eigenforms $\{A_n(r)\}_{n=1}^N$ can be obtained by employing the Green's function solution of a zero-order *Bessel equation* [22].For a unique solution the Green's function must obey a set of boundary condition consisting of the jump discontinuity of the first derivative at the source position and the radiation condition for $r \to \infty$. Eventually, the solutions are given in terms of Hankel functions of the first kind as

$$A_n(r) = \frac{i}{4\rho(z_s)} H_0^{(1)}(k_n r) u_n(z_s). \tag{2.23}$$

Thus,the modal solution for the generic waveguide described in Figure 2.1 is expressed by

MODAL SOLUTION

$$p(r,z) = \frac{i}{4\rho(z_s)} \sum_{n=1}^{N} H_0^{(1)}(k_n r) u_n(z_s) u_n(z). \tag{2.24}$$

Furthermore, if we replace the Hankel function by its asymptotic approximation for large arguments, we get a closed form for the acoustic pressure by means of fundamental mathematical functions as following:

$$p(r,z) = \frac{i}{4\rho(z_s)} \sum_{n=1}^{N} \sqrt{\frac{2}{\pi k_n r}} \exp\left[i\left(k_n r - \frac{\pi}{4}\right)\right] u_n(z_s) u_n(z). \tag{2.25}$$

2.4 Range Dependent (RD) Environments

In order to deal with more realistic oceanic or shallow-water environments, we have to relax the assumption of the constant properties of the media with the range.

In the subsequent analysis we will present a method for treating this problem based on the notion of "coupled modes". This approach is used in this work for obtaining the acoustic field in range dependent environments.

Here, we assume that the range dependent properties of the environment are of compact support and are confined between r_1 and r_J. Furthermore we assume that the marine environment is divided into $J+1$ cylindrical domains defined by the regions between two successive cylinders as illustrated in Figures 2.6. Each vertical slice in this axially symmetric environment is shown in Figures 2.7, 2.8. In each region defined by the cylinders the interfaces are taken to be flat, and horizontal, and the parameters of the environment constant with respect to range, similar to the range-independent case. This geometry is a step-wise representation of range-dependent general geometry.

2.4.1 The depth-problem in the j-th ring

In each region with index j ($j \neq 0$) we define a localized depth problem defined by the equation

$$\frac{d^2 u_j(z)}{dz^2} + \left[\kappa_j^2(z) - k_{jn}\right] u_j(z) = 0 \tag{2.26}$$

accompanied with boundary conditions as in 2.15-2.18.

Neglecting the continuous spectrum of the above eigenvalue problem, we express the acoustic pressure in the j-th ring by the sum

$$p_j(r,z) = \sum_{n=1}^{N} \Phi_{jn}(r) u_{jn}(z),\ j = 1, \ldots, J+1 \tag{2.27}$$

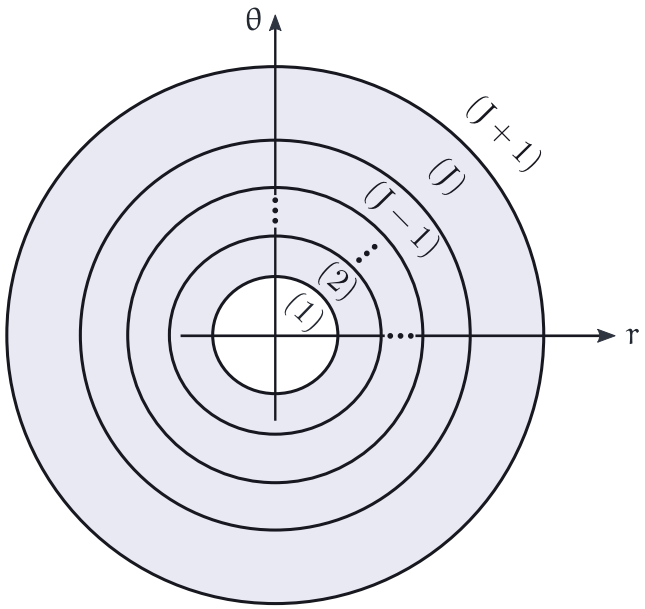


Figure 2.6: Top-view of the discretization of the range dependent environment in $J+1$ regions with each one range properties similar to the range independent case. Note that the regions 2 to J refer to the region including the singularity, whereas 1 and $J+1$ refer to the range-independent background environment.

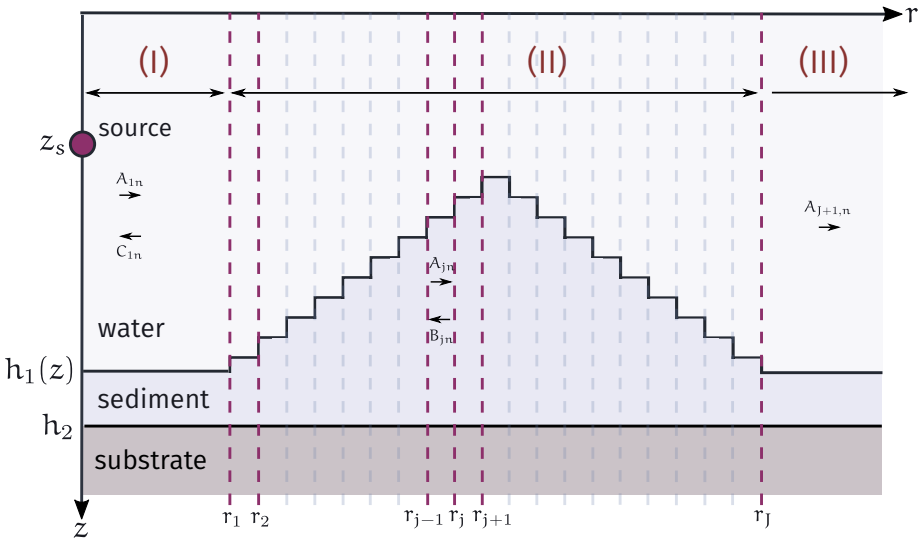


Figure 2.7: Approaching an axisymmetric sea mount using $J+1$ cylindrical ring areas of range independent properties.

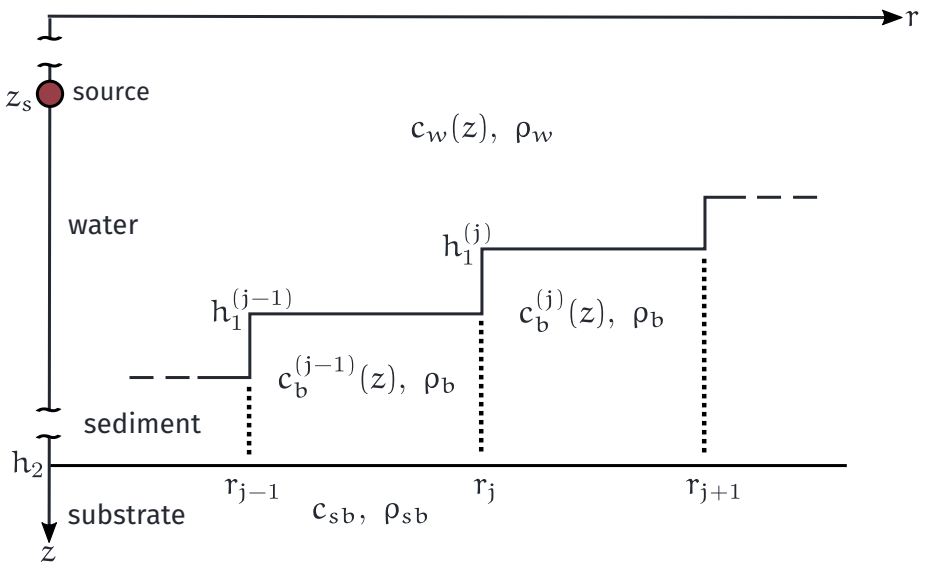


Figure 2.8: A zoomed version of Figure 2.7 around the area of the j, $j+1$ adjacent cylindrical rings.

where u_{jn} are the eigenfunctions of the depth-problem satisfying the following orthogonal relations

$$\int_0^{+\infty} \frac{1}{\rho_j(z)} u_{jn}(z) u_{jm}(z) dz = \delta_{nm},\ n,m = 1,\ldots,N. \tag{2.28}$$

In order to obtain a unique solution to the acoustic propagation problem we have to apply additional interface conditions, These express the continuity of the pressure p_j and the normal component of the particle velocity at the vertical boundaries defined at r_j. Thus,

$$p_j(r_j,z) = p_{j+1}(r_j,z) \tag{2.29}$$

$$\frac{1}{\rho_j(z)} \frac{\partial p_j(r_j,z)}{\partial r} = \frac{1}{\rho_{j+1}(z)} \frac{\partial p_{j+1}(r_j,z)}{\partial r}. \tag{2.30}$$

Note that, the eigenforms $\Phi_{jn}(r)$ take a different form in each region (I),(II), and (III), as they are shown in Figure 2.7 . In particular, we have

Following [19, 20] the acoustic pressure the subsequent forms in each on of the regions

REGION (I) $0 \le r \le r_1$: The acoustic field in the first region can be decomposed into a radiation field $p_r(r,z)$ plus a diffraction field $p_d(r,z)$ and it takes the form

$$\begin{aligned} p_1(r,z) &= \frac{i}{4\rho_1(z_s)} \sum_{n=1}^{N} u_{1n}(z_s) u_{1n}(z) H_0^{(1)}(k_{1n} r) \\ &+ \sum_{n=1}^{N} C_{1n} J_0(k_{1n} r) u_{1n}(z). \end{aligned} \tag{2.31}$$

REGION (II) $r_1 \leq r \leq r_J$: For the j-th ring, $2 \leq j \leq J$ we have an homogeneous Helmholtz equation, so the acoustic field is given by the following formula

$$\begin{aligned} p_j(r,z) &= \sum_{n=1}^{N} A_{jn} H_0^{(1)}(k_{jn} r) u_{jn}(z) \\ &+ \sum_{n=1}^{N} B_{jn} H_0^{(2)}(k_{jn} r) u_{jn}(z) \end{aligned} \tag{2.32}$$

REGION (III) $r \geq r_J$: Beyond the irregularity according to the Sommerfield radiation condition, there is no radiation from infinity, therefore

$$p_{J+1}(r,z) = \sum_{n=1}^{N} A_{J+1,n} H_0^{(1)}(k_{J+1,n} r) u_{J+1,n}(z) \tag{2.33}$$

2.4.2 Discussion for the calculation of the coupling coefficients

Using the conditions of continuity of pressure and particle velocity across the interfaces between the segments 2.29, 2.30, we obtain a linear system of equations with respect to the unknown coefficients of the series expansions 2.31-2.33.

equal in number with the unknowns of the problem and therefore have a unique solution due to the non singularity of the matrix of the linear system.

In this thesis, we skip the details about the treatment of that linear system of equations. However, in order to form the matrix of the linear system we have to find numerical approximations for the following integrals

$$C_{1jmn} = \int_0^{+\infty} \frac{1}{\rho_j(z)} u_{jm}(z) u_{j+1,n}(z) dz \tag{2.34}$$

$$C_{2jmn} = \int_0^{+\infty} \frac{1}{\rho_{j+1}(z)} u_{jm}(z) u_{j+1,n}(z) dz. \tag{2.35}$$

The quantities C_{1jmn}, C_{2jmn} are known by the term *coupling coefficients* and they express the amount of the energy, exchanged between any pair of modes belonging in the adjacent rings j and $j+1$.

2.5 Broadband Modelling

In this section we will present how the normal mode theory can be generalized when a broadband point source is considered. Since the spatial component of the acoustic pressure $p(r,z;\omega)$ consists the impulse response of the propagation system (Figure 2.9), then for the acoustic field $p'(r,z;\omega)$ due to a general point source is given by

$$p'(r,z;\omega) = p(r,z;\omega) S(\omega), \tag{2.36}$$

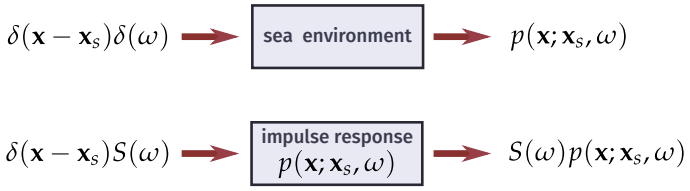


Figure 2.9: The linear systems governing the sound propagation in a marine environment.

where $S(\omega)$ is the source excitation function in the frequency domain and typically has its compact support in an closed interval $[\omega_I, \omega_F] \subset [0, +\infty)$.

2.5.1 Fourier synthesis

We can express the previous field in the time domain using the inverse Fourier transform:

$$P'(r,z;t) = \frac{1}{2\pi}\int_{\omega_I}^{\omega_F} S(\omega)p(r,z;\omega)\exp(-i\omega t)d\omega. \tag{2.37}$$

For practical applications we need to descretize the signals in the frequency and the time domain. Using a frequency step $\Delta\omega > 0$, we define the discrete frequencies

$$\omega_k = k\Delta\omega,\ k = 1,\dots,N, \tag{2.38}$$

where the number of frequencies N is typically chosen, so that $N = 2^J$ for a $J \in \mathbb{Z}^+$ and also $\omega_N > \omega_F$.

Then, we form the discrete signal $\hat{\mathbf{s}}$ consisting of the N values of the acoustic pressure in the frequency domain, so that

$$\hat{\mathbf{s}}[k] = \begin{cases} p'(r,z;\omega_k), & \text{if } \omega_k \in [\omega_I,\omega_F] \\ 0, & \text{otherwise} \end{cases},\ k = 1,\dots,N. \tag{2.39}$$

The discrete signal in the time domain $\mathbf{s}[n],\ n = 1,\dots,N$ is given by the real part of the inverse fast Fourier transform of $\hat{\mathbf{s}}$

$$\mathbf{s} = \mathrm{Re}\left\{\mathrm{FFT}^{-1}(\hat{\mathbf{s}})\right\}, \tag{2.40}$$

Note that the signal $\mathbf{s}$ is also composed by N samples. Each sample is given at a time interval $\Delta t = \dfrac{2\pi}{N * \Delta\omega}$.

2.5.2 Choosing the frequency step

The choice of the frequency step is very important for the adequate representation of the acoustic signal. If the frequency step is very large then cut-off effects will appear, while a small frequency step will increase significantly the computational cost.

The sampling angular frequency ω_s corresponding to the discrete signal $\mathbf{s}$ is given by $\omega_s = N\Delta\omega$.

Furthermore, according to the Nyquist criterion [23], the sampling frequency ω_s must be at least twice the highest frequency component ω_F, otherwise the frequency components corresponding to the spectrum $\omega \in [\omega_s, \omega_F]$ will not be correctly represented in the digital version of the signal.

Therefore, a necessary condition to get a discretized signal of adequate quality is

$$N\Delta\omega \geq 2\omega_F. \tag{2.41}$$

Note that an equivalent relation can be derived for frequencies in hertz.

2.5.3 Modal dispersion

Dispersion has an important effect on the propagation of sound through an ocean waveguide, as it is responsible for modulations of the shape of the original pulse emitted by the sound source. In normal mode theory the acoustic field is decomposed in a set of the so called *modal energy packets*, each one of these is due to an associated normal mode and it propagates with a velocity that varies with frequency.

The *phase velocity* of the m-th order mode is defined as

PHASE VELOCITY
$$u_p^{(m)} = \frac{\omega}{k_m}, \tag{2.42}$$

and it is always greater than the sound speed. This expresses the horizontal velocity at a specific frequency of the modal packet as defined by the plane wave expression of the propagating mode.

The energy carried by the normal mode m is propagating with a horizontal speed $u_g(m;\omega)$, called the *group velocity*. This velocity is given as follows

GROUP VELOCITY
$$\frac{1}{u_g^{(m)}} = \frac{dk_{rm}}{d\omega}. \tag{2.43}$$

The group velocity of a normal mode can be estimated using a finite-difference scheme for the estimation of derivative or in terms of an integral over its corresponding eigenfunction u_n as described by Chapman et al. [24].

The arrival time of the m-th modal energy packet at range r is given by

MODAL TIME ARRIVALS
$$t^{(m)} = \frac{r}{u_g^{(m)}}. \tag{2.44}$$

It should be noted that not all the modal packets are resolvable using a time-frequency analysis, if we have to deal with propagation through a highly dispersive channel, small differences in modal group velocities and at small ranges [25, 26, 27].

2.5.4 Source excitation functions concerned in this thesis

In the context of this thesis, we have considered two forms of point broadband acoustic sources. The first form simulates a Gaussian point source, whereas the second form was developed to model a light-bulb implosion witch has been used as the acoustic source for a real experiment.

— Gaussian source excitation function —

In our simulated inversion test cases, we have considered source excitation functions modelling Gaussian pulses in the frequency domain, given by

$$S(\omega) = \exp\left\{-4\pi \frac{(\omega - \omega_c)^2}{\Delta\Omega^2}\right\}, \tag{2.45}$$

where ω_c is the *central frequency* and $\Delta\Omega$ is the half-bandwidth of the source.

— Light-bulb implosion source excitation function —

For the case of the considering a light-bulb implosion that took place in SW06 experiment, the modeled source excitation function obeys the following formula

$$S(\omega) = \frac{\omega_c}{(\alpha + i\omega)^2 + \omega_c^2}, \tag{2.46}$$

where α and ω_c are determined by the conditions of the experiment. The details of the derivation can be found in Appendix A.

chapter 3

Inverse Problems in Underwater Acoustics (State of the Art)

Abstract

The previous chapter presented the forward problem of calculating the acoustic pressure in an ocean waveguide given the environmental and operational properties of some underwater acoustic application. Here, we present the reciprocal process of estimating the model parameters by exploiting the information given by the measurements of the acoustic field.

3.1 Introduction

The term *inverse problem* in contrast to *forward problem* describes the process of estimating parameters of some physical model by using data obtained through some appropriate experiment. Data and parameters are related through a mathematical model. We will refer to those parameters by the term *model parameters*.

Underwater inverse problems to be considered in this thesis can be classified as:

SOURCE LOCALIZATION : Source localization is referred to the problem of identifying the location of a sound source.

OCEAN ACOUSTIC TOMOGRAPHY : Ocean acoustic tomography is typically referred to the recovery of the sound speed profile $c(\mathbf{x})$ of the sea, through measurements of the acoustic field.

GEOACOUSTIC INVERSION : Geoacoustic inversion is referred to classification of the seabed of research by using acoustic means. Typically, a sea-bed is classified for its geoacoustic pa-

rameters which include the density of the sentiment layers ρ_i, the velocities of compressional and shear waves c_{b_i}, c_{s_i}, and the attenuation coefficients α_i.

The next section is dedicated to the formulation of discrete inverse problem.

3.2 General Formulation

First, we describe the measurements of an event by means of a D-dimensional vector $\mathbf{d}$:

$$\mathbf{d} = [d_1, \ldots, d_D]^T. \tag{3.1}$$

Similarly, the model parameters are represented by a M-dimensional vector $\mathbf{m}$:

$$\mathbf{m} = [m_1, \ldots, m_M]^T. \tag{3.2}$$

The measurements are related to the model parameters, through a generally nonlinear function $\mathbf{g} : \mathbb{R}^M \times \mathbb{R}^D \to \mathbb{R}^L$, such that

$$\mathbf{g}(\mathbf{m}, \mathbf{d}) = 0. \tag{3.3}$$

The above function called the *model* of the inverse problem and the problem is $g(\mathbf{m}, \mathbf{d})$ can consist of arbitrarily complicated (nonlinear) functions of the data and model parameters [28].

3.2.1 Linear inverse problems

Sometimes, it is possible to separate the measurements from the model parameters through a matrix $\mathbf{G} \in \mathbb{R}^{D \times M}$, such that

$$\mathbf{g}(\mathbf{m}, \mathbf{d}) = \mathbf{G}\mathbf{m} - \mathbf{d}. \tag{3.4}$$

This form of the function $\mathbf{g}$ leads to the following matrix equation

$$\mathbf{G}\mathbf{m} = \mathbf{d}. \tag{3.5}$$

The estimated values of the model parameters is to be given by means of a matrix $\mathbf{G}^{-\mathbf{g}} \in \mathbb{R}^{M \times D}$, called the *generalized inverse* of $\mathbf{G}$, as follows

$$\mathbf{m}^{\text{est}} = \mathbf{G}^{-\mathbf{g}}\,\mathbf{d}. \tag{3.6}$$

The form of the $\mathbf{G}^{-\mathbf{g}}$ depends on the properties of the matrix $\mathbf{G}$. Here, we shall restrict our analysis to three common classes of the explicit linear inverse problems, when no prior constraints of any type are applied.

EVEN-DETERMINED PROBLEMS : The simplest case is when we have the same number of linearly independent rows of and columns of $\mathbf{G}$. This condition is summarized by the following condition for the rank of the matrix

$$\text{rank}(\mathbf{G}) = D = M,$$

then as it is known from the linear algebra, the problem has a unique solution simply given through the inverse matrix $\mathbf{G}^{-1}$ of $\mathbf{G}$, thus the 3.6 takes the following form

$$\mathbf{m}^{\text{est}} = \mathbf{G}^{-1}\mathbf{d}. \tag{3.7}$$

UNDER-DETERMINED PROBLEMS : This case occurs when the matrix $\mathbf{G}$ does not provide enough information in order for a unique solution of the inverse problem to be determined. Therefore

$$rank(\mathbf{G}) = D < M.$$

Since there are infinity number of solutions, with the sense of the zero representation error, of such an inverse problem, to obtain a unique estimation we have to consider additive prior information for the solution. A common choice of an additive assumption is to consider as the solution of the inverse problem the one with the smallest length with respect the ℓ^2 norm. Thus

$$\mathbf{m}^{est} = \arg\min_{\mathbf{m}} \Big\{ \|\mathbf{m}\|_2 \mid \mathbf{G}\mathbf{m} = \mathbf{d} \Big\}. \tag{3.8}$$

Then by using basic linear algebra, we can obtain for the generalized inverse $\mathbf{G}^{-\mathbf{g}} = \mathbf{G}^T(\mathbf{G}\mathbf{G}^T)^{-1}$, thus

$$\mathbf{m}^{\text{est}} = \mathbf{G}^T(\mathbf{G}\mathbf{G}^T)^{-1}\mathbf{d}. \tag{3.9}$$

OVER-DETERMINED PROBLEMS : On the contrary to the previous case, in this class the information being carried by the matrix $\mathbf{G}$ is too much for the existence of an exact solution of that linear system. Here, we have the following condition for the rank of $\mathbf{G}$

$$\text{rank}(\mathbf{G}) = M < D.$$

Since we always have a non-zero estimation error with respect to the Euclidean norm, its make sense to seek for the model parameter vector $\mathbf{m}^{\text{est}}$ that minimizes this error. Therefore, we have to solve the following minimization problem for the least square error

$$\mathbf{m}^{\text{est}} = \arg\min_{\mathbf{m}} \Big\{ (\mathbf{G}\mathbf{m} - \mathbf{d})^T(\mathbf{G}\mathbf{m} - \mathbf{d}) \Big\}. \tag{3.10}$$

Then, the so called *least square solution* is given by the following closed formula

$$\mathbf{m}^{\text{est}} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{d}. \tag{3.11}$$

MIXED-DETERMINED PROBLEMS : In practice, most of the real inverse problems that appear are neither completely overdetermined nor underdetermined. In such problems it is common for some of the recoverable parameters to appear only in combinations (e.g. mean values) in the formulation of the forward problem. In this case these combined model parameters can not be recovered individually. The solution of a Mixed-determined problem can be obtained by transforming the original problem $\mathbf{Gm} = \mathbf{d}$ to a new one $\mathbf{G}'\mathbf{m}' = \mathbf{d}'$, where $\mathbf{m}'$ is partitioned into an upper part $\mathbf{m}'_o$ which is overdetermined and a lower part $\mathbf{m}'_u$ that is underdetermined. After the partitioning the inverse problem can be written as:

$$\begin{bmatrix} \mathbf{G}'_o & \mathbf{0} \\ \mathbf{0} & \mathbf{G}'_u \end{bmatrix} \begin{bmatrix} \mathbf{m}'_o \\ \mathbf{m}'_u \end{bmatrix} = \begin{bmatrix} \mathbf{d}'_o \\ \mathbf{d}'_u \end{bmatrix}. \tag{3.12}$$

The partitioning process can be performed by applying singular value decomposition (SVD) on $\mathbf{G}$.

3.2.2 Non-linear inverse problems

When the linearization of an inverse problem is not possible, we have to consider alternative methods for solving the non-linear optimization problem described by the equation 3.3. In most case nonlinear are formulated as optimization problems that can be solved by a variety of methods. Some common choices of such methods are the exhaustive searching, the Newton's method, the gradient descent approach, non-linear programming techniques, genetic algorithms, simulated annealing, and neural networks.

Here we will describe in brief how we shall deal with such non-linear inverse problems in this work.

First, we introduce a norm (or pseudo-norm) function $\nu : \mathbb{R}^L \to \mathbb{R}^+$. We define a scalar function $f : \mathbb{R}^{M\times D} \to \mathbb{R}^+$ through that norm as

$$f(\mathbf{m}, \mathbf{d}) \equiv (\nu \circ \mathbf{g})(\mathbf{m}, \mathbf{d}). \tag{3.13}$$

We shall refer to this function f as the *objective function* of the inverse problem. The inverse problem is then transformed to the following optimization problem

$$\mathbf{m}^{\text{est}} = \arg\min_{\mathbf{m}} f(\mathbf{m}, \mathbf{d}). \tag{3.14}$$

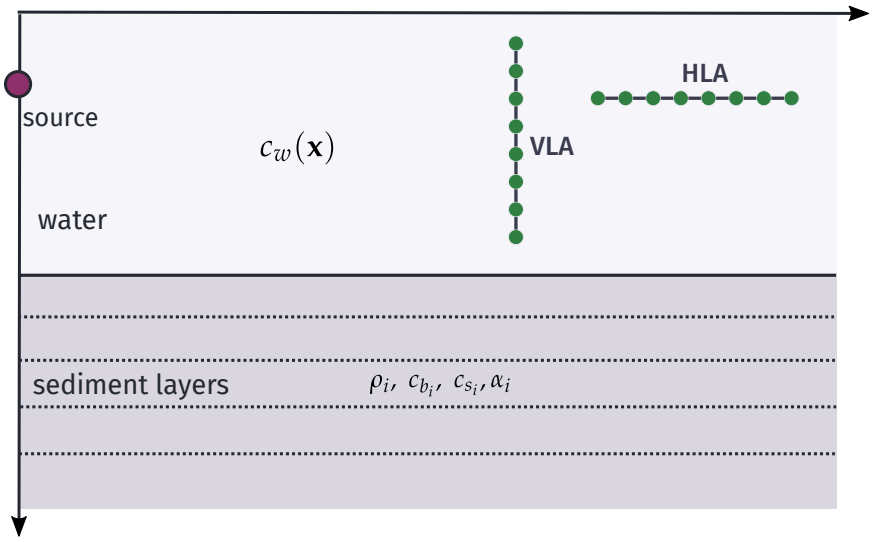


Figure 3.1: VLA and HLA of receivers.

The method that we will employ for solving this problem is a Genetic algorithm, the implementation of which discussed in section 4.7.

3.3 Geoacoustic Inversion and Ocean Acoustic Tomography

Most geoacoustic inversion and ocean tomography studies utilize data obtained by a vertical (VLA) [29, 30, 31, 32, 33] or horizontal (HLA) [34, 35] arrays of receivers (Figure 3.1). Using an array of receivers (hydrophones) we have the possibility to exploit data obtained at different locations, thus enabling the use of the spatial properties of the acoustic field as additional conditions for solving the inverse problem. The drawbacks of using such a setup for estimating unknown environmental parameters, include the cost of the equipment as well as the operational difficulties. A popular inversion method associated with the use of an array of hydrophones is the matched field processing [36].

A less expensive setup exploits a single receiver for inversions. Using a single-receiver the data which are the input of the inverse problem are associated with specific observables of the acoustic field which are measured in the time domain. The observables are identified other directly in the received signal or indirectly after some processing of the signal. Among the observables obtained directly in the measured signal we mention ray arrivals, modal arrivals and among the observables obtained after post processing we will mention dispersion curves and statistical features. Some of the methods to utilized for the solution of the inverse problem using both configurations as above are briefly sketched at the following sections. It should be noted that the analysis presented in this and the following chapter is referred to a geometry at a single vertical slice. Thus the parameters to be estimated may function of range and depth

but not the azimuthal angle. Reconstructions of the properties in 3D can be obtained by combining results obtained at several vertical slices as in Fig 1.4. Also, methods presenting here exclude those applied with data obtained by means of typical sonar equipment mounted at toed or autonomous underwater vehicles. It should be noted that geoacoustic inversions can also be obtained by means of Ocean Bottom Seismograms (OBS) or equivalent equipment, by exploiting Scholte waves [37, 38].

In the subsequent sections we will briefly describe matched-field inversion methods as well as methods related to observables obtained at a single hydrophone (ray inversion, modal travel time, statistic).

3.3.1 Matched-Field Processing (MFP)

In matched-field processing the observables are correlated with replicas from a propagation model for various values of the recoverable parameters using a suitable objective function. Here, we present briefly the basic concept of MFP considering the Bartlett linear processor as the objective function to provide similarity measurements between the observables (acoustic pressure in the frequency domain) with the replica data produced by a simulated geometry and a parameter vector $\mathbf{m}$.

This linear processor is given by

BARTLETT LINEAR PROCESSOR

$$\mathcal{P}_{\text{lin}}(\mathbf{m};\omega) = \|\mathbf{w}^H(\mathbf{m};\omega)\mathbf{d}(\omega)\|_2^2. \tag{3.15}$$

where $\mathbf{d}$ is the vector of the measured acoustic fields corresponding to the angular frequency ω at the VLA (or HLA) hydrophones, $\mathbf{w}(\mathbf{m};\omega) = \frac{p(\mathbf{m};\omega)}{|p(\mathbf{m};\omega)|}$ are the normalized replica fields for the model defined from the parameter vector $\mathbf{m} \in \mathcal{M}$ and index H denotes complex conjugate transpose.

$$\mathbf{m}^{\text{est}} = \arg\max_{\mathbf{m}} \sum_{\omega_i \in \Omega} \mathcal{P}_{\text{lin}}(\mathbf{m};\omega_i) \tag{3.16}$$

where $\mathcal{M}$ is the set of all different combinations of the model parameters in the search space and Ω is the set of the considered frequencies. For a detailed presentation of MPF an interest reader can refer to some early papers [36, 39, 40, 41, 42, 43, 44].

3.3.2 Bayesian inversions

The Bayesian approach for estimating the model parameters is widely used in inverse problems of acoustical oceanography, exploiting characteristic observables derived from measurements of the acoustic field. The estimate solution to a inverse problem using a Bayesian ap-

proach is given in terms of the posterior probability distribution over the recoverable parameters.

The Bayesian approach is based on the Bayes dependency between the observables $\mathbf{d}$ and the recoverable parameters $\mathbf{m}$. The posterior probability distribution of the model parameters $\mathbf{m}$ can be proportionally expressed as the product of the likelihood function $\mathcal{L}(\mathbf{d}|\mathbf{m})$ and the prior information $P(\mathbf{m})$ of the model parameters, therefore

$$P(\mathbf{m}|\mathbf{d}) \propto \mathcal{L}(\mathbf{d}|\mathbf{m})P(\mathbf{m}). \tag{3.17}$$

Moreover, the likelihood function can be expressed in terms of an appropriate misfit function $E(\mathbf{m})$ as

$$\mathcal{L}(\mathbf{d}|\mathbf{m}) \propto \exp[-E(\mathbf{m})]. \tag{3.18}$$

The posterior probability distribution of $\mathbf{m}$ is eventually occurred to be

$$P(\mathbf{m}|\mathbf{d}) = \frac{\exp[-E(\mathbf{m})]P(\mathbf{m})}{\int_{\mathcal{M}} \exp[-E(\mathbf{m}')]P(\mathbf{m}')d\mathbf{m}'}, \tag{3.19}$$

therefore, the marginal probability distribution are given mathematically as

$$P(m_i|\mathbf{d}) = \int_{\mathcal{M}} \delta(m_i' - m_i)P(\mathbf{m}'|\mathbf{d})d\mathbf{m}', \tag{3.20}$$

where δ is the Dirac delta function.

A proper estimation of the mode parameters can be the vector $\mathbf{m}^{\text{est}}$ which minimizes the joint posterior distribution, thus

$$\mathbf{m}^{\text{est}} = \arg\max_{m\in\mathcal{M}} p(\mathbf{m}|\mathbf{d}). \tag{3.21}$$

In order Bayesian inference to be applicable, it is necessary the consideration of a random variable that expresses the misfit between the measurements and the replica data obtained by solving the forward model.

This brief outline presents just the basic concepts of the Bayesian inference. For a more detailed presentation on Bayesian inversions one can refer to papers by Dosso et al. [45, 46, 47, 48, 49].

3.3.3 Modal-phase inversions

The *modal phase* at range r is defined by the following expression

$$\Phi_n(r) = \int_0^r k_n(r')dr', \tag{3.22}$$

where $k_n(r)$ is the n-th eigenvalue of the "depth problem" at range r. Using this expression it is assumed that eigenvalues of the depth problem change with range.

The modal phase can be calculated using the acoustic field measured at a vertical array of receivers through an appropriate "mode filtering technique" as suggested by Lo et al. [50].

Once the modal phase is measured, the variation or the phase, $\delta\Phi_n(r)$ with respect to a reference (background) environment, can be associated with the variance of the sound speed profile $\delta c(r,z)$ in the water column and the seabed using the following formula:

$$\delta\Phi_n(r) = \int_0^r \int_0^{+\infty} Q_n(z;r')\delta c(r',z)dzdr', \tag{3.23}$$

The kernel $Q_n(z;r)$ is calculated for the model parameters of the reference environment.

Following a proper discretization of the sound speed variation in range and depth Equation 3.23 is replaced by a linear equation that formulates a discrete linear problem of the form:

$$\mathbf{G}_1\delta\mathbf{c} = \delta\mathbf{\Phi}. \tag{3.24}$$

which can be solved to obtain the vector of the discretized sound speed variations $\delta\mathbf{c}$. Note that, in order that this approach is applied, an array of hydrophones is needed to determine the modal structure in the frequency domain. Both theoretical [26, 51, 52] and experimental studies [53] of the method have been reported. Despite the good results obtained by using this method for ocean acoustic tomography and geoacoustic inversion problems, the method suffers from problems related to the accuracy of obtaining the modal phase which is essential for the accuracy of the inversion results as well as the ill-conditioning of the kernel matrix in Equation 3.24.

3.3.4 Dispersion analysis

In this class of inversion techniques the observables are obtained after utilizing certain group velocity dispersion characteristics of the acoustic field. Dispersion-based methods for acoustical oceanography problems have been applied by Potty et al. [54, 55] who solved an optimization problem based on the difference between the measured arrival times of the propagated modes at various frequencies using measured and predicted dispersion curves. The dispersion curves are obtained using time-frequency analysis of the acoustic signal measured in the time domain and are based on the group velocities of the propagating modes withing the frequency range of the acoustic signals. Typical dispersion curves of a acoustic signal is shown in Figure 3.2. A problem with the use of the dispersion curves for the inversions arise when separability among the different dispersion curves is a difficult task. Taroudakis and Tzagkarakis suggested a method to improve mode separability [56] and recently, Bonnel et al. [57, 58] have

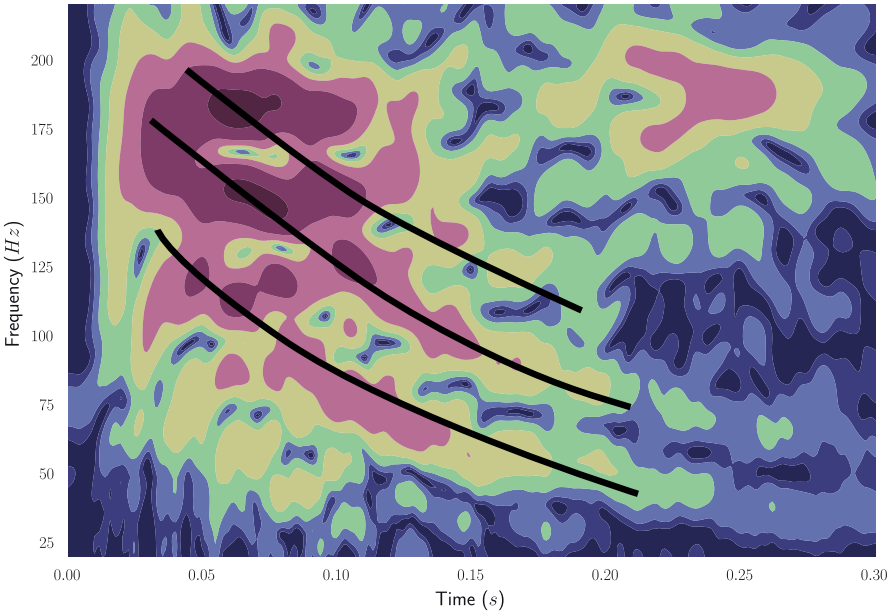


Figure 3.2: Typical dispersion curves of an actual recorded underwater acoustic signal.

introduced an alternative method called time-warping technique, and associated with an appropriate optimization problem for the efficient identification of the dispersion curves.

3.3.5 Travel-time tomography

The following methods are referred to an experimental setup employing a single receiver.

— Ray inversions —

Traditionally, inverse problems in underwater acoustics were based on ray theory. In this method, it was the ray travel time due to a broadband source, which gave the necessary information for the estimation of the recoverable parameters. This method has been extensively used especially in deep-water sea environments with good results.

Here, we present a typical application of this method for retrieving the sound speed profile. By linearizing the problem with respect to a known background state, the travel time variation $\delta\tau_i$ along a certain ray Γ_i is associated with the sound speed variation $\delta c(\mathbf{X})$ through the following formula

$$\delta\tau_i = \int_{\Gamma_i} \frac{\delta c(\mathbf{x})}{c_0^2(\mathbf{x})} ds, \tag{3.25}$$

with c_0 denoting the sound speed profile for the background state and $\mathbf{x}$ are the spatial variables.

Provided that ray arrivals could be identified (automatically or manually) in the recorded acoustic signal, a set of measurements can be used for the recovery of the sound speed profile along the specific ray path. The problem is normally solved by discretization of the ray path and sometimes employing empirical orthogonal functions to describe the variations of the sound speed profile. The inverse problem is therefore formulated as a linear inverse problem of the form:

$$\delta\boldsymbol{\tau} = \mathbf{G}_2^0 \delta\mathbf{m} \tag{3.26}$$

where $\mathbf{G}_2^0$ is the kernel matrix, and $\delta\mathbf{m}$ is the perturbations of the recoverable parameters. This method is still used by many researchers for ocean acoustic tomography. For more information about the method the reader can refer to publications [59, 25, 27].

— Peak travel time inversions —

The amplitude of the acoustic pressure at the receiver location depends of course on the model parameters $\mathbf{m}$ of the environment. Therefore, we denote the amplitude of a signal received in the time domain as $a(t;\mathbf{m})$.

The local maxima of the signal calculated of the values of the model parameters of the reference environment satisfy the equation

$$\frac{\partial}{\partial t}\alpha(t_\ell; \mathbf{m}^0) = 0, \tag{3.27}$$

where $\mathbf{m}^0$ are the model parameters of the reference environment. A similar expression holds for the peaks of the measured signal which corresponds to the model parameters $\mathbf{m}$. Assuming that the reference environment is closed to the actual one, the relationship of the form can be written associating the perturbations of the arrival times of the peaks measured with the peaks of the signal calculated for the reference environment

$$\delta\boldsymbol{\tau} = \mathbf{G}_3^0 \delta\mathbf{m}. \tag{3.28}$$

where the kernel $\mathbf{G}_3^0$ has elements based on the properties of the background environment, and they are given by

$$(G_3^0)_{i,j} = \frac{\partial \alpha_i}{\partial m_j}(\mathbf{m}^0). \tag{3.29}$$

Using this approach, a linear relationship between travel time variations of the peaks and associated sound speed variations is defined [60]. This approach has been used extensively for processing the data of the THETIS and THETIS-2 experiments in the Mediterranean Sea with good results [61, 60, 62].

— Modal travel time inversions —

An alternative approach is to identify modal arrivals instead of rays. The peaks of the signal corresponding to modal arrivals are denoted as t_n. Using the notion of the group velocity

$$u_g^{(n)} = \left.\frac{\partial \omega}{\partial k_n}\right|_{\omega_0}, \tag{3.30}$$

where k_n is the n-th eigenvalue of the depth problem, when the normal mode solution is considered.

The arrival time of the n-th mode is $r/u_g^{(n)}$ assuming a range-independent, An integration over range defines the arrival time in range-dependent environments.

It can be shown that the modal travel time variations in terms of the sound speed variations is given by the following formula

$$\delta\tau_n = \int_S \left.\frac{\partial Q_n}{\partial \omega}\right|_{\omega_0} \delta c(\mathbf{x}) d\mathbf{x}, \tag{3.31}$$

and the integration is over the area of the sound speed variation. The kernel Q_n calculated for the parameters of the background environment [26] is the same as in Equation 3.23.

Following a proper discretization of the sound speed variation in range and depth Equation 3.23 is replaced by a linear equation that formulates a discrete linear problem of the form:

$$\delta\boldsymbol{\tau} = \mathbf{G}_4^0 \delta\mathbf{c} \tag{3.32}$$

which can be solved to obtain the vector of the discretized sound speed variations $\delta\mathbf{c}$, where $\delta\boldsymbol{\tau}$ is the vector of the travel time variations by comparing actual modal arrivals and modelled modal arrivals for the reference environment. The system of Equation 3.32 is solved as usual.

In order that this method is applicable a certain number of model arrival should be identified. The identification of the modal arrivals is not always easy. A method of semi-automatic identification of the modal arrivals was presented by Taroudakis [63]. Modal arrivals can also be exploited in terms of an optimization scheme [64] or a hybrid scheme where a nonlinear inversion based on modal arrivals leads to a background environment which can be utilized to a linear scheme presented above [65].

3.3.6 The Statistical Signal Characterization Scheme (SSCS)

The SSCS was introduced by Taroudakis et al. [1] as a way to define signal observables especially in cases that typical observables such as ray arrivals or modal arrivals cannot be identified in the recorded signals. Moreover, the setting of the associated inverse problem requires just a single hydrophone, which makes its application cheap in comparison with signal inversion methods requiring reception at an array of hydrophones. The SSCS is based on a wavelet transform of the signal at various decomposition levels, followed by the statistical modeling of the wavelet sub-band coefficients.

The objective of the SSCS is the modeling of the signal wavelet sub-band coefficients by suitable statistical distributions. Motivated by the corresponding modeling of texture images, Taroudakis et al. demonstrated that the symmetric alpha-stable (SaS) distributions are capable of modeling the wavelet coefficients of underwater acoustic signals typically used in applications of ocean acoustic tomography.

Similarity measurements between two acoustic signals are obtained by means of the Kullback-Leibler divergence (KLD) [66].

So far, three alternative inversion approaches have been applied for the scheme. The first one was based on neural networks [67] and although first applications for geoacoustic inversions gave reliable results, are considered not appropriate for solving multidimensional inverse

problems. In subsequent applications, the optimization procedure has been controlled by a genetic algorithm (GA), providing good results using both synthetic and real data. Furthermore, a preliminary study of a Bayesian driven optimization has been tested with promising results [68].

Taroudakis et al. have also presented hybrid inversion schemes that use the SSCS over a wide search space for deriving either a narrower search space followed by a typical matched field processing [69]. or by defining a rough initial guess of the recoverable parameters followed by a fine-tuning procedure using the identification of the modal arrivals in case that is possible [70].

The SSCS is actually the motivation of the PhD thesis. In this respect it is analytically described in the fifth chapter of the thesis. Also, the sixth chapter includes some characteristic applications of the scheme in inverse problems of acoustical oceanography.

chapter 4

MATHEMATICAL BACKGROUND

Abstract

In this chapter, we review the mathematical frameworks which we will employ in the subsequent analysis of the proposed characterization and inversion schemes.

4.1 Fourier Transform

In this chapter, t will denote time and ω will denote angular frequency.

L^p-NORM : If $p > 0$ and if $f(t)$ is a measurable function on $\mathbb{R}$, the following quantity is called the L^p-norm of f:

$$\|f\|_p = \left\{ \int_{-\infty}^{\infty} |f(t)|^p dt \right\}^{1/p}. \tag{4.1}$$

L^p-SPACE : The set of all measurable functions of finite L^p-norm consist the $L^p(\mathbb{R})$ space. Thus:

$$L^p(\mathbb{R}) = \{f \text{ measurable on } \mathbb{R} \ : \ \|f\|_p < \infty\}. \tag{4.2}$$

4.1.1 Continuous Fourier Transform (CFT)

Given a function $f(t) \in L^1(\mathbb{R})$, the function $\hat{f}(\omega)$ calculated by the following integral is well defined for all $\omega \in \mathbb{R}$, and it is called the *Fourier Transform* of f.

(CFT)
$$\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt, \quad \omega \in \mathbb{R}. \tag{4.3}$$

Moreover, if $f(t), \hat{f}(\omega) \in L^1(\mathbb{R})$ then the integral $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{f}(\omega) e^{i\omega t} d\omega$ is well defined and it equals to $f(t)$ almost everywhere[1] and it is called the *Inverse Fourier Transform (ICFT)* of f, thus:

(ICFT)
$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{f}(\omega) e^{-i\omega t} d\omega, \quad \text{for almost all } t \in \mathbb{R}. \tag{4.4}$$

4.1.2 Discrete Fourier Transform (DFT)

Let us consider a discrete signal of finite duration signal $\mathbf{s}[n],\ n = 0, \ldots, N-1$. The discrete analog of (4.3) is given as follows:

(DFT)
$$\hat{\mathbf{s}}[k] = \sum_{n=0}^{N-1} \mathbf{s}[n] \exp\left(-i\frac{2\pi k}{N} n\right), \quad k = 0, \ldots, N-1. \tag{4.5}$$

Using the discrete Fourier transform the signal is sampled at frequencies

$$\omega_k = \frac{2\pi k}{N}.\ k = 0, \ldots, N-1. \tag{4.6}$$

Additionally, we have the following reconstruction formula, which is the discrete equivalent of inverse Fourier transform:

(IDFT)
$$\mathbf{s}[n] = \frac{1}{N} \sum_{k=0}^{N-1} \hat{\mathbf{s}}[k] \exp\left(i\frac{2\pi k}{N} n\right), \quad n = 0, \ldots, N-1. \tag{4.7}$$

4.1.3 Discrete Windowed Fourier Transform

Consider again a digital signal $\mathbf{s}[n],\ n = 0, \ldots, N-1$. As we have seen before the discrete Fourier transform provides information about the signal at various discrete frequencies. This is useful if the signal has similar form over time (stationary signal).

In case when the signal's spectral characteristics vary in time, we are interested in adopting a representation which provides information in both time and frequency domains. The simplest treatment is given by means of the *windowed Fourier transform*, where the signal $\mathbf{s}$ is represented by its transformed version $\mathbf{S}(m,k)$ by evaluating for any $m = 0, \ldots, N-1$, the discrete Fourier transform corresponding to the signal when it is observed through a narrow window with energy concentrated around its m-th sample.

The window Fourier transform is mathematically described by the following expression

(DWFT)
$$\mathbf{S}(m,k) = \sum_{n=0}^{N} \mathbf{s}[n] \mathbf{w}[n-m] \exp\left(-i\frac{2\pi k}{N} n\right), \quad k, m = 0, \ldots, N-1. \tag{4.8}$$

1 Two functions f and g defined on a set E are equal almost everywhere if the set $\{x \in E : f(x) \neq g(x)\}$ has measure zero.

where $\mathbf{w}[n]$ denotes a window function with energy concentrated around zero.

The corresponding inverse formula is written as:

(IDWFT)
$$\mathbf{s}[n] = \sum_{m=0}^{N-1} \mathbf{w}[n-m] \sum_{k=0}^{N-1} \mathbf{S}(m,k) \exp\left(i\frac{2\pi k}{N}n\right), \quad n = 0, \ldots, N-1 \tag{4.9}$$

4.2 Wavelet Transform

The Fourier transform is a widely used tool for decomposing a signal to a superposition of sine waves, each one with a specific frequency. However, real world signals often exhibit time variations of their properties such as specific trends and abrupt changes. By applying the Fourier transform in such signals, we miss important information about the variation of their structures. Therefore, if we would like to characterize signals with structures that vary in time, it is necessary the decomposition uses a class of functions that are well localized in both time and frequency domains. In this work, we shall use the Wavelet transform for extracting features of underwater acoustic signals.

4.2.1 Continuous Wavelet Transform (CWT)

(*CWT*) If a function $\psi(t) \in L^2(\mathbb{R})$ satisfies the condition,

$$C_\psi = \int_0^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega < +\infty, \tag{4.10}$$

where $\hat{\psi}(\omega)$ is the Fourier transform of $\psi(t)$, then $\psi(t)$ is called a *mother wavelet*. The continuous wavelet transform of a function $f(t)$, relative to the mother wavelet, is defined by the following expression:

(CWT)
$$(W_\psi f)(u,s) = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{s}} \psi^*\left(\frac{t-u}{s}\right) dt, \; f \in L^2(\mathbb{R}), \tag{4.11}$$

where $u \in \mathbb{R}$ expresses the time and $s \in \mathbb{R}^+$ the scale parameter. Also, ψ^* denotes the conjugate of the ψ.

Furthermore, by setting

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right), \tag{4.12}$$

the formula 4.11 takes the compact form:

$$(W_\psi f)(u,s) = \int_{-\infty}^{+\infty} f(t) \psi^*_{u,s}(t) dt = \langle f, \psi_{u,s} \rangle. \tag{4.13}$$

Note that, provided that ψ is centered at 0, $(W_\psi f)(u,s)$ carries local information of f in the following time-frequency window

$$[u - s\Delta_\psi, u + s\Delta_\psi] \times \left[\frac{\omega_c}{s} - \frac{1}{s}\Delta_{\hat{\psi}}, \frac{\omega_c}{s} + \frac{1}{s}\Delta_{\hat{\psi}}\right], \tag{4.14}$$

where ω_c denotes the central frequency of $\hat{\psi}(\omega)$ which is given by

$$\omega_c = \frac{1}{2\pi}\int_0^{+\infty} \omega|\hat{\psi}(\omega)|^2 d\omega. \tag{4.15}$$

We can reconstruct any function $f \in L^2(\mathbb{R})$ from its CWT using the following formula

(ICWT)
$$f(t) = \frac{2}{C_\psi}\int_0^{+\infty}\int_{-\infty}^{+\infty} (W_\psi)f(u,s)\psi_{u,s}(t)du\frac{ds}{s^2}. \tag{4.16}$$

4.2.2 Wavelet bases

(*Multi-Resolution Analysis (MRA)*) A sequence of closed subspaces $\{V_j\}_{j\in\mathbb{Z}}$ of $L^2(\mathbb{R})$ is called a MRA if it satisfies the following properties:

1. $V_j \subset V_{j-1}$, for all $j \in \mathbb{Z}$
2. $lim_{j\to-\infty} V_j = \text{closure}(\bigcup_{j\in\mathbb{Z}} V_j) = L^2(\mathbb{R})$
3. $lim_{j\to\infty} V_j = \bigcap_{j\in\mathbb{Z}} V_j = \emptyset$
4. $f(t) \in V_j \iff f(2^j t) \in V_0$, for all $j \in \mathbb{Z}$
5. $f(t) \in V_0 \iff f(t-n) \in V_0$, for all $n \in \mathbb{Z}$
6. There exists a function $\phi(t) \in V_0$, called as *scaling function*, such that the set $\{\phi(t-k)\}_{k\in Z}$ to be an orthonormal basis in V_0.

Let $\{V_j\}_{j\in\mathbb{Z}}$ be a MRA and $\phi(t) \in V_0$ be its corresponding scaling function. We define a family of modified scaling function $\phi_{k,j}$ such that

$$\phi_{n,j}(t) = 2^{-j/2}\phi(2^{-j}t - n),\ n, j \in \mathbb{Z}. \tag{4.17}$$

Using properties 4, 5 and 6 of the MRA one can prove that the set $\{\phi_{k,j}\}_{k\in\mathbb{Z}}$ constitutes an orthonormal base in the so called *approximation space* V_j. The projection of $f(t)$ onto V_j is given as follows:

$$P_{V_j} f(t) = \sum_{k=-\infty}^{+\infty} \mathbf{a}_j[k]\phi_{k,j}(t)\ , \tag{4.18}$$

where $\mathbf{a}_j[n] = \langle f(t), \phi_{n,j}\rangle$ the *approximation coefficients* of the function f in the resolution 2^{-j}.

Using the definitions for convolution and $\phi_{n,j}$, we see that the approximation coefficients can be expressed as follows:

$$\mathbf{a}_j[n] = \Big(f * \phi_{0,j}\Big)[2^j n]. \tag{4.19}$$

Knowing that the spectrum of ϕ is generally concentrated in a subset of $[-\pi, \pi]$, it is easy to prove that $\phi_{n,j}$ has its effective spectrum in a subset of $[-2^{-j}\pi, 2^{-j}\pi]$. Therefore, $\mathbf{a}_j[n]$ can be considered as a discrete approximation of f after filtering by the (low-pass) filter $\phi_{0,j}(t)$ sampled at intervals 2^j.

Now, let W_j be the orthogonal complement of V_j in V_{j-1}, thus

$$V_j = V_{j+1} \oplus W_{j+1} \tag{4.20}$$

So, the orthogonal projection of any function $f \in L^2(\mathbb{R})$ onto V_j can be decomposed into two partial projection on spaces V_{j+1} and W_{j+1}, hence

$$P_{V_j} f(t) = P_{V_{j+1}} f(t) + P_{W_{j+1}} f(t). \tag{4.21}$$

Moreover, according to the properties of the MRA, we can find a index $J \in \mathbb{Z}$ small enough such that P_{V_J} approximates f up to a certain precision, therefore

$$f \approx P_{V_J} = P_{V_{J+1}} f + P_{W_{J+1}} f. \tag{4.22}$$

Mallat and Meyer [71, 72, 73] have associated each scaling function $\phi(t)$ with a wavelet $\psi(t)$, such that the set $\{\psi(t-k)\}_{k\in\mathbb{Z}}$ to form an orthonormal basis of W_0, via the following relation

$$\psi(t) = \sum_{k=-\infty}^{+\infty} (-1)^k h[1-n]\phi_{k,-1}(t), \tag{4.23}$$

where $h[n] = \langle \phi, \phi_{n,-1} \rangle$.

Similar to the case of the scaling function we define the following modulated versions of the wavelet functions:

$$\psi_{k,j}(t) = \{2^{-j/2}\psi(2^{-j}t - n)\}_{k,j\in\mathbb{Z}}. \tag{4.24}$$

Using once more the properties of the MRA in connection with the previous formula we can prove that the set $\{\psi_{k,j}(t)\}_{k\in\mathbb{Z}}$ constitutes an orthonormal base in W_j. The projection in this space is characterized by the *detail coefficients*, denoted by $\mathbf{d}_j[n] = \langle f, \psi_{n,j} \rangle$.

Therefore, we can obtain projection of f onto the space W_j analogous to 4.18, such as

$$P_{W_j} f = \sum_{k=-\infty}^{+\infty} \mathbf{d}_j[k]\psi_{k,j}. \tag{4.25}$$

We would like to mention that the scaling function $\phi(t)$ has the characteristics of a low-pass filter, whereas the wavelet function $\psi(t)$ has the characteristics of a high-pass filter.

4.2.3 Decomposition of discrete signals

Let $s[n]$ denote a discrete signal, recorded with a sampling frequency $f_s = 2^J$ for some $J \in \mathbb{Z}$. We construct a function $\mathcal{G}(t)$ with respect to the signal as follows:

$$\mathcal{G}(t) = \sum_{n=-\infty}^{+\infty} (f_s)^{-1/2} \mathbf{s}[n] \phi_{n,-J}(t) \in V_{-J} \tag{4.26}$$

Using the fact that the set $\{\phi_{n,-J}\}_{n=-\infty}^{+\infty}$ is an orthonormal basis in V_{-J}, in connection with the properties of the scaling function we can make the following approximation

$$s[n] = \mathbf{a}_{-J}[n] \approx \mathcal{G}(2^{-J}n). \tag{4.27}$$

Of course $\mathcal{G}(2^{-J}n)$ convergences to $s[n]$ as $J \to +\infty$.

— Fast Wavelet Transform (FWT) —

We adapt the following notations for the approximation and detail sub-band coefficients of the wavelet transform of a discrete signal with sampling frequency $f_s = 2^J$,

$$\mathcal{A}_\ell[n] \equiv \mathbf{a}_{-J+\ell}[n], \tag{4.28}$$

$$\mathcal{D}_\ell[n] \equiv \mathbf{d}_{-J+\ell}[n]. \tag{4.29}$$

We can get the approximation and the detail sub-band coefficients at level $\ell + 1$ by convolving the approximation sub-band coefficients at level ℓ with a pair of proper filters and then performing down-sampling. Hence:

$$\mathcal{A}_{\ell+1}[n] = (\mathcal{A}_\ell * L)[2n] \tag{4.30}$$

$$\mathcal{D}_{\ell+1}[n] = (\mathcal{A}_\ell * H)[2n], \tag{4.31}$$

where $L[n] = h[-n]$ is a low-pass filter with respect to the scaling function ϕ, whereas $H[n] = (-1)^{1-n} h[1-n]$ is a high-pass filter with respect to the wavelet function ψ.

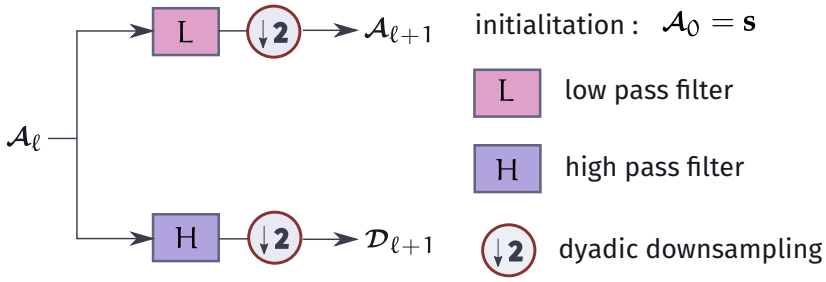


Figure 4.1: Wavelet decomposition of an approximation sub-band coefficient

4.2.4 Wavelet Packet Decomposition (WPD)

The WPD [74, 73] consists of an expansion of the MRA providing a more extensive time-frequency analysis of the signals. In the WD, at each step, the approximation sub-band coefficients split into a pair of coefficient vectors, one containing the approximation sub-band coefficients of the next level of the decomposition, and the other one the corresponding detail sub-band coefficients.

In the WPD, the detail sub-band coefficients are also decomposed into approximation and detail sub-band coefficients by applying a similar filtering procedure.

As we have presented in the previous section, an approximation space V_j belonging to a MRA is capable of being split to a lower resolution approximation space V_{j+1} and a detail space W_{j+1}.

Coefman, Meyer, and Wichershauser [75] have generalized this decomposition to any space U_j whose an orthonormal basis consists of functions translated by $n2^j$, for all $n \in \mathbb{Z}$.

Starting with the approximation space V_J, the WPD performs an iterative splitting of both the approximation and detail coefficients for a certain number of levels, forming a decomposition binary tree such as the one in Figure 4.2. Any node of this tree is labeled by a pair of indexes (j, p), where p denotes the number of the nodes that are on its left at the same decomposition level.

We assigning to each node (j, p) a space U_j^p, which has an orthonormal basis $\{\theta_j^p(t - 2^j n)\}_{n\in\mathbb{Z}}$. At the root of the tree presented in Figure 4.2, we set $U_J^0 = V_J$ and $\theta_J^0 = \phi_J$, as well. Then, for each node (j, p) the space U_j^p can be decomposed into an approximation space U_{j+1}^{2p} plus a

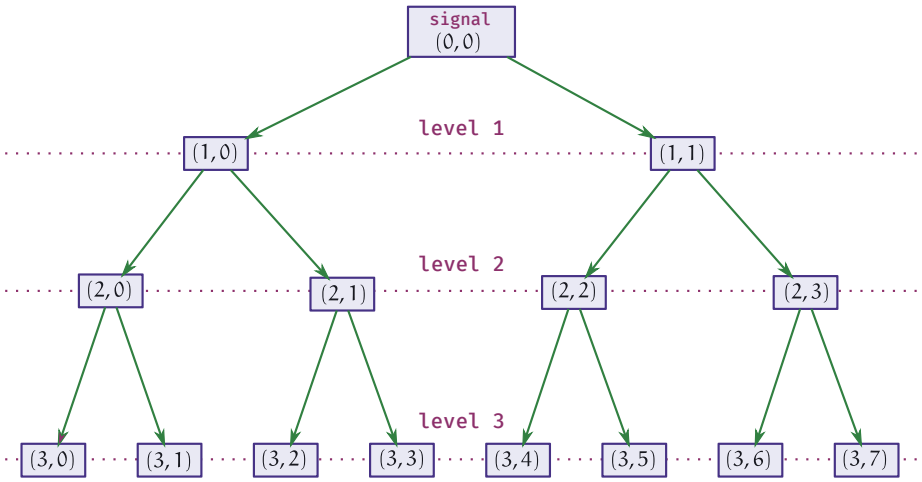


Figure 4.2: WPD decomposition tree over 3 levels.

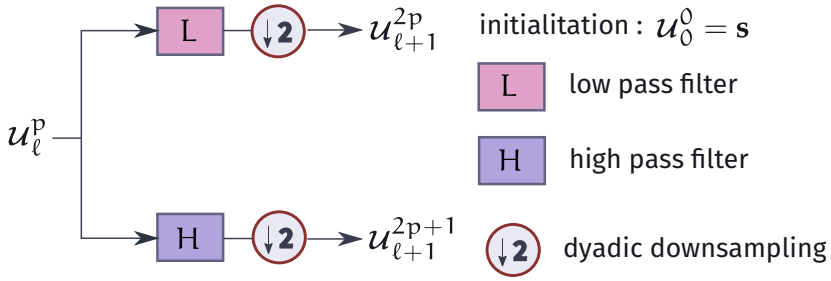


Figure 4.3: WPD of a sub-band coefficient vector

detail space U_{j+1}^{2p+1}. Furthermore, the base functions of these spaces can be proven to follow the iterative forms

$$\theta_{j+1}^{2p}(t) = \sum_{k=-\infty}^{+\infty} h[k]\theta_j^p(t - 2^j k), \tag{4.32}$$

$$\theta_{j+1}^{2p+1}(t) = \sum_{k=-\infty}^{+\infty} (-1)^k h[1-k]\theta_j^p(t - 2^j k). \tag{4.33}$$

— Fast Wavelet Packet Transform (FWPT) —

Let $u_j^p[n]$ denote the wavelet sub-band coefficients of the signal $\mathbf{s}[n]$ corresponding to the orthogonal space U_j^p. Using the base functions 4.32, 4.33 we can conclude to the following iterative relations, for the wavelet packet sub-band coefficients which are similar to those of the wavelet transform,

$$u_{\ell+1}^{2p}[n] = (u_\ell^p * L)[2n], \tag{4.34}$$

$$u_{\ell+1}^{2p+1}[n] = (u_\ell^p * H)[2n]. \tag{4.35}$$

where $u_0^0[n] = \mathbf{s}[n]$.

4.2.5 Frequency ordering of the wavelet packet coefficients

Due to some properties of the filterbank of the WPT, the resulting multi-level decomposition of a signal is not frequency ordered. To this end, a simple treatment from literature is adopted [76].

We denote by $\{\boldsymbol{\nu}_j^k\}$ a permutated version of the wavelet packet coefficient vector sequence $\{\boldsymbol{u}_j^p\}$.

The indexes k calculate through a permutation function G of the indexes p of the original coefficient vector $\boldsymbol{u}_j^p$ through a formula:,

$$k = G(p), \tag{4.36}$$

where the permutation function G associating the indexes k and p is given recursively as:

$$G(2p) = \begin{cases} 2G(p), & \text{if } G[p] \text{ is even} \\ 2G(p)+1, & \text{if } G[p] \text{ is odd} \end{cases} \tag{4.37}$$

$$G(2p+1) = \begin{cases} 2G(p)+1, & \text{if } G[p] \text{ is even} \\ 2G(p), & \text{if } G[p] \text{ is odd} \end{cases} \tag{4.38}$$

Thus, the wavelet packet coefficient vector $\boldsymbol{\nu}_j^k$ carries information for the spectrum of the signal around the central frequency of its corresponding wavelet function θ_j^k given as

$$f_c(\theta_j^k) = (k+1/2)\pi 2^{J-j-2} \quad \text{(in Hz)}. \tag{4.39}$$

The resulting WPD is then frequency ordered at each decomposition level. So, for any pair of coefficients $(\boldsymbol{\nu}_j^{k_1}, \boldsymbol{\nu}_j^{k_2})$ belonging to the same decomposition level j, we have $f_c(\theta_j^{k_1}) < f_c(\theta_j^{k+2})$.

4.2.6 Stationary Wavelet Decomposition (SWD)

The DWT and the DWPT suffer due to the absence of the time invariant property. Specifically, the DWT of a translated signal is not a similar translated version of the transform applied to the original signal.

Consider the case of the WPD of a discrete signal $\mathbf{s}[n]$ in L levels. It is clear that the same conclusions can also be made for the case of the WD due to the fact that the full decomposition tree corresponding to the WPT includes all the nodes of the WD.

Let $\mathbf{s}_{n_0}[n]$ be a translated version of the original signal $\mathbf{s}$ by n_0 samples:

$$\mathbf{s}_{n_0}[n] = \mathbf{s}[n + n_0]. \tag{4.40}$$

It is easy to observe, the following relation is true for any given discrete filter F and a positive integer n_0:

$$(\mathbf{s}_{n_0} * F)[n] = (\mathbf{s} * F)[n - n_0]. \tag{4.41}$$

Using the above relation, we can get for the first level of the decomposition of$\mathbf{s}[n]$ and $\mathbf{s}_{n_0}[n]$ the following relations:

$$\mathcal{u}_1^{2p} = (\mathcal{u}_0^0 * L)[2n] \quad , \quad \mathcal{u}_{n_0,1}^{2p} = (\mathcal{u}_0^0 * L)[2n + k], \tag{4.42}$$

$$\mathcal{u}_1^{2p+1} = (\mathcal{u}_0^0 * L)[2n] \quad , \quad \mathcal{u}_{n_0,1}^{2p+1} = (\mathcal{u}_0^0 * L)[2n + k]. \tag{4.43}$$

Using these formulas we can conclude that if $n_0 = 2k,\ k \in \mathbb{Z}$, then

$$\mathcal{u}_{n_0,1}^{2p}[n] = \mathcal{u}_1^{2p}[n - k], \tag{4.44}$$

$$\mathcal{u}_{n_0,1}^{2p+1}[n] = \mathcal{u}_1^{2p+1}[n - k]. \tag{4.45}$$

On the other hand, if $n_0 = 2k + 1\ k \in \mathbb{Z}$, the corresponding coefficients are generally unrelated. This result can be extended up to the last level of the decomposition as follows: In case that $n_0 = k2^L$, for some $k \in \mathbb{Z}$ then

$$\mathcal{u}_{n_0,L}^{2p}[n] = \mathcal{u}_L^{2p}[n - k], \tag{4.46}$$

$$\mathcal{u}_{n_0,L}^{2p+1}[n] = \mathcal{u}_L^{2p+1}[n - k], \tag{4.47}$$

otherwise, we obtain two totally unrelated decomposition trees.

As a result, to any discrete signal, corresponds to a total of 2^L unrelated decomposition trees, where $L \in \mathbb{Z}^+$ is the maximum level of the decomposition. To treat the lack of the time invariant property, the *stationary wavelet decomposition* takes account of the coefficients of all these unrelated decompositions.

— Fast Stationary Wavelet Packet Transform (FSWPT) —

Without mentioning the derivation process of the decomposition, we will refer to how it works. At the first level for the given discrete signal the approximation and the detail coefficients are evaluated by convolving the signal with the same filters as the wavelet decomposition but without down-sampling. At an arbitrary level ℓ the coefficients of the very previous

level convolve the with the up-sampled version of the previous level filters. This paragraph can be mathematically expressed as follows:

$$\mathcal{U}_{\ell+1}^{2p} = (\mathcal{U}_{\ell}^{p} * L_{\ell})[n], \tag{4.48}$$

$$\mathcal{U}_{\ell+1}^{2p+1} = (\mathcal{U}_{\ell}^{p} * H_{\ell})[n], \tag{4.49}$$

$$L_{\ell+1}[n] = \begin{cases} L_{\ell}[2^{-1}n], \text{ if } n \text{ is even,} \\ 0, \text{ otherwise} \end{cases}, \tag{4.50}$$

$$H_{\ell+1}[n] = \begin{cases} H_{\ell}[2^{-1}n], \text{ if } n \text{ is even,} \\ 0, \text{ otherwise} \end{cases}, \tag{4.51}$$

where $L_0 \equiv L$, and $H_0 \equiv H$.

4.2.7 Discussion about the wavelet decomposition of finite signals

So far, we have studied only decompositions of infinite (continuous and discrete) signals. To be capable of decomposing any finite signal, it is enough to construct an orthogonal wavelet basis of $L^2[0,1]$. To create such a bases we have to consider certain modifications of the wavelet and scaling function and the introduction of the so called *boundary wavelets*.

For a given discrete signal $\mathbf{s}[n]$ with N samples and sampling frequency $f_s = 2^J$ for a $J \in \mathbb{Z}$, we have to associate to the signal an approximation function $\mathcal{G} \in L^2[0,1]$ similar to the case of the infinity discrete signals. The scaling and wavelet function will have support in $[0,1]$ and will be constructed by periodizing the corresponding wavelet and scaling functions in $L^2(\mathbb{R})$.

Thus, the values of the corresponding coefficients except theirs boundaries values can be calculated according to the fast wavelet or fast wavelet packet transform. We shall skip the details about the calculations of the boundary coefficients, but it should be mentioned that these calculations depend on the introduction of properly modified scaling and wavelet functions [77, 78].

4.3 Concepts of Probability Theory

The purpose of this section is to present the concepts of the probability theory that we will use throughout the thesis.

In probability theory a *random variable* is a variable whose possible values quantify the possible outcomes of a random process with a sample space S. Each random variable is associated

with a *probability distribution* P which specifies the probability of any possible observation of the random variable to be realized. A probability distribution inherits its properties from the measure theory. In brief, a probability distribution associated with a sample space S must be equipped with the following properties [79]:

$$P(\emptyset) = 0 \tag{4.52}$$

$$P(S) = 1 \tag{4.53}$$

$$\text{For } \{A_1 \subset S, A_2 \subset S, \ldots\} \text{ with } A_i \cap A_j = \emptyset, i \neq j \implies P(\cup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} P(A_j), \tag{4.54}$$

where the subsets A_j called *events*.

Here, we review the basic rules of probability theory. The *joint probability distribution* of two events A and B is defined as follows:

JOINT PROBABILITY DISTRIBUTION

$$P(A, B) = P(A \cap B) = P(A|B)P(B) \tag{4.55}$$

where for $P(B) \neq 0$ the quantity $P(A|B)$ is called the *conditional probability distribution* of A given that event B is true and given by:

CONDITIONAL PROBABILITY DISTRIBUTION

$$P(A|B) = \frac{P(A, B)}{P(B)}. \tag{4.56}$$

Moreover, given the joint probability distributions $P(A, B)$ of the events A, B, we define the *marginal probability distribution* of A by:

MARGINAL PROBABILITY DISTRIBUTION

$$P(A) = \sum_B P(A, B) = \sum_B P(A|B)P(B). \tag{4.57}$$

The random variables can be classified into two main categories [80]:

DISCRETE RANDOM VARIABLES : A random variable $\mathbf{X}$ is said to be discrete if there is either a finite set of values $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ or a countable infinite set of values $\{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$ such that $P(\mathbf{X} = \mathbf{x}_j) = 1$ for some j. Furthermore, we call as the *probability mass function* of the discrete random variable $\mathbf{X}$ the following function:

PROBABILITY MASS FUNCTION

$$p(\mathbf{x}) = P(\mathbf{X} = \mathbf{x}). \tag{4.58}$$

CONTINUOUS RANDOM VARIABLES : A random variable $\mathbf{X}$ is said to be continuous if it is associated with a probability distribution P for which $P(\mathbf{X} \leq \mathbf{x})$ is differentiable (or continuous, and differentiable except finite many points). For a continuous random variable $\mathbf{X}$ with a probability distribution P, the *probability density function* of $\mathbf{X}$ is given as:

PROBABILITY DENSITY FUNCTION

$$p(\mathbf{x}) = \frac{d}{d\mathbf{x}} P(\mathbf{X} \leq \mathbf{x}). \tag{4.59}$$

For reasons of simplicity, we can refer to both probability mass and probability density functions simply as *probability* of the random variable. Also, we will adopt the terminology used by Bishop [81] who denote by p both the probability distribution and the probability density (or mass) function. Furthermore, we will denote by a lower-case letter a single random variable and by a upper-case letter a set of random variables.

4.3.1 Maximum Likelihood Estimation (MLE)

This thesis is restricted to cases of the so called *parametric distributions* which are governed by a small number of adaptive parameters, such as the mean value and variance in the one dimensional Gaussian distribution (Section 4.3.3).

Let $\mathbf{x}$ a discrete or a continuous random variable and $p(\mathbf{x}|\boldsymbol{\lambda})$ its corresponding probability mass or density function, where $\boldsymbol{\lambda}$ is the vector with the adaptive parameters which control the distribution of $\mathbf{x}$. Let also consider a set of N observations $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ of $\mathbf{x}$. We need a procedure for determining the most suitable values for the parameters given the observation $\mathbf{X}$. For this end, we employ the MLE as follows:

(MLE)
$$\boldsymbol{\lambda}^{\text{MLE}} = \arg\max_{\boldsymbol{\lambda}} \mathcal{L}(\boldsymbol{\lambda}; \mathbf{X}), \tag{4.60}$$

where $\mathcal{L}$ is the *likelihood function* which is given by:

LIKELIHOOD FUNCTION
$$\mathcal{L}(\boldsymbol{\lambda}; \mathbf{X}) = \prod_{n=1}^{N} p(\mathbf{x}_n|\boldsymbol{\lambda}) \tag{4.61}$$

4.3.2 Multinomial Distribution

A multinomial random variable is a form of a discrete random variable that can take one of a total of $K \in \mathbb{Z}^+$ different values also called *states*. In this work, we adopt the 1-of-K formulation, where the variable is represented with a K-dimensional binary vector $\mathbf{x}$ in which one and only one of the elements x_k is equal to one.

$$\mathbf{x} = [x_1, x_2, \ldots, x_K]^T, \tag{4.62}$$

Note that for such vectors we have

$$x_k \in \{0, 1\},\ k = \{1, 2, \ldots, K\}, \tag{4.63}$$

with

$$\sum_{k=1}^{K} x_k = 1. \tag{4.64}$$

Denoting the distribution of $x_k = 1$ by the parameter π_k, the probability of $\mathbf{x}$ is then given as follows

$$p(\mathbf{x}|\boldsymbol{\pi}) = \prod_{k=1}^{K} \pi_k^{x_k}. \tag{4.65}$$

where $\boldsymbol{\pi} = [\pi_1, \ldots, \pi_K]$ with $\sum_{k=1}^{K} \pi_k = 1$.

Let us consider a set of independent observations $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$. The ML estimations are given by:

$$\pi_k^{\mathrm{ML}} = \frac{1}{N} \sum_{n=1}^{N} x_{nk}, \tag{4.66}$$

where x_{nk} denotes the k-th element of $\mathbf{x}_n$.

Such distributions will play a crucial role in the generative models that we will introduce into the next chapter.

4.3.3 Gaussian Distribution

The Gaussian is by far the most widely used distribution when we deal with continuous variables, due to the central limit theorem which states that under certain conditions the sum of a set of random variables tends to obey Gaussian distribution as the number of terms in the sum increases [82].

In the case of an one dimensional variable x, the probability density function is given by

$$p(x|\mu, \sigma^2) \equiv \mathcal{N}(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}, \tag{4.67}$$

where μ is the mean value and σ^2 is the variance.

When a d-dimensional vector $\mathbf{x}$ is under consideration, the density function takes the following form

$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \equiv \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2}(\det \boldsymbol{\Sigma})^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \tag{4.68}$$

where $\boldsymbol{\mu}$ is the d-dimensional mean value vector, and $\boldsymbol{\Sigma}$ is the $d \times d$ covariance matrix.

Given $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ be a set of independent observations of a random variable $\mathbf{x}$ that obeys a multivariate Gaussian distribution, the ML estimations are given by:

$$\boldsymbol{\mu}^{\mathrm{ML}} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_n \tag{4.69}$$

$$\boldsymbol{\Sigma}^{\mathrm{ML}} = \frac{1}{N} \sum_{n=1}^{N} (\mathbf{x}_n - \boldsymbol{\mu}^{\mathrm{ML}})(\mathbf{x}_n - \boldsymbol{\mu}^{\mathrm{ML}})^T. \tag{4.70}$$

These estimations are still valid for one dimensional data $(d = 1)$.

4.3.4 The Symmetric α-Stable Distribution (SαS)

This distribution consists of a generalization of the Gaussian distribution. A symmetric α-stable distribution is defined through its characteristic function as follows:

$$\phi(t) = \exp(i\delta t - \gamma|t|^{\alpha}), \tag{4.71}$$

The characteristic function of the random variable $\mathbf{x}$, it is actually the expectation of the e^{itx} [83]:

$$\phi(t) = \int_{-\infty}^{\infty} p(x)e^{itx}dt, \tag{4.72}$$

where $\alpha \in (0, 2]$ is the characteristic exponent, $\delta \in \mathbb{R}$ is the location parameter, and $\gamma > 0$ is the dispersion of the distribution. The characteristic exponent controls the thickness of the tails of the density function. The dispersion parameter determines the amount of spreading of the distribution around its location parameter, similar to the variance of the Gaussian distribution.

Note that there are two special cases of $S\alpha S$, for $\alpha = 1$ the characteristic function is associated with a Cauchy distribution whereas for $\alpha = 2$ a Gaussian distribution is associated.

In order to derive the parameters of the distribution that describe to a greater extend a given set of observations, we use the consistent ML method described by Nolan [83], which gives reliable estimations. These distributions have been used for extracting statistical features from the wavelet coefficients of underwater acoustic signals. An overview of this procedure will describe in the next chapter.

4.3.5 Kullback-Leibler Divergence (KLD)

In this section, we present some basic concepts of information theory, that will be used to compare probability distributions and probabilistic models as well.

We consider a continuous random variable $\mathbf{x}$ which obeys a probability distribution with a probability density function p. We also assume an approximation q of the probability density function p. We wish to measure how close to the actual one is this approximation.

— Entropy —

Following the definitions presented in [81, 84], we define the following quantities

$$H(p) = -\int p(\mathbf{x}) \ln p(\mathbf{x}) d\mathbf{x}, \tag{4.73}$$

$$H(p, q) = -\int p(\mathbf{x}) \ln q(\mathbf{x}) d\mathbf{x}. \tag{4.74}$$

The first one called the *entropy* of the probability density function p and expresses the expected amount of information that this density transmits. $H(p,q)$ is called the *cross-entropy* between the probability density functions p and q and expresses the expected amount of information needed to characterize an event if the coding system that was optimized considering the estimation q instead of p.

— Relative entropy (KLD) —

In information theory, the KLD can be interpreted as the expected amount of extra information that is needed in order to cover the handicap created by using an estimated optimized distribution q rather than the actual one p. This handicap is expressed as the difference between the two entropies defined above:

$$\begin{aligned} \mathrm{KLD}(p||q) &= H(p,q) - H(p) \\ &= -\int p(\mathbf{x}) \ln\left(\frac{q(\mathbf{x})}{p(\mathbf{x})}\right) d\mathbf{x}. \end{aligned} \tag{4.75}$$

In many cases the KLD can be given in a close mathematical form. Unfortunately this is not the rule. When there is no a such closed relation, numerical estimations of the above integration should be considered.

Although, the KLD satisfies $\mathrm{KLD}(p,q) \geq 0$ with equality if and only if, $p \equiv q$, this is not considered as distance with the mathematical sense. Indeed, it is easy to be proved that in the general case, $\mathrm{KLD}(p,q) \neq \mathrm{KLD}(q,p)$.

4.4 Probabilistic Models with Hidden Variables

This section is dedicated to probabilistic models which include hidden (unobserved) variables. These variables are considered to be responsible for producing the observed data. As we will see later on this section, the existence of the hidden variables makes the determination of the model parameters harder comparing to models which consist only of observed data. We will refer to the procedure of adapting the model parameters to fit the observed data by the term *training*.

Here, we shall study models in which the probability distributions $p(\mathbf{z})$ and $p(\mathbf{x}|\mathbf{z})$ can be described by a set of model parameters $\boldsymbol{\lambda}$, and so their joint probability distribution can be factorized as follows:

$$p(\mathbf{x},\mathbf{z}|\boldsymbol{\lambda}) = p(\mathbf{z}|\boldsymbol{\lambda})p(\mathbf{x}|\mathbf{z},\boldsymbol{\lambda}). \tag{4.76}$$

Depending on the form of the probability distributions $p(\mathbf{z}|\boldsymbol{\lambda})$ and $p(\mathbf{x}|\mathbf{z},\boldsymbol{\lambda})$, we can create a large variation of probabilistic models. One of the simplest are the Mixture Models and they will be described in Section 4.4.2.

4.4.1 The Expectation-Maximization (EM) Algorithm

The goal of the EM algorithm [85] is to estimate the maximum likelihood solution for probabilistic models that include hidden variables.

Let us denote by $\mathbf{X}$ the set $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ of N realizations of a random variable $\mathbf{x}$ and by $\mathbf{Z}$ the set $\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N\}$ which includes the corresponding realizations of a random variable $\mathbf{z}$. In this work, we assume that $\mathbf{z}$ is a discrete random variable but the same approach can also be applied when $\mathbf{z}$ is a continuous random variable by simply changing summations with integrals. Note that, in general $\mathbf{x}$ and $\mathbf{z}$ are not independent random variables.

Lets consider that the joint distribution of $\mathbf{x}, \mathbf{z}$ is controlled by a set $\boldsymbol{\lambda}$ and we have observed only the set $\mathbf{X}$ then the log-likelihood function is given by

$$\ln p(\mathbf{X}|\boldsymbol{\lambda}) = \ln\left\{\sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\lambda})\right\}. \tag{4.77}$$

We denote the sets $\mathbf{X}$ and $\mathbf{Z}$ as observed and hidden data sets, respectively. The optimal set of model parameters $\boldsymbol{\lambda}^{\text{opt}}$ could be given by maximizing the log-likelihood function. However, the sum inside the logarithm on the right side of the formula 4.77, makes in general this maximization problem intractable.

Starting with an initial guess $\boldsymbol{\lambda}^0$, EM performs a two-phase iterative procedure. Assuming that we have calculated the estimation $\boldsymbol{\lambda}^t$ of the model parameters after t iterations, then the next estimation $\boldsymbol{\lambda}^{t+1}$ will be obtained after performing the expectation step, followed by the maximization step.

— The Expectation step (E-step) —

Using the current estimation of the set of the model parameters $\boldsymbol{\lambda}^t$, we evaluate the conditional probabilities of the hidden data given the set of the observed data as follows

$$\gamma^t(\mathbf{Z}) = p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\lambda}^t) = \frac{p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\lambda}^t)}{p(\mathbf{X}|\boldsymbol{\lambda}^t)} = \frac{p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\lambda}^t)}{\sum_Z p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\lambda}^t)}. \tag{4.78}$$

These quantities called *responsibilities* of $\mathbf{Z}$. Then, we denote by $\mathcal{Q}(\boldsymbol{\lambda}, \boldsymbol{\lambda}^n)$ the *expectation of the complete-data log-likelihood* evaluated for a set of parameters $\boldsymbol{\lambda}$. This expectation is given by

$$\begin{aligned} \mathcal{Q}(\boldsymbol{\lambda}, \boldsymbol{\lambda}^t) &= \mathbb{E}_{\mathbf{Z}|\boldsymbol{\lambda}^t}\{\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\lambda})\} \\ &= \sum_{\mathbf{Z}} \gamma^t(\mathbf{Z}) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\lambda}). \end{aligned} \tag{4.79}$$

— The Maximization step (M-step) —

In the M-step, we obtain a new estimation based on the current estimation of the model parameters $\boldsymbol{\lambda}^t$, $\boldsymbol{\lambda}^{t+1}$ by maximizing the function $\mathcal{Q}$, hence

$$\boldsymbol{\lambda}^{t+1} = \arg\max_{\boldsymbol{\lambda}} \mathcal{Q}(\boldsymbol{\lambda}, \boldsymbol{\lambda}^t). \tag{4.80}$$

This maximization typically is performed analytically by calculating the derivatives of $\mathcal{Q}$ with respect to the model parameters.

— Discussion about the convergence of the EM algorithm —

In this section we will show that EM algorithm constantly increases the observed data log likelihood function until it reaches a local (or its global) maximum value.

Using the definitions of the entropy, the relative entropy and the expectation of the compete-data log-likelihood $\mathcal{Q}$, in connection with the Bayes' theorem, we can associate these quantities with the observed-data log likelihood as follows:

$$\begin{aligned} \ln p(\mathbf{X}|\lambda) &= \mathcal{Q}(\boldsymbol{\lambda}, \boldsymbol{\lambda}^t) + H[\mathbf{Z}|\mathbf{X}, \boldsymbol{\lambda}^t] \\ &+ \mathrm{KLD}\Big(p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\lambda}^t), p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\lambda})\Big). \end{aligned} \tag{4.81}$$

For the analytical derivation of this expression the interested reader can refer to [81, 84].

Knowing that $\mathrm{KLD}(\mathrm{p}, \mathrm{q}) \geq 0$ with equality if and only if $p \equiv q$ enables us to write the following properly,

$$\ln p(\mathbf{X}|\boldsymbol{\lambda}) \geq \mathcal{Q}(\boldsymbol{\lambda}, \boldsymbol{\lambda}^t) + H[\mathbf{Z}|\mathbf{X}, \boldsymbol{\lambda}^t], \tag{4.82}$$

with the equality occuring when only $\boldsymbol{\lambda} \equiv \boldsymbol{\lambda}^t$.

Using the above property and keeping in mind that $\boldsymbol{\lambda}^{t+1}$ maximizes the quantity $\mathcal{Q}(\boldsymbol{\lambda}, \boldsymbol{\lambda}^{t+1})$ with respect to $\boldsymbol{\lambda}$, we can conclude

$$\begin{aligned} \ln p(\mathbf{X}|\boldsymbol{\lambda}^t) &= \mathcal{Q}(\boldsymbol{\lambda}^t, \boldsymbol{\lambda}^t) + H[\mathbf{Z}|\mathbf{X}, \boldsymbol{\lambda}^t] \\ &\leq \mathcal{Q}(\boldsymbol{\lambda}^{t+1}, \boldsymbol{\lambda}^t) + H[\mathbf{Z}|\mathbf{X}, \boldsymbol{\lambda}^t] \\ &\leq \ln p(\mathbf{X}|\boldsymbol{\lambda}^{t+1}). \end{aligned} \tag{4.83}$$

Since this sequence $\{\ln p(\mathbf{X}|\boldsymbol{\lambda}^t)\}_{t\in\mathbb{Z}^+}$ is bounded by the quantity $\ln p(\mathbf{X}|\boldsymbol{\lambda}^{\text{opt}})$ and it is monotonically increasing, the *monotone convergence theorem* [86] implies that this sequence converges, thus

$$\ln p(\mathbf{X}|\boldsymbol{\lambda}^t) \to \ln p(\mathbf{X}|\boldsymbol{\lambda}^*), \text{ as } t \to +\infty \tag{4.84}$$

where $\boldsymbol{\lambda}^*$ can be either a local or the total maximum of $\ln p(\mathbf{X}|\boldsymbol{\lambda})$.

4.4.2 Gaussian Mixture Model (GMM)

In the general form of a mixture model [87], its joint probability distribution over both the observed and hidden variables is modelled through a pair of parameters $\boldsymbol{\pi}$ and $\boldsymbol{\phi}$, where $\boldsymbol{\pi}$ is a K-dimensional vector controlling the probability distribution of the hidden variable and $\boldsymbol{\phi}$ is a set of parameters responsible for the conditional distribution of the observed random variable given the hidden variable. Following the notations for the probabilistic models, in a Mixture Model the set of model parameters $\boldsymbol{\lambda}$, that we have previously introduced, gets the form

$$\boldsymbol{\lambda} = \{\boldsymbol{\pi}, \boldsymbol{\phi}\}. \tag{4.85}$$

The joint distribution is written as follows

$$p(\mathbf{x}, \mathbf{z}|\boldsymbol{\lambda}) = p(\mathbf{z}|\boldsymbol{\pi})p(\mathbf{x}|\mathbf{z}, \boldsymbol{\phi}). \tag{4.86}$$

More specifically, $\mathbf{z}$ consists of a multinomial random variable with K states such that $p(z_k = 1) = \pi_k$ and $\sum_{k=1}^{K} \pi_k = 1$.

Furthermore, due to the conditional independence [88] of the observed data with respect to the hidden variables, there is a partition[2] $\{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K\}$ of $\boldsymbol{\phi}$ such that

$$p(\mathbf{x}|z_k = 1, \boldsymbol{\phi}) \equiv p(\mathbf{x}|\boldsymbol{\phi}_k), \text{ for } k = 1, \ldots, K \tag{4.87}$$

Therefore,

$$p(\mathbf{x}|\mathbf{z}, \boldsymbol{\phi}) = \prod_{k=1}^{K} \Big(p(\mathbf{x}|\boldsymbol{\phi}_k) \Big)^{z_k}. \tag{4.88}$$

As a result, the joint probability distribution can be given by:

$$\begin{aligned} p(\mathbf{x}, \mathbf{z}|\boldsymbol{\lambda}) &= p(\mathbf{z}|\boldsymbol{\pi})p(\mathbf{x}|\mathbf{z}, \boldsymbol{\phi}) \\ &= \prod_{k=1}^{K} \Big(\pi_k p(\mathbf{x}|\boldsymbol{\phi}_k) \Big)^{z_k}. \end{aligned} \tag{4.89}$$

2 $\cup_{k=1}^{K} \boldsymbol{\phi}_k = \boldsymbol{\phi}$ such that $\cap_{k=1}^{K} \boldsymbol{\phi}_k = \emptyset$

The marginal probability density of the observed random variable $\mathbf{x}$ is then given by marginalizing as follows

$$\begin{aligned} p(\mathbf{x}|\boldsymbol{\lambda}) &= \sum_{\mathbf{z}} p(\mathbf{x}, \mathbf{z}|\boldsymbol{\lambda}) \\ &= \sum_{k=1}^{K} \pi_k p(\mathbf{x}|\boldsymbol{\phi}_k). \end{aligned} \tag{4.90}$$

Note that the last expression justifies the term *mixture model*. Furthermore, the weights π_k express the percentage of the contribution of each $\boldsymbol{\phi}_k$ in the mixture.

Now, let us consider the superposition of K Gaussian distributions, such as

$$p(\mathbf{x}|\boldsymbol{\lambda}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k). \tag{4.91}$$

Note that, in this case we have $\boldsymbol{\phi}_k = \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}$ for mean value vectors $\boldsymbol{\mu}_k$ and covariance matrices $\boldsymbol{\Sigma}_k$.

— The EM algorithm for the GMM —

As we have seen, the k-th state of the variable $\mathbf{z}$ is associated with the k-th Gaussian of the mixture. Specifically, assuming that $\mathbf{z}$ is in its k-th state, then the k-th distribution to be responsible for generating a realization of $\mathbf{x}$. Thus,

$$p(\mathbf{x}|z_k = 1) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k). \tag{4.92}$$

The above conditional distribution can also be written in the form

$$p(\mathbf{x}|\mathbf{z}) = \prod_{k=1}^{K} \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_k}. \tag{4.93}$$

In *E-step*, we use the currently estimated model parameters $\boldsymbol{\lambda}^t$ to calculate the responsibilities γ^t for each observation $\mathbf{x}_n$, hence

$$\begin{aligned} \gamma^t(z_{nk}) &= p(z_{nk} = 1|\mathbf{x}_n, \boldsymbol{\lambda}^t) \\ &= \frac{\pi_k^t \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k^t, \boldsymbol{\Sigma}_k^t)}{\sum_{j=1}^{K} \pi_j^t \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_j^t, \boldsymbol{\Sigma}_j^t)}. \end{aligned} \tag{4.94}$$

Then we use these responsibilities to evaluate the expectation of the complete-data log likelihood

$$\begin{aligned} \mathcal{Q}(\boldsymbol{\lambda}, \boldsymbol{\lambda}^t) &= \mathbb{E}_{\mathbf{Z}|\boldsymbol{\lambda}^t}\Big\{ \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\lambda}^t) \Big\} \\ &= \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma^t(z_{nk}) \Big\{ \ln \pi_k + \ln \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \Big\}. \end{aligned} \tag{4.95}$$

In *M-step*, we get the updated model parameters $\boldsymbol{\lambda}^{t+1}$ by solving the following maximization problem with respect to the elements of $\boldsymbol{\lambda}$.

$$\boldsymbol{\lambda}^{t+1} = \arg\max_{\boldsymbol{\lambda}} \mathcal{Q}(\boldsymbol{\lambda}, \boldsymbol{\lambda}^t), \tag{4.96}$$

Maximizations with respect to $\boldsymbol{\pi}$ and $\boldsymbol{\phi}$ are performed by using proper Langrange multipliers, and the updated model parameters are given by the following formulas

$$\pi_k^{t+1} = \frac{\sum_{n=1}^N \gamma^t(z_{nk})}{N}, \tag{4.97}$$

$$\boldsymbol{\mu}_k^{t+1} = \frac{\sum_{n=1}^N \gamma^t(z_{nk})\mathbf{x}_n}{\sum_{n=1}^N \gamma^t(z_{nk})}, \tag{4.98}$$

$$\boldsymbol{\Sigma}_k^{t+1} = \frac{\sum_{n=1}^N \gamma^t(z_{nk})(\mathbf{x}_n - \boldsymbol{\mu}_k^{t+1})(\mathbf{x}_n - \boldsymbol{\mu}_k^{t+1})^T}{\sum_{n=1}^N \gamma^t(z_{nk})}. \tag{4.99}$$

This iterative procedure continues until no significant improvement of the expectation of the complete-data log likelihood function $\mathcal{Q}$.

4.5 Hidden Markov Model (HMM)

As we have seen in the previous section, a Mixture model is a tool for associating one set of observations with a batch of statistical parameters (features). In this procedure the members of the set were considered as independent realizations of the model and therefore such a model is not suitable for revealing any sequential information of the data. A HMM similarly characterizes the observation using a mixture of a probability distribution but it also takes into account sequential aspects of the data. Thus, we can consider these models as an extension of the mixture models. The mathematical infrastructure of the HMMs developed by L. Baum et al [89, 90, 91, 92] in the late 60s.

4.5.1 Definition

Lets suppose that we have observed a time series of data $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, each of which is associated with an unobserved (hidden) multinomial random variable with K states $\{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$. Let us denote these times series by $\mathbf{X}$ and $\mathbf{Z}$, respectively.

In HMMs the hidden data are set to obey the Markov property that says that the state of the variable $\mathbf{z}_n$ depends only on the state of the previous variable $\mathbf{z}_{n-1}$, therefore

$$p(\mathbf{z}_n|\mathbf{Z} - \{\mathbf{z}_n\}) = p(\mathbf{z}_n|\mathbf{z}_{n-1}). \tag{4.100}$$

The joint probability of the model is given as follows:

$$p(\mathbf{X}, \mathbf{Z}) = p(\mathbf{z}_1)\left(\prod_{n=2}^N p(\mathbf{z}_n|\mathbf{z}_{n-1})\right)\prod_{n=1}^N p(\mathbf{x}_n|\mathbf{z}_n), \tag{4.101}$$

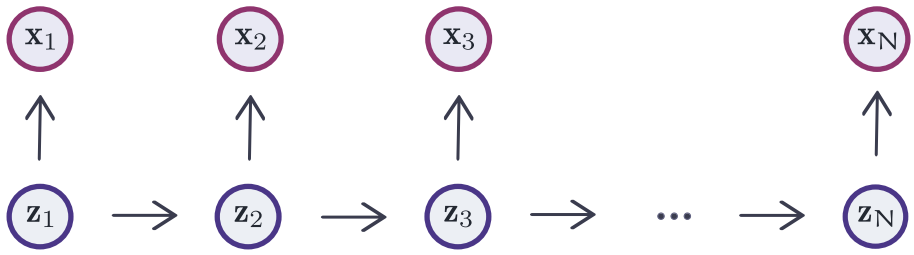


Figure 4.4: A graphical representation of a HMM. The hidden variables form a Markov Chain, and each observed variable is conditioned on the state of its corresponding hidden variable.

where the factor $p(\mathbf{z}_1)$ is called the *prior probabilities* of the first hidden variable, the factor $\prod_{n=2}^{N} p(\mathbf{z}_n|\mathbf{z}_{n-1})$ includes the transition probabilities for the hidden states, and the last factor $\prod_{n=1}^{N} p(\mathbf{x}_n|\mathbf{z}_n)$ contains the so called *emission probability distributions*, each one of which expresses the conditional probability distribution of an observed random variables $\mathbf{x}_n$, given its corresponding hidden variable $\mathbf{z}_n$.

— Prior probabilities —

Due to the hidden variable $\mathbf{z}_1$ is the first node of the chain, its marginal distribution $p(\mathbf{z}_1)$ appears directly in the joint distribution 4.101. Since this distribution is multinomial, it is represented by means of a vector $\boldsymbol{\pi}$ with elements is given by

$$\pi_k = p(z_{1k} = 1),\ k = 1, \ldots, K \tag{4.102}$$

where $\sum_{k=1}^{K} \pi_k = 1$. Therefore we have

$$p(\mathbf{z}_1) = p(\mathbf{z}_1|\boldsymbol{\pi}) = \prod_{k=1}^{K} \pi_k^{z_{1k}}. \tag{4.103}$$

The elements of the vector $\boldsymbol{\pi}$ are denoted as "prior probabilities". Some authors are referred to them as "initial probabilities".

— Transition probabilities —

If we assume that the conditional distributions $p(\mathbf{z}_n|\mathbf{z}_{n-1})$ are independent of the choice of n (time invariant) they are controlled by a $K \times K$ matrix $\mathbf{A}$ which is called the **transition matrix** of the model and its elements A_{ij} express the probability of passing from the i-th to the j-th state, hence

$$A_{ij} = p(z_{nj} = 1|z_{n-1,i} = 1),\ i, j = 1, \ldots, K. \tag{4.104}$$

As a result, all the conditional distributions share a common transition distribution defined by means of the transition matrix A. Moreover, we can express the transition probability distribution in the following explicit form:

$$p(\mathbf{z}_n|\mathbf{z}_{n-1}) = p(\mathbf{z}_n|\mathbf{z}_{n-1}, \mathbf{A}) = \prod_{i=1}^{K}\prod_{j=1}^{K} A_{ij}^{z_{n-1,i}z_{nj}}. \tag{4.105}$$

— Emission distribution —

We define the conditional distribution of any observed variable $\mathbf{x}_n$ given the value of the corresponding hidden variable $\mathbf{z}_n$.

Similar to the case of the transition probabilities were under consideration, we also assume that the emission distributions are time invariant. This assumption leads us to consider K sets of parameters $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K$, one per distinct hidden state. Each set controls the conditional distribution of the observations, so that

$$p(\mathbf{x}_n|z_{nk} = 1,\ \boldsymbol{\phi}_k),\ k = 1, \ldots, K. \tag{4.106}$$

Let us denote the hyperset of all the above sets by $\boldsymbol{\Phi}$, so $\boldsymbol{\Phi} = \{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K\}$, then we can write the emission distribution in the following form:

$$p(\mathbf{x}_n|\mathbf{z}_n) = p(\mathbf{x}_n|\mathbf{z}_n, \boldsymbol{\Phi}) = \prod_{k=1}^{K} p(\mathbf{x}_n|z_{nk} = 1, \boldsymbol{\phi}_k)^{z_{nk}}. \tag{4.107}$$

— Model parameters —

The three previous subsections introduced the set of parameters that can specify uniquely a HMM. Staying consistent with the notation introduced in Section 4.4, we introduce the set $\boldsymbol{\lambda}$ that consists of the following set of parameters:

$$\boldsymbol{\lambda} = \{\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\Phi}\}. \tag{4.108}$$

The joint probability distribution over both observed and hidden variables is given with respect to the set $\boldsymbol{\lambda}$ by

$$p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\lambda}) = p(\mathbf{z}_1|\boldsymbol{\pi})\left(\prod_{n=2}^{N} p(\mathbf{z}_n|\mathbf{z}_{n-1}, \mathbf{A})\right)\prod_{n=1}^{N} p(\mathbf{x}_n|\mathbf{z}_n, \boldsymbol{\Phi}) \tag{4.109}$$

In the sequel, we refer to a specific HMM by its set of model parameters $\boldsymbol{\lambda}$.

In this work, we are interested in the following two problems relative to the HMMs:

- **Training :** Estimation of the optimal values for the parameters $\boldsymbol{\lambda}$ of the model using one or more time series of observations.

- **Similarity measurements :** Definition of a proper similarity measure between two models comparing their model parameters $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$.

4.5.2 The EM algorithm for the HMM

Lets consider a time series of observations $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$. We wish to find the optimal model $\boldsymbol{\lambda}^{\text{opt}}$ that maximizes the posterior probability over the observed data, hence:

$$\boldsymbol{\lambda}^{\text{opt}} = \arg\max_{\boldsymbol{\lambda}}(\ln p(\mathbf{X}|\boldsymbol{\lambda})). \tag{4.110}$$

As we mentioned before, if a model including hidden variables is under concern, a common treatment is to employ the EM algorithm in order to get an estimation of its maximum likelihood solution.

The EM algorithm, as discussed earlier, is a two-phase iterative procedure in which after setting an initial guess $\boldsymbol{\lambda}^0$ for the model parameters we perform the E-step followed by the M-step producing a new improved estimation for the model parameters. We will refer by $\boldsymbol{\lambda}^t$ to the estimation of the model parameters after t iterations of the algorithm.

In our implementation of EM algorithm, we are using scaling factors in order to address the computational issues relative to low numbers due to the large number of multiplications among probabilities involved, as it has been proposed by Bishop [81].

— Initialization —

Initially, in order to set the initial guess of the model, it is common to forget any sequential connection among the observations and to handle each $\mathbf{x}_n$ independently. With this assumption we can simply set $\pi_k^0 = 1/K$ and $A_{ij}^0 = 1/K$ for all possible values of the indexes.

The initial emission parameters $\boldsymbol{\Phi}^0$ can be chosen via a hard assigned clustering procedure such as mixture models or Kmeans [93]. In this work we make use of the former procedure.

— Smoothed posterior distributions —

Here, we will provide definitions of two auxilary functions that will appear in the expression of the expected complete likelihood function of the model.

We denote by $\gamma^t(\mathbf{z}_n)$ the smoothed marginal posterior distribution for the hidden variable $\mathbf{z}_n$ and by $\xi^t(\mathbf{z}_{n-1}, \mathbf{z}_n)$ the smoothed joint posterior distribution of the two successive hidden variables $\mathbf{z}_{n-1}, \mathbf{z}_n$. Thus

$$\gamma^t(\mathbf{z}_n) = p(\mathbf{z}_n|\mathbf{X}, \boldsymbol{\lambda}^t), \tag{4.111}$$

$$\xi^t(\mathbf{z}_{n-1}, \mathbf{z}_n) = p(\mathbf{z}_{n-1}, \mathbf{z}_n|\mathbf{X}, \boldsymbol{\lambda}^t). \tag{4.112}$$

Since the variable $\mathbf{z}_n$ can take K possible values and there are K^2 different combinations of the pair $\mathbf{z}_{n-1}, \mathbf{z}_n$, it also makes sense to introduce notations for the posterior distribution for each candidate scenarios. So, we denote the following functions for $i, j, k = 1, \ldots, K$.

$$\gamma^t(z_{nk}) = p(z_{nk} = 1|\mathbf{X}, \boldsymbol{\lambda}^t), \tag{4.113}$$

$$\xi^t(z_{n-1,i}, z_{nj}) = p(z_{n-1,i} = 1, z_{nj} = 1|\mathbf{X}, \boldsymbol{\lambda}^t). \tag{4.114}$$

— Expected complete data log-likelihood function —

Let us recapitulate what the expected complete log-likelihood function is, and how it can be used to estimate the log-likelihood solution over only the observed data.

Consider that we have already performed t iterations of the EM algorithm, therefore we have acquired an estimation $\boldsymbol{\lambda}^t$ of the model parameters. Thus, (see Equation 4.79),

$$\mathcal{Q}(\boldsymbol{\lambda}, \boldsymbol{\lambda}^t) = \mathbb{E}_{\mathbf{Z}|\boldsymbol{\lambda}^t}\{\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\lambda}^t)\} = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\lambda}^t) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\lambda}) \tag{4.115}$$

$$= \sum_{k=1}^{K} \gamma^t(z_{1k}) \ln \pi_k + \sum_{n=2}^{N} \sum_{j=1}^{K} \sum_{k=1}^{K} \xi^t(z_{n-1,j}, z_{nk}) \ln A_{jk}$$

$$+ \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma^t(z_{nk}) \ln p(\mathbf{x}_n|\boldsymbol{\phi}_k). \tag{4.116}$$

In the E-step, we calculate the quantities γ^t and ξ^t for all the combinations of hidden states and then we obtain the next approximation $\boldsymbol{\lambda}^{t+1}$ by maximizing the function $\mathcal{Q}$ with respect to $\boldsymbol{\lambda}$, thus

$$\boldsymbol{\lambda}^{t+1} = \arg\max_{\boldsymbol{\lambda}} \mathcal{Q}(\boldsymbol{\lambda}, \boldsymbol{\lambda}^t). \tag{4.117}$$

— The E-step (Forward-Backward algorithm) —

Making use of the d-separation property [94, 95] applied to HMM, we can get the following conditional independencies which are being exploited for the calculation of the γ^t and ξ^t functions

$$
\begin{aligned}
p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{z}_n,\boldsymbol{\lambda}^t) &= p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n,\boldsymbol{\lambda}^t) && (4.118)\\
p(\mathbf{X}|\mathbf{z}_n,\boldsymbol{\lambda}^t) &= p(\mathbf{x}_1,\ldots,\mathbf{x}_n|\mathbf{z}_n,\boldsymbol{\lambda}^t)\\
&\quad\; p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{z}_n,\boldsymbol{\lambda}^t)\\
&= p(\mathbf{x}_1,\ldots,\mathbf{x}_n|\mathbf{z}_n,\boldsymbol{\lambda}^t)\\
&\quad\; p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n,\boldsymbol{\lambda}^t) && (4.119)\\
p(\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}|\mathbf{z}_{n-1},\mathbf{z}_n,\boldsymbol{\lambda}^t) &= p(\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}|\mathbf{z}_{n-1},\boldsymbol{\lambda}^t) && (4.120)\\
p(\mathbf{x}_n|\mathbf{z}_{n-1},\mathbf{z}_n,\boldsymbol{\lambda}^t) &= p(\mathbf{x}_n|\mathbf{z}_n,\boldsymbol{\lambda}^t) && (4.121)\\
p(\mathbf{X}|\mathbf{z}_{n-1},\mathbf{z}_n,\boldsymbol{\lambda}^t) &= p(\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}|\mathbf{z}_{n-1},\mathbf{z}_n,\boldsymbol{\lambda}^t)\\
&\quad\; p(\mathbf{x}_n|\mathbf{x}_1,\ldots,\mathbf{x}_{n-1},\mathbf{z}_{n-1},\mathbf{z}_n,\boldsymbol{\lambda}^t)\\
&\quad\; p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{z}_{n-1},\mathbf{z}_n,\boldsymbol{\lambda}^t)\\
&= p(\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}|\mathbf{z}_{n-1},\boldsymbol{\lambda}^t)\\
&\quad\; p(\mathbf{x}_n|\mathbf{x}_1,\ldots,\mathbf{x}_{n-1},\mathbf{z}_n,\boldsymbol{\lambda}^t)\\
&\quad\; p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{z}_n,\boldsymbol{\lambda}^t) && (4.122)
\end{aligned}
$$

Moreover, the posterior probability distribution over the observed data factors using Bayes' theorem is as follows

$$
p(\mathbf{X}|\boldsymbol{\lambda}^t) = p(\mathbf{x}_1,\ldots,\mathbf{x}_n|\boldsymbol{\lambda}^t)\,p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{x}_1,\ldots,\mathbf{x}_n,\boldsymbol{\lambda}^t) \tag{4.123}
$$

for an arbitrary n that belongs to the set $\{1,\ldots N\}$.

Let us define the functions α^t, β^t over the hidden variables and the constants c_n^t. We wish to clarify that the variables $\mathbf{x}_n$ are **observed** and therefore, they can be considered as constants to following relations

$$
\begin{aligned}
\alpha^t(\mathbf{z}_n) &= p(\mathbf{z}_n|\mathbf{x}_1,\ldots,\mathbf{x}_n,\boldsymbol{\lambda}^t), && (4.124)\\
\beta^t(\mathbf{z}_n) &= \frac{p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n,\boldsymbol{\lambda}^t)}{p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{x}_1,\ldots\mathbf{x}_n,\boldsymbol{\lambda}^t)}, && (4.125)\\
c_n^t &= p(\mathbf{x}_n|\mathbf{x}_1,\ldots,\mathbf{x}_{n-1},\boldsymbol{\lambda}^t). && (4.126)
\end{aligned}
$$

Using the conditional independence property 4.119 in connection with the factorization 4.123 and the definitions of $\alpha^t(\mathbf{z}_n)$, $\beta^t(\mathbf{z}_n)$, we obtain

$$\begin{aligned}
\gamma^t(\mathbf{z}_n) &= p(\mathbf{z}_n|\mathbf{X},\boldsymbol{\lambda}^t) = \frac{p(\mathbf{X}|\mathbf{z}_n,\boldsymbol{\lambda}^t)p(\mathbf{z}_n|\boldsymbol{\lambda}^t)}{p(\mathbf{X}|\boldsymbol{\lambda}^t)} \\
&= \frac{p(\mathbf{x}_1,\ldots,\mathbf{x}_n|\mathbf{z}_n,\boldsymbol{\lambda}^t)p(\mathbf{z}_n|\boldsymbol{\lambda}^t)}{p(\mathbf{x}_1,\ldots,\mathbf{x}_n|\boldsymbol{\lambda}^t)} \frac{p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n,\boldsymbol{\lambda}^t)}{p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{x}_1,\ldots,\mathbf{x}_n,\boldsymbol{\lambda}^t)} \\
&= p(\mathbf{z}_n|\mathbf{x}_1,\ldots,\mathbf{x}_n,\boldsymbol{\lambda}^t)\frac{p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n,\boldsymbol{\lambda}^t)}{p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{x}_1,\ldots,\mathbf{x}_n,\boldsymbol{\lambda}^t)} \\
&= \alpha^t(\mathbf{z}_n)\beta^t(\mathbf{z}_n) \qquad (4.127)
\end{aligned}$$

Similarly, we can derive a factorization of $\xi^t(\mathbf{z}_n,\mathbf{z}_{n-1})$ using the property 4.122 as well as the definitions 4.124-4.126

$$\begin{aligned}
\xi^t(\mathbf{z}_{n-1},\mathbf{z}_n) &= p(\mathbf{z}_{n-1},\mathbf{z}_n|\mathbf{X},\boldsymbol{\lambda}^t) = \frac{p(\mathbf{X}|\mathbf{z}_{n-1},\mathbf{z}_n,\boldsymbol{\lambda}^t)p(\mathbf{z}_{n-1},\mathbf{z}_n|\boldsymbol{\lambda}^t)}{p(\mathbf{X}|\boldsymbol{\lambda}^t)} \\
&= \frac{1}{p(\mathbf{x}_n|\mathbf{x}_1,\ldots,\mathbf{x}_{n-1},\boldsymbol{\lambda}^t)} \frac{p(\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}|\mathbf{z}_{n-1},\boldsymbol{\lambda}^t)p(\mathbf{z}_{n-1}|\boldsymbol{\lambda}^t)}{p(\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}|\boldsymbol{\lambda}^t)} \\
&\quad \frac{p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n,\boldsymbol{\lambda}^t)}{p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{x}_1,\ldots,\mathbf{x}_n,\boldsymbol{\lambda}^t)} p(\mathbf{z}_n|\mathbf{z}_{n-1},\boldsymbol{\lambda}^t)p(\mathbf{x}_n|\mathbf{z}_n,\boldsymbol{\lambda}^t) \\
&= (c_n^t)^{-1}\alpha^t(\mathbf{z}_{n-1})\beta^t(\mathbf{z}_n)p(\mathbf{z}_n|\mathbf{z}_{n-1},\boldsymbol{\lambda}^t)p(\mathbf{x}_n|\mathbf{z}_n,\boldsymbol{\lambda}^t). \qquad (4.128)
\end{aligned}$$

Recall that the goal of the E-step is to evaluate the $\gamma^t(\mathbf{z}_n)$ and $\xi^t(\mathbf{z}_{n-1},\mathbf{z}_n)$ to calculate the quantity $\mathcal{Q}(\boldsymbol{\lambda},\boldsymbol{\lambda}^t)$ for each set of model parameter $\boldsymbol{\lambda}$. The relations 4.127, 4.128 indicate that these evaluations could be given in terms of the functions $\alpha^t(\mathbf{z}_n)$, $\beta^t(\mathbf{z}_n)$ and the constants c_n^t. As we will see, these quantities can be calculated through two recursive relations.

— Forward process ($n = 1,\ldots,N-1$) —

From the d-separation property, the following conditional independencies are obtained

$$p(\mathbf{x}_1,\ldots,\mathbf{x}_n|\mathbf{z}_{n+1},\mathbf{x}_{n+1},\boldsymbol{\lambda}^t) = p(\mathbf{x}_1,\ldots,\mathbf{x}_n|\mathbf{z}_{n+1},\boldsymbol{\lambda}^t) \qquad (4.129)$$

$$p(\mathbf{z}_{n+1}|\mathbf{z}_n,\mathbf{x}_1,\ldots,\mathbf{x}_n,\boldsymbol{\lambda}^t) = p(\mathbf{z}_{n+1}|\mathbf{z}_n,\boldsymbol{\lambda}^t). \qquad (4.130)$$

Using these two relations and rules of the probability theory, we are able to express the product $c_{n+1}^t\alpha^t(\mathbf{z}_{n+1})$ in terms of the $\alpha^t(\mathbf{z}_n)$ as follows:

$$c_{n+1}^t\alpha^t(\mathbf{z}_{n+1}) = p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1},\boldsymbol{\lambda}^t)\sum_{\mathbf{z}_n}\alpha^t(\mathbf{z}_n)p(\mathbf{z}_{n+1}|\mathbf{z}_n,\boldsymbol{\lambda}^t). \qquad (4.131)$$

Now, let's consider that we have already calculated the values $\alpha^t(\mathbf{z}_n)$ for all the possible values of $\mathbf{z}_n$ as well as the coefficient c_n. We can use the recursive relation 4.131 to calculate at first the product $c_{n+1}^t\alpha^t(\mathbf{z}_{n+1})$ and then due to the $\sum_{\mathbf{z}_{n+1}}\alpha^t(\mathbf{z}_{n+1}) = 1$ we can separately obtain the coefficient c_{n+1}^t and the K values of $\alpha^t(\mathbf{z}_{n+1})$.

— Backward process $(n = N-1, \ldots, 1)$ —

We can also derive the next conditional independent properties for the model

$$p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N | \mathbf{z}_n, \mathbf{z}_{n+1}, \boldsymbol{\lambda}^t) = p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N | \mathbf{z}_{n+1}, \boldsymbol{\lambda}^t) \tag{4.132}$$

$$p(\mathbf{x}_{n+2}, \ldots, \mathbf{x}_N | \mathbf{z}_{n+1}, \mathbf{x}_{n+1}, \boldsymbol{\lambda}^t) = p(\mathbf{x}_{n+2}, \ldots, \mathbf{x}_N | \mathbf{z}_{n+1}, \boldsymbol{\lambda}^t). \tag{4.133}$$

and then to write the $\beta^t(\mathbf{z}_n)$ by means of $\beta^t(\mathbf{z}_{n+1})$ as follows

$$\beta^t(\mathbf{z}_n) = (c_{n+1}^t)^{-1} \sum_{\mathbf{z}_{n+1}} \beta^t(\mathbf{z}_{n+1}) p(\mathbf{x}_{n+1} | \mathbf{z}_{n+1}, \boldsymbol{\lambda}^t) p(\mathbf{z}_{n+1} | \mathbf{z}_n, \boldsymbol{\lambda}^t). \tag{4.134}$$

When the forward process is over, we have obtained the values of each coefficient c_n^t. By assuming the values $\beta(\mathbf{z}_{n+1})$ known, we are able to evaluate the corresponding values $\beta(\mathbf{z}_n)$ by applying the formula 4.134.

— Initialization —

It is simple to find the initial values c_1^t and $\alpha^t(\mathbf{z}_1)$. Indeed, we have

$$c_1^t = p(\mathbf{x}_1 | \boldsymbol{\lambda}^t), \tag{4.135}$$

$$\begin{aligned} \alpha^t(\mathbf{z}_1) &= p(\mathbf{z}_1 | \mathbf{x}_1) = (c_1^t)^{-1} c_1^t p(\mathbf{z}_1 | \mathbf{x}_1) \\ &= (c_1^t)^{-1} p(\mathbf{z}_1) p(\mathbf{x}_1 | \mathbf{z}_1). \end{aligned} \tag{4.136}$$

Therefore,

$$c_1^t = \sum_{k=1}^{K} \{\pi_k p(\mathbf{x}_1 | \boldsymbol{\lambda}_k)\}^{z_{1k}}, \tag{4.137}$$

$$\alpha^t(\mathbf{z}_1) = (c_1^t)^{-1} \prod_{k=1}^{K} \{\pi_k p(\mathbf{x}_1 | \boldsymbol{\lambda}_k)\}^{z_{1k}}. \tag{4.138}$$

In case of $\beta^t(\mathbf{z}_N)$, we first have to notice that for $n = N$, we have $\gamma^t(\mathbf{z}_N) = \alpha^t(\mathbf{z}_N)$. So, in combination with 4.125 we obtain:

$$\beta^t(\mathbf{z}_N) = \frac{\gamma^t(\mathbf{z}_N)}{\alpha^t(\mathbf{z}_N)} = 1. \tag{4.139}$$

— The M-step —

The objective of the M-step is to maximize $\mathcal{Q}(\boldsymbol{\lambda}, \boldsymbol{\lambda}^t)$ with respect to the model parameter set $\boldsymbol{\lambda}$.

Maximizations with respect to $\boldsymbol{\pi}$ and $\mathbf{A}$ are performed by using proper Langrange multipliers, and the solutions are given independently of the type of the emission distributions as follows

$$\pi_k^{t+1} = \frac{\gamma^t(z_{1k})}{\sum_{j=1}^K \gamma^t(z_{1j})}, \tag{4.140}$$

$$A_{jk}^{t+1} = \frac{\sum_{n=2}^N \xi^t(z_{n-1,j}, z_{nk})}{\sum_{\ell=1}^K \sum_{n=2}^N \xi^t(z_{n-1,j}, z_{n\ell})}. \tag{4.141}$$

By considering Gaussian emission densities, such that $\boldsymbol{\phi}_k = \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}$ and $p(\mathbf{x}|\boldsymbol{\phi}_k) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, the maximization of $\mathcal{Q}$ with respect each pair $\{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}$, $k = 1, \ldots, K$ at a time, leads to

$$\boldsymbol{\mu}_k^{t+1} = \frac{\sum_{n=1}^N \gamma^t(z_{nk})\mathbf{x}_n}{\sum_{n=1}^N \gamma^t(z_{nk})}, \tag{4.142}$$

$$\boldsymbol{\Sigma}_k^{t+1} = \frac{\sum_{n=1}^N \gamma^t(z_{nk})(\mathbf{x}_n - \boldsymbol{\mu}_k^{t+1})(\mathbf{x}_n - \boldsymbol{\mu}_k^{t+1})^T}{\sum_{n=1}^N \gamma^t(z_{nk})}. \tag{4.143}$$

— Inference —

Here, we infer the posterior probability distribution over the observed data $\mathbf{X}$ given the model parameters $\boldsymbol{\lambda}^t$. Since, posterior probability can be obtained by the product of c_n^t, such an inference demands just performing the forward procedure. Indeed,

$$\begin{aligned} p(\mathbf{X}|\boldsymbol{\lambda}^t) &= p(\mathbf{x}_1, \ldots, \mathbf{x}_N|\boldsymbol{\lambda}^t) \\ &= \prod_{n=1}^N p(\mathbf{x}_n|\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}, \boldsymbol{\lambda}^t) \\ &= \prod_{n=1}^N c_n^t. \end{aligned} \tag{4.144}$$

and therefore, for the log-scaled posterior probability distribution, we obtain:

$$\ln p(\mathbf{X}|\boldsymbol{\lambda}^t) = \sum_{n=1}^N \ln c_n^t. \tag{4.145}$$

— Convergence criterion —

If we assume again that we have the model $\boldsymbol{\lambda}^t$ in the t-th iteration of the EM algorithm, we can get the improved model $\boldsymbol{\lambda}^{t+1}$ after performing one more iteration. Lets denote by $\mathcal{E}^t$ the difference of the posterior probabilities with respect to the successive models $\boldsymbol{\lambda}^t$ and $\boldsymbol{\lambda}^{t+1}$, hence:

$$\mathcal{E}^t = \ln p(\mathbf{X}|\boldsymbol{\lambda}^{t+1}) - \ln p(\mathbf{X}|\boldsymbol{\lambda}^t) \geq 0. \tag{4.146}$$

According to 4.84, for an arbitrary small $\epsilon > 0$, we can find $t_\epsilon \in \mathbb{Z}^+$ such that

$$\mathcal{E}^{t_\epsilon} < \epsilon \tag{4.147}$$

— Feature extraction (Encoding) —

To summarize the training process, by choosing $\epsilon > 0$ and the number of possible hidden states K and having a time series of observations $\mathbf{X}$, we find a initial guess $\boldsymbol{\lambda}^0$ of the optimal HMM that describes these observations using a Mixture model. Then we obtain a sequence of models $\boldsymbol{\lambda}^t$ using iterations of the EM algorithm. We stop obtaining updated models once we have achieved the convergence criterion in t_ϵ-th iteration. We apparently associate the observations $\mathbf{X}$ with the final model $\boldsymbol{\lambda}^{t_\epsilon}$. Therefore,

FEATURE EXTRACTION

$$\mathbf{X} \to \boldsymbol{\lambda}^{t_\epsilon} \tag{4.148}$$

— Viterbi algorithm (Decoding) —

The Viterbi algorithm [96, 97] is used to compute the most probable sequence of the hidden states given the observed data by solving the following maximization problem

$$\mathbf{Z}^* = \arg\max_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\lambda}). \tag{4.149}$$

Using the Bayes' rule in the logarithm scale, the above maximization is equivalent to the maximization with respect to hidden data of the probability density function over both hidden and observed data $(\mathbf{X}, \mathbf{Z})$, thus

$$\mathbf{Z}^* = \arg\max_{\mathbf{Z}} \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\lambda}). \tag{4.150}$$

Let us introduce the following quantities $\omega(\mathbf{z}_n) \in \mathbb{R}^K,\ n = 1, \ldots, N$ as

$$\omega(\mathbf{z}_n) = \max_{\mathbf{z}_1, \ldots, \mathbf{z}_{n-1}} p(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{z}_1, \ldots, \mathbf{z}_n|\boldsymbol{\lambda}), \tag{4.151}$$

where K is again the number of hidden states.

Making use of the sum-product algorithm [98], we obtain a recursive expression for the calculation of the $\omega(\mathbf{z}_n),\ n = 1, \ldots, N$, as follows:

$$\omega(\mathbf{z}_n) = \ln p(\mathbf{x}_n|\mathbf{z}_n, \boldsymbol{\lambda}) + \max_{\mathbf{z}_{n-1}} \Big\{ \ln p(\mathbf{x}_{n-1}|\mathbf{z}_{n-1}, \boldsymbol{\lambda}) + \omega(\mathbf{z}_{n-1}) \Big\},\ n = 2, \ldots, N, \tag{4.152}$$

in connection with the initialization for the first hidden state $\mathbf{z}_1$

$$\omega(\mathbf{z}_1) = \ln p(\mathbf{z}_1|\boldsymbol{\lambda}) + \ln p(\mathbf{x}_1|\mathbf{z}_1|\boldsymbol{\lambda}) \tag{4.153}$$

4.5.3 Similarity measurements

So far in this section, we have described a procedure of associating a series of observations $\mathbf{X}$ with a set of features $\boldsymbol{\lambda}$, which defines a HMM.

Consider two set of observations $\mathbf{X}_1, \mathbf{X}_2$ and lets denote by $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2$ their corresponding HMMs. We wish to measure how similar these series are by means of their corresponding models. For this end, we consider the use of the KLD [66] that consists a suitable similarity measure among probabilistic models. So, we define the following similarity measure $d(\mathbf{X}_1, \mathbf{X}_2)$ of the series of observations as

$$d(\mathbf{X}_1, \mathbf{X}_2) = \mathrm{KLD}(\boldsymbol{\lambda}_1 \| \boldsymbol{\lambda}_2). \tag{4.154}$$

We recall that the KLD is given by considering the following integration over the observed data

$$\mathrm{KLD}(\boldsymbol{\lambda_1} \| \boldsymbol{\lambda_2}) = -\int_{\mathbf{X}} p(\mathbf{X}|\boldsymbol{\lambda_1}) \ln \frac{p(\mathbf{X}|\boldsymbol{\lambda_2})}{p(\mathbf{X}|\boldsymbol{\lambda_1})} d\mathbf{X}. \tag{4.155}$$

Unfortunately, there is no a closed form estimation for this integral, therefore a numerical approximation technique has to be employed. In the next subsection, we will present a numerical approximation of the KLD.

— Numerical approximation of the KLD —

A set of $\mathcal{I}$ time series of observations $\{\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(\mathcal{I})}\}$ with N' samples each, is drawn from the probability distribution $p(\mathbf{X}|\boldsymbol{\lambda}_1)$ using the so called importance sampling technique presenting in refs [99, 100, 101].

Then, we consider a *Monte-Carlo* numerical approximation over these generated sequences as

$$\mathrm{KLD}(\boldsymbol{\lambda}_1 \| \boldsymbol{\lambda}_2) \approx \frac{1}{\mathcal{I}N'} \sum_{i=1}^{\mathcal{I}} [\ln p(\mathbf{X}^{(i)}|\boldsymbol{\lambda}_1) - \ln p(\mathbf{X}^{(i)}|\boldsymbol{\lambda}_2)] \tag{4.156}$$

Of course, the larger the numbers $\mathcal{I}, N'$ are, the better are the approximations. On the other hand as the more accuracy required, the calculations become more and more demanding. In view of this, a trade off between the computational cost and the precision of the estimation has to be considered in real world applications.

4.6 Sparse Dictionary Decomposition

Consider a set of observations $\mathbf{S} = \{\mathbf{s}_j \in \mathbb{R}^L\}_{j=1}^M$. For the problems treated in this thesis these observations consist of the samples of a signal belonging to certain windows. Our goal is to

find a, so called "dictionary", matrix $\mathbf{D} \in \mathbb{R}^{L,M}$ so that each vector $\mathbf{s}_j$ to be estimated in a proper form as $\mathbf{s}_j^*$ given by

$$\mathbf{s}_j^* = \mathbf{D}\mathbf{a}_j \tag{4.157}$$

where the coefficients $\mathbf{a}_j$ are appropriate vectors $\{\mathbf{a}_j \in \mathbb{R}^M\}_{j=1}^M$ to be also defined. If we are able to define the coefficients vectors $\mathbf{a}_j$ to have many zeros, the $\mathbf{s}_j^*$ will be assumed as a sparse approximations of the $\mathbf{s}_j$.

We wish to define dictionary and coefficient vectors so that the approximation error $\frac{1}{M}\sum_{j=1}^N \|\mathbf{s}_j - \mathbf{s}_j^*\|_2^2$ is small, subject to a penalty regularization function that keeps the representation as sparse as possible.

We can now introduce the general form of the optimization problem that satisfies the sparse model prerequisites. We have considered as the total objective function a combination of the approximation error and sparsity term in order to get an effective projection of the signal frames from the noisy space onto a low-dimensional space with low noise component. Following Marial et al. [102] our optimization problem takes the form

$$\min_{\mathbf{D}\in\mathbb{R}^{L,M}} \frac{1}{M}\sum_{j=1}^{M} \min_{\mathbf{a}_j\in\mathbb{R}^M} \Big(g(\mathbf{a}_j; \mathbf{D}, \mathbf{s}_j) + h(\mathbf{a}_j) \Big), \tag{4.158}$$

where $g(\mathbf{a}_j; \mathbf{D}, \mathbf{s}_j)$ is the convex quadratic function representing the approximation error term of a single signal window frame $\mathbf{s}_j$, given by

$$g(\mathbf{a}_j; \mathbf{D}, \mathbf{s}_j) = \frac{1}{2}\|\mathbf{s}_j - \mathbf{D}\mathbf{a}_j\|_2^2, \tag{4.159}$$

and $h(\mathbf{a}_j)$ is a proper penalty regularization function which controls the sparsity character of the representation. We set

$$h(\mathbf{a}_j) = h(\mathbf{a}_j; \lambda) = \lambda\|\mathbf{a}_j\|_p, \tag{4.160}$$

where λ is a "sparsity coefficient" and $\|\mathbf{a}_j\|_p$ is a p-th order norm of $\mathbf{a}_j$.

A rational choice for the order of the norm of $\mathbf{a}_j$ would be zero in which case the zero order norm (actually pseudonorm) is the cardinal number (number of non-zero elements) of each coefficient vector $\mathbf{a}_j$:

$$\|\mathbf{a}_j\|_0 = \#\{k = 1, \ldots, M \mid \mathbf{a}_j[k] \neq 0\}, \tag{4.161}$$

where symbol # stands for the cardinal number of the set. λ should be a small number in order that the term $h(\mathbf{a}_j)$ is kept small with respect to $g(\mathbf{a}_j; \mathbf{D}, \mathbf{s}_j)$ for reasons to be explained later.

Minimization including the ℓ_0-pseudonorm is not computationally tractable due to the non-convexity[3] of the $\|\cdot\|_0$. For this reason, we have chosen to exploit the best available convex approximation of the above pseudonorm, which has been geometrically explained by Ramirez et al. [103] to be the ℓ_1 norm. Thus,

$$h(\mathbf{a}_j;\lambda) = \lambda\|\mathbf{a}_j\|_1, \tag{4.162}$$

with λ to be manually adjusted with trial-and-error strategy.

Once the minimization problem has been defined we can write down the two step adaptation scheme of the dictionary $\mathbf{D}$ and coefficients $\mathbf{a}_j$ which is performed at each training stage:

- For a fixed dictionary matrix D, adapt each coefficient vector $\mathbf{a}_j$ to $\hat{\mathbf{a}}_j$ so that minimum error is achieved.
- Keep $\hat{\mathbf{a}}_j$ fixed and focus on seeking a new dictionary D for which the total error is minimized.

4.6.1 The Iterative Shrinkage and Thresholding Algorithm (ISTA)

Assuming that the dictionary matrix $\mathbf{D} \in \mathbb{R}^{L,M}$ is given, we define $\ell(\mathbf{a}_j;\mathbf{D},\mathbf{s}_j,\lambda)$ as follows:

$$\ell(\mathbf{a}_j;\mathbf{D},\mathbf{s}_j,\lambda) = g(\mathbf{a}_j;\mathbf{D},\mathbf{s}_j) + h(\mathbf{a}_j;\lambda). \tag{4.163}$$

We estimate the coefficient vectors $\mathbf{a}_j$, solving the following minimization problem

$$\hat{\mathbf{a}}_j = \underset{\mathbf{a}_j\in\mathbb{R}^M}{\arg\min}\, \ell(\mathbf{a}_j;\mathbf{D},\mathbf{s}_j,\lambda). \tag{4.164}$$

One critical issue is the fact that the gradient of the function $\ell(\mathbf{a}_j;\mathbf{D},\mathbf{s}_j,\lambda)$ is not differentiable at zero. So, the use of a classical gradient descent method is not applicable in this case.

To solve the optimization problem (4.164) we introduce an iterative procedure: For a given dictionary matrix $\mathbf{D}$ we estimate the coefficient vectors $\hat{\mathbf{a}}_j$ in order to minimize the objective function (4.164). This procedure is realized by means of an implementation of the ISTA similar to that presented by Marial et al. [102].

4.6.2 Dictionary Learning Algorithm

We revisit the original problem as stated in (4.158). Having already adapted the coefficient vectors, our present goal is to choose the best possible dictionary $\mathbf{D}$ for the most appropriate

3 f is a convex function in $[a,b]$ iff $f(\lambda a + (1-\lambda)b) \leq \lambda f(a) + (1-\lambda)f(b)$, for all $\lambda \in [0,1]$.

description of the signal samples. To this end, we shall employ the algorithm of the dictionary learning with mini-batch extension proposed by Marial et al. [102] which has been proved by the same author to converge. According to this approach the best dictionary $\mathbf{D}$ is determined by means of the following formula :

$$\min_{\mathbf{D}} \frac{1}{M} \sum_{j=1}^{M} \min_{\mathbf{a}_j} \ell(\mathbf{a}_j; \mathbf{D}, \mathbf{s}_j). \tag{4.165}$$

Let us assume that $\mathbf{a}_j$ does not depend on the dictionary choice. This is generally not true but if we consider that $\mathbf{D}$ changes slowly enough we can use this simplification without further consideration. So, the minimization problem is written in the form

$$\min_{\mathbf{D}} \frac{1}{M} \sum_{j=1}^{M} \frac{1}{2} \|\mathbf{s}_j - \mathbf{D}\hat{\mathbf{a}}_j\|_2^2. \tag{4.166}$$

Note that only the approximation error contributes to the dictionary learning procedure.

It is crucial that we choose matrix $\mathbf{D}$ to have all columns with unit Euclidean norm. Note that the columns of $\mathbf{D}$ need not be orthogonal (linear independence is not required).

This procedure is repeated for large number of steps. The pseudocode for the implemented algorithm is described in Algorithm 2 in Appendix B.

4.7 The Genetic Algorithm (GA)

Genetic Algorithms (GAs) are optimization algorithms based on the mechanics of natural selection and evolution. They deal with problems of minimization (or maximization) of an objective function on the form:

$$\mathbf{m}^{\text{opt}} = \arg\min_{\mathbf{m}} f(\mathbf{m}) \tag{4.167}$$

where $\mathbf{m} = [m_1, \ldots, m_M]^T$, the vector of unknown parameters to be estimated. These algorithms are fundamentally different from traditional optimization procedures in the following aspects

- They deal with a coded version of the unknown variables, usually in the binary system.
- They need no information about derivatives with respect to the variables.
- The provide a number of candidate estimations (individuals) instead of a single one.
- They employ stochastic operators rather than deterministic ones.

Note that, there are many variations of GAs. Here, we will restrict the description only to the implementation that we have used and it is based on approaches presented in [104, 105].

4.7.1 Encoding

Initially, we have to define a search space for each recoverable variable m_j, thus $m_j \in [a_j, b_j] \subset \mathbb{R}$. Bellow, we define the functions h_j, each one mapping an binary variable $\mathbf{z}_j$ with d_j digits to a value belonging to the search space of the m_j

$$h_j(\mathbf{z}_j) = a_j + \frac{\langle \mathbf{z}_j, \mathbf{u} \rangle}{2^{d_j} - 1}(b_j - a_j) \in [a_j, b_j]. \tag{4.168}$$

where $\mathbf{u} = [2^{d_j-1}, 2^{d_j-2}, \ldots, 2^0]^T$. Note that the inner product simply maps the binary variable $\mathbf{z}_j$ to its decimal arithmetic form.

We define the set $\mathbf{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_M\}$ that consists of the binary variables for each recoverable variable, then instead of dealing with the initial problem 4.167, we seek an estimation of a discrete optimization problem that is defined by means of $\mathbf{Z}$:

$$\mathbf{Z}^{\text{opt}} = \arg\min_{\mathbf{Z}} (f \circ \mathbf{h})(\mathbf{Z}) \tag{4.169}$$

where $\mathbf{h}(\mathbf{Z}) = \Big[h_1(\mathbf{z}_1), \ldots, h_M(\mathbf{z}_M)\Big]^T$.

The number of the digits d_j controls the density of the grid that forms the candidate values that can take the corresponding parameter $\mathbf{m}_j$.

4.7.2 Terminology

A GA maintains a fixed number of estimations that evolve through an iteration procedure that includes several evolutionary operators. Before presenting how our implementation works, we would like to make readers familiar with some terminology.

POPULATION : A set of encoded estimations POP $= \{\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(Q)}\}$. We will also refer with the same term to the corresponding set of decoded estimations $\{\mathbf{m}^{(1)}, \ldots, \mathbf{m}^{(Q)}\}$. Note that, the population in a GA has an analogous meaning as to the population for animals or human beings.

INDIVIDUAL : Is an element of the population, that is an encoded candidate solution described by the set $\mathbf{Z}$.

ATTRIBUTE : With this term we refer to each encoded parameter $\mathbf{z}_j$ in an individual.

GENE : A gene is the value of any element $z_{jk} \in \{0, 1\}$ of an attribute $\mathbf{z}_j$.

FITNESS RATE : It defined by the value of the objective function which takes each individual as input and produces the suitability of the individual as the output. In the case of minimization problems, smaller values of the fitness rate means more fitter individual.

4.7.3 Evolutionary Operators

The population in a GA involves with the following genetic operators

REPRODUCTION : A selection of individuals is performed based on their fitness. In particular, GA select a so called mating pool a set consisting of N_{mp} individuals so that $N_{\text{mp}} = N_{\text{pop}} - N_{\text{hof}}$, where N_{hof} individuals are to be cloned via the Hall of Fame operator to be described bellow. Note that, this is a stochastic procedure in which the fittest individuals are favoured but not always chosen.

CROSSOVER : Each member of the mating pool has p_c probability to produce descendants, mating with another one. If the algorithm selects odd number of individuals, we pull out one at random. The genes of a pair of individuals are combined and create two descendants, that replace their parents.

MUTATION : Each of the genes of an individual can be mutated with a so called mutation probability. Mutation is used to avoid being trapped in local minima.

ELITISM MECHANISM : It keeps track of a group of the elite individuals which illustrate the best fitness rate and at the same time innovative characteristics.

— Reproduction —

TOURNAMENT SELECTION : Randomly selects k individuals (tournament size) from the population and then selects the best out of these to survive.

RANK BASED ROULETTE WHEEL : Rank selection first ranks the population and then every individual is associated with a piece of the roulette with area proportional to its fitness rate. For example, the worst individual could have area 1, second worst individual 2 and the best will have fitness Q (size of the population). After this all the individuals have a chance to be selected. Rank-based selection schemes can avoid premature convergence but it is computationally expensive because of the sorting of the populations based on the individuals' fitness rate.

— Crossover —

GA replaces the parents of the mating pool created by the reproduction operator by the produced children directly in the population

— Mutation —

Each individual in the mating pool has p_{mut} probability to become a mutant. We will refer to this probability as the mutant probability. Once an individual has been marked as a mutant, then each of its attributes is to be flipped (from 0 to 1 and vice versa) with a quite small mutation probability p_m, typically between 0.01 and 0.05. Larger values of this probability could turn the optimization procedure into a random search.

— Hall of Fame (HOF) —

Here, we introduce a new form of an elitism mechanism based on the one implemented in [105] that gets a set of N_{hof} of informative individuals evolved from generation to generation, the elements of which after possible modifications, are survived in each new population throughout the optimization process. This elite population is divided into two subsets. The first subset which we will be referred as the *innovative population*, composed by N_{in} individuals and takes into account, in addition to the fitness rate of the individuals, the diversity of the members.

The insertion of a new individual in this subset is made so that old members of that set have priority over new individuals. The individuals are sorted based on their fitness values. A new candidate individual is to be inserted only if it is not similar to the last member (worst fitness rate) of the set and also illustrates a better fitness score.

We consider two individuals $\mathbf{m}^{(1)}, \mathbf{m}^{(2)}$ similar if and only if the absolute normalized difference, as described in the following equivalence, is less than a certain threshold ϵ. Hence

$$\mathbf{m}^{(1)} \text{ is similar to } \mathbf{m}^{(2)} \iff \frac{1}{M}\sum_{j=1}^{M}\frac{|m_j^{(1)} - m_j^{(2)}|}{b_j - a_j} < \epsilon. \tag{4.170}$$

Additionally, in our implementation we perform a forced mutation ($p_{\text{mut}} = 1$) to the innovative population in order to promote the exploration in possibly interesting areas.

As it is clear, the above procedure does not assure that the best individual that ever exists is part of the innovative population. Therefore, in order to make sure that we keep alive the individuals with the best fitness rates ever occurred, we also have decided to track a small

number N_{el} of those individuals. This sub-population would be the second subset of the HOF and we call it by the term *elite population* [106].

To summarize, at each iteration, HOF consists of a totally $N_{\text{hof}} = N_{el} + N_{in}$ number of individuals, where

- The first N_{el} are introduced such that to be the best individuals up to now. Keeping these individuals has been proved to give a significant boost to the convergence speed.
- The last N_{in} have been collected with respect to a trade off between their fitness rates and the similarity measure 4.170 which supports the diversity of the population.

chapter 5

Inverse Problems of Acoustical Oceanography using Statistical and Probabilistic Features

Abstract

In this chapter we describe two inversion schemes that can be applied to problems of ocean acoustic tomography and geoacoustic inversions (see Chapter 3). The inversion schemes exploit statistical or probabilistic features of a measured acoustic signal to estimate parameters of the marine environment through an appropriate optimization procedure that associates features of the measured acoustic signal with the features of a simulated one calculated on the basis of possible values of the recoverable parameters (replica fields). We will make use of the mathematical background presented in Chapter 4. The inversion scheme based on the statistical features has already been applied in problems of acoustical oceanography, while the scheme based on the probabilistic features of the signals is a novel one. As both schemes are affected by the noise contained in the measured signal, it is essential that a denoising procedure is applied prior to its exploitation. The first section of the chapter is dedicated to the presentation of a denoising scheme. The second section presents the statistical signal characterization scheme while the third one is devoted to the probabilistic signal characterization scheme. The final section describes the inversion procedure based on the two alternative characterization schemes.

5.1 De-noising Scheme

An acoustic signal in the time domain $\mathbf{s}(t)$ is considered here in its digital form as a set of discrete samples

$$\mathbf{s}[n] = \mathbf{s}'[n] + \mathbf{w}[n],\ n = 1, \ldots, N \tag{5.1}$$

where $\mathbf{s}'[n]$ is the noise-free part of the signal and $\mathbf{w}[n]$ is added noise.

In inverse problems formulated on the basis of observables which are not directly related to the spectrum of the signal (e.g. ray or modal arrivals), there is also a reduction to the performance of the inversion scheme, due to presence of noise.

However, in both the statistical and probabilistic characterization schemes, the features of the signals are extracted taking into account the energy content of the signal in the whole bandwidth. Therefore, the noise in those schemes has a more serious affect to the feature extraction procedure applied to the recorded signal.

It should be noted that recently, an additional possible contamination factor of the recorded underwater acoustic signals has been studied. Specifically, the signal was considered to be both blurred and noisy and it was modeled mathematically as follows:

$$\mathbf{s} = \mathbf{A}\,\mathbf{s}'[n] + \mathbf{w}[n],\ n = 1, \ldots, N, \tag{5.2}$$

where the matrix $\mathbf{A}$ quantifies the blurring mechanism.

Deblurring of the signal can be achieved by means of a technique introduced by Taroudaki and O' Leary [107] for image deblurring, and it is based on a statistical near optimal spectral filtering technique that takes advantage of the singular values of the approximated blurring matrix and the Picard Parameter of the signal that allows for estimation of the additive noise properties and estimation of the error. This deblurring-denoising technique has been studied using both synthetic and actual signals with promising results [108, 109], but will not be applied in the case studies presenting in this thesis. In the framework of this thesis, we will apply a denoising scheme to enhance the quality of the inversion results. Due to the nature of this scheme it will be denoted as the **Sparse Denoising Scheme (SDS)**. Given a noisy acoustic signal in form of (5.1), we divide this into overlapping windows $\mathbf{s}_j$, each one of length L using the maximum overlapping rate of $L-1$ samples [110]. Hence, the $j-th$ window $\mathbf{s}_j$ is given in vector form by

$$\mathbf{s}_j = \{\mathbf{s}'[n] + \mathbf{w}[n],\ n \in \mathbb{Z} \cap [j, j+L-1]\}, \tag{5.3}$$

where the index $j \in \mathbb{Z} \cap [1, N-L+1]$.

In Figure 5.1 an example of overlapping windowing is illustrated. For simplicity reasons, we will denote $N-L+1$ as M. Also, in the rest of the chapter we will omit the argument $[n]$.

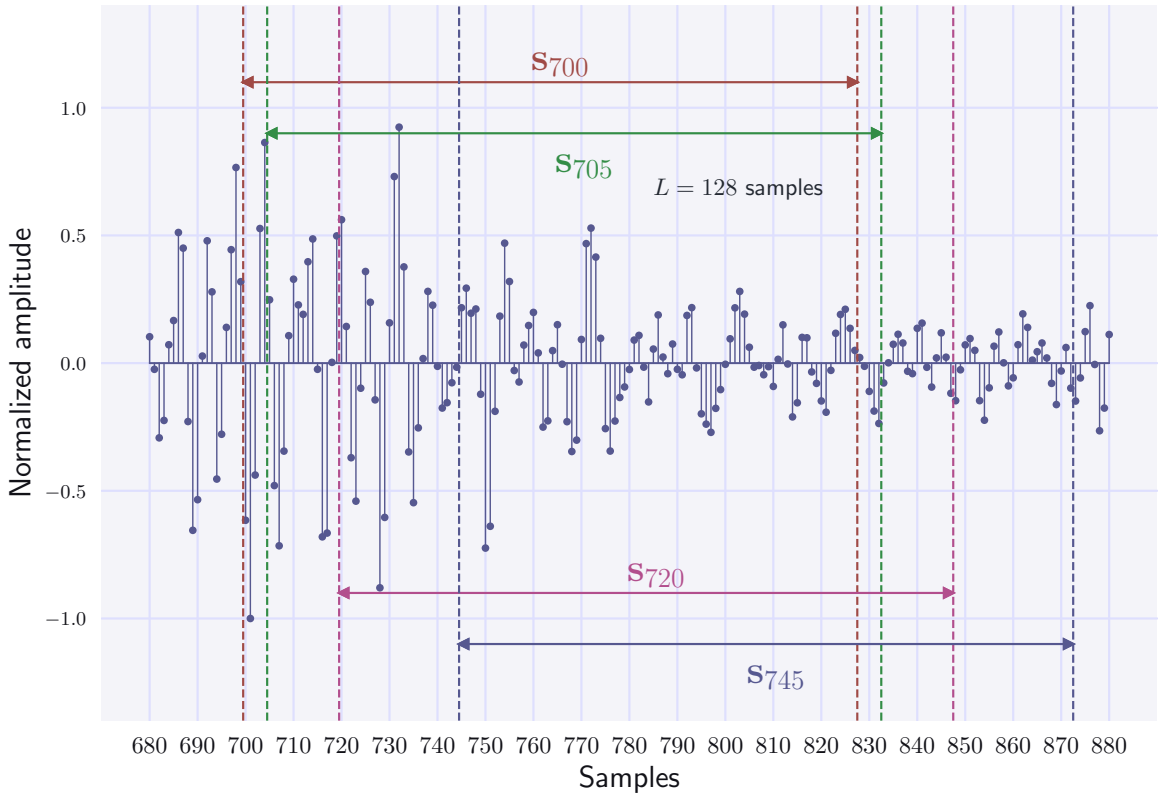


Figure 5.1: A zoomed version of the discrete signal from the SW06 experiment. This figure depicts the $\mathbf{s}_{600}$, $\mathbf{s}_{601}$, $\mathbf{s}_{610}$, $\mathbf{s}_{640}$ overlapping windows (each of length 128 samples) of the signal.

Thus, we define a family of overlapping windows $\mathbf{S} = \{\mathbf{s}_j \in \mathbb{R}^L\}_{j=1\ldots M}$. Following the theory presented in Chapter 4.6 we define the dictonary matrix $\mathbf{D} \in \mathbb{R}^{L,M}$, so that, each specific window $\mathbf{s}_j$ is estimated in the form as $\tilde{\mathbf{s}}_j$ given by the following formula

$$\tilde{\mathbf{s}}_j = \mathbf{D}\mathbf{a}_j, \tag{5.4}$$

where the coefficients $\mathbf{a}_j$ are appropriate vectors $\{\mathbf{a}_j \in \mathbb{R}^M\}_{j=1\ldots M}$ to be also defined.

The proper dictionary matrix $\mathbf{D}$ is obtained by solving the following sparse dictionary learning problem:

$$\min_{\mathbf{D}\in\mathbb{R}^{L,M}} \frac{1}{M} \sum_{j=1}^{M} \min_{\mathbf{a}_j\in\mathbb{R}^M} \left\{ \frac{1}{2}\|\mathbf{s}_j - \mathbf{D}\mathbf{a}_j\|_2^2 + \lambda\|\mathbf{a}_j\|_1 \right\} \tag{5.5}$$

The solution of an optimization problem of this form was also discussed in Section 4.6.

Once the dictionary $\mathbf{D}$ has been defined and the coefficient vectors $\hat{\mathbf{a}}_j$ have been calculated for all windows $\mathbf{s}_j$, the approximation of a denoised version $\hat{\mathbf{s}}_j$ for each window frame $\mathbf{s}_j$ is given by:

$$\hat{\mathbf{s}}_j = \mathbf{D}\,\hat{\mathbf{a}}_j, \quad j = 1, \ldots, M. \tag{5.6}$$

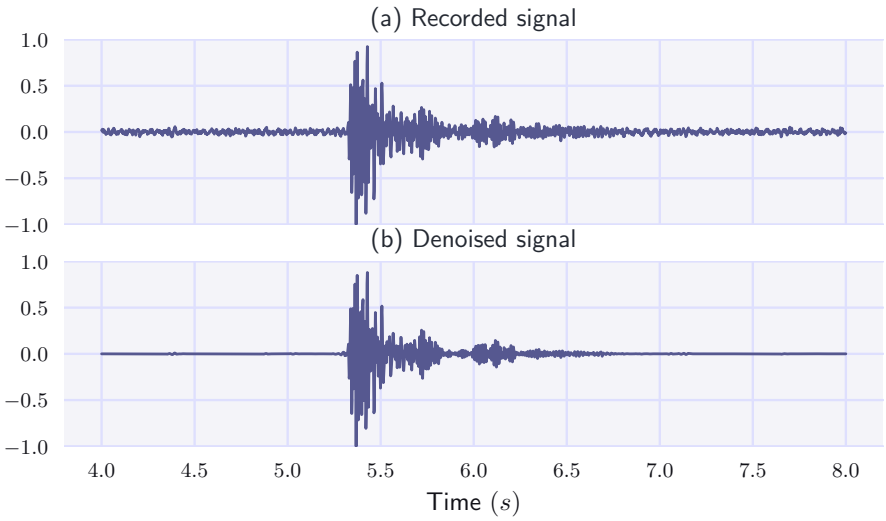


Figure 5.2: The raw and denoised signal from the SW06 experiment.

Thus, we have $c(n)$ segments each of them including an estimation of the $n-$ th sample of the signal. $c(n)$ is given by the function :

$$c(n) = \begin{cases} n & \text{if } n \in [1, L) \\ L & \text{if } n \in [L, N-L] \\ N-n+1 & \text{if } n \in (N-L, N]. \end{cases} \tag{5.7}$$

We define an approximation of the actual signal $\mathbf{s}'$ composed by digits, each one being the expected value of all the approximations as described above. Assuming that every approximation has the same impact to the final estimated value we get the approximation $\hat{\mathbf{s}}$ as a typical mean value of the observations. Hence

$$\hat{\mathbf{s}}[n] = \frac{1}{c(n)} \sum_{k=\max(1,n-L+1)}^{\min(n,M)} \hat{\mathbf{s}}_k[n-k+1]. \tag{5.8}$$

The signal $\hat{\mathbf{s}}[n]$ is a good approximation of the noise-free signal $\mathbf{s}'[n]$. This statement, although not rigorously proved, is justified by the fact that the procedure degrades the noisy component of the signal $\mathbf{w}[n]$ and emphasizes its energy significant part. Note that, there is a sharing of the dictionary matrix $\mathbf{D}$ among all the window frames of the signal forcing its columns to reflect patterns of similar frames. This observation is the key factor behind the denoising strategy as by the procedure described above, the dictionary is forced to describe quite well all the frames in a sparse and compressible way. When the sparsity terms $\lambda\|\mathbf{a}_j\|$ tend to zero, the coefficients $\mathbf{a}_j$ are in dense form and take into account the redundancy of the dictionary, the iterative procedure yields the noisy signal. Generally, if a frame is described in a dense

representation, then a lot of the dictionary vectors would be involved to the representation and the stochastic feature of the noise could be efficiently described.

On the other hand, if h has a strong impact in the cost function, the coefficients will be in extreme sparse forms. Therefore, the approximation will omit significant signal parts including parts associated with the noise-free signal. In addition, similar frames will tend to have same representation and the reconstructed signal will be considered as a bad approximation of the real noise-free one.

As is easily understood, the method implemented in this work constitutes a trade off between sparsity and proper reconstruction, giving an estimation of the signal in which most of the components of the noise-free signal are kept, and omitting most of its noisy parts. In other words, the fewer vectors are used in the estimation of the noise-free version of the signal, the less external structures (noise) are revealed with some compromise on quality of the approximation error of the observed signal. Figure 5.2 presents an example of signal denosing applied to a measured during the SW06. Detals of SW06 experiment presented in Chapter 6 of this thesis.

5.2 The Statistical Signal Characterization Scheme (SSCS)

This scheme associates a signal with statistical features of the coefficients of a transformation of the acoustic signals via an one-dimensional wavelet decomposition and then by fitting the empirical distributions of the sub-band coefficients using an appropriate statistical distribution.

Taroudakis et al. [1] have shown that for a general class of acoustic signals used in applications of acoustical oceanography these empirical statistical distributions are symmetric, have heavy algebraic tails, and then can be statistical described by Symmetrical alpha-Stable (SαS) distributions. The symmetric alpha-stable distributions, have proven to be efficient in describing the content of many texture images [111, 112].

The similarity measurements between two acoustic signals statistical characterized as above, is performed through a proper version of the KLD between their corresponding SαS distributions.

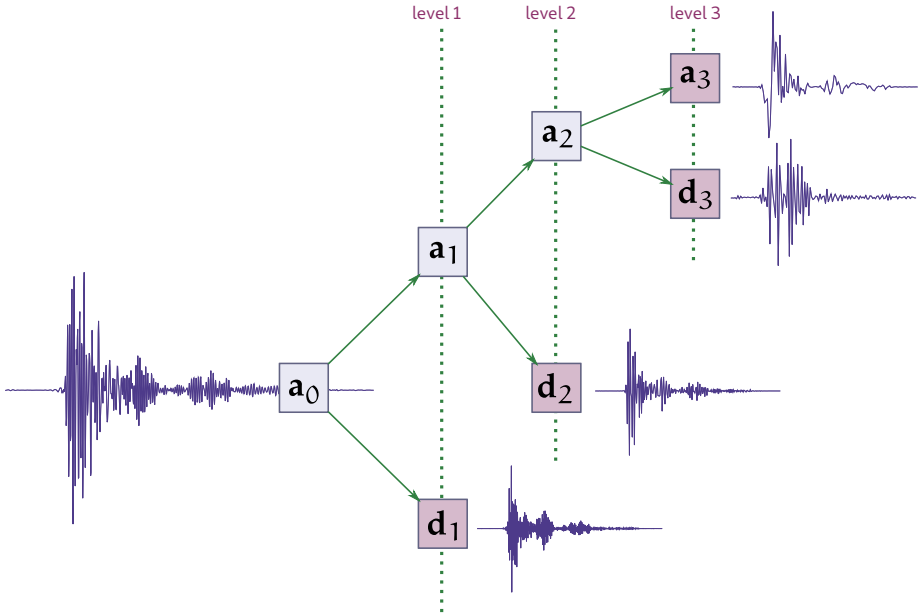


Figure 5.3: Wavelet multiresolution analysis at 3 levels of a recorded acoustic signal obtained by the SW06 experiment, corresponding to a light-bulb implosion after a denoising procedure.

5.2.1 Feature extraction

— Wavelet multiresolution analysis —

Consider a digital acoustic signal $\mathbf{s}[n]$, $n = 1, \ldots, N$ measured at a specific location in the water column. By applying a 1-D Discrete Wavelet Transform (DWT) to the signal $< \mathbf{s}, \psi_{a,b} >$, where $\psi_{a,b}$ is an appropriately chosen wavelet, we obtain an N-dimensional vector of coefficients $\mathcal{A}_0$ representing the signal ($\mathbf{s}[n] \longleftrightarrow \mathcal{A}_0$). The signal can in principle be reconstructed by applying the inverse wavelet transform to its transformed version. For applications of signal characterization, a single (zero order) level analysis of the signal is not enough. Following the work by Mallat [73], we proceed with a multiresolution analysis employing the 1D Discrete Wavelet Transform. This procedure was described more analytically in Section 4.2.3, and it is summarized here. Starting with the discrete signal $\mathbf{s}[n]$ we calculate the first level of decomposition which consists of two vectors $\mathcal{A}_1$ and $\mathcal{D}_1$, each one of dimension about $N/2$. These vectors are obtained by convolution of the signal with suitable discrete filters followed by down sampling.

$$\mathcal{A}_1[n; \mathbf{s}] = \mathcal{A}_0 * L[2n] \tag{5.9}$$

$$\mathcal{D}_1[n; \mathbf{s}] = \mathcal{A}_0 * H[2n] \tag{5.10}$$

The vector $\mathcal{A}_1$ is the result of the passage of the signal from a low-pass filter L and is called the first level approximation coefficients vector of the decomposition. Similarly, the $\mathcal{D}_1$ results from the passage of the signal from a high-pass filter H and is called the first level detail coefficients vector of the decomposition.

At the second level of the decomposition, the vector $\mathcal{A}_1$ of approximation coefficients is decomposed in two coefficients vectors using the same process and replacing only $\mathbf{s}[n]$ by $\mathcal{A}_1$ and then producing $\mathcal{A}_2$ and $\mathcal{D}_2$ vectors each one with dimension about $N/2^2$. At an arbitrary level j we get:

$$\mathcal{A}_{j+1}[n;\mathbf{s}] = \mathcal{A}_j * L[2n] \tag{5.11}$$

$$\mathcal{D}_{j+1}[n;\mathbf{s}] = \mathcal{A}_j * H[2n] \tag{5.12}$$

This procedure continues in the same way up to the k-th level of decomposition, where the $(L-1)$-th level approximation coefficients vector is $\mathcal{A}_{L-1}$ and producing $\mathcal{A}_L$ and $\mathcal{D}_L$ the L-th level approximation and detail coefficients vectors respectively, each one with dimension about $N/2^L$.

— Statistical modelling of wavelet sub-band coefficients —

The next step of the scheme is the statistical characterization of the signal wavelet coefficients. Initially we calculate a-posteriori statistical distributions of the wavelet coefficients $\mathcal{D}_1, \mathcal{D}_2, ..., \mathcal{D}_L, \mathcal{A}_L$ and associate each one of them with a Symmetric Alpha Stable distribution $(S\alpha S)$ using the characteristic function:

$$\Phi(t) = \exp(i\delta t - \gamma^\alpha |t|^\alpha), \tag{5.13}$$

where $0 < \alpha \leq 2$ is the characteristic exponent which controls the marginal behavior of tails, $-\infty < \delta < \infty$ is the location parameter and $\gamma > 0$ is the dispersion of the distribution which determines the spread of the distribution around the location parameter δ [113].

We know that the statistical distributions of the wavelet coefficients are concentrated around $t = 0$, whenever each coefficient is characterized only by the parameters α and γ. Accordingly, an arbitrary acoustic signal S is characterized by a feature vector $\mathbf{d}$ of dimension $2L + 2$ as follows

$$\mathbf{s} \leftrightarrow \{\Phi^0, ..., \Phi^L\} \leftrightarrow \mathbf{d} = [\alpha^0, \gamma^0, \alpha^1, \gamma^1, ..., \alpha^L, \gamma^L]^T, \tag{5.14}$$

where a particular pair α_i, γ_i consists of the estimated model parameters that correspond, for $i = 0$ to the approximation subband at the L-th decomposition level and for $i = 1, ..., L$ to the detail subband at the i-th decomposition level.

— A feature extraction example —

In this section we introduce an example of the signal characterization technique described above, using a real acoustic signal in the time domain (Figure 5.4). The example signal was

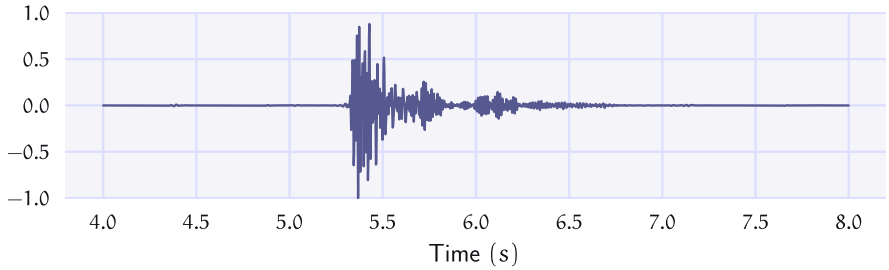


Figure 5.4: The denoised recorded acoustic signal. The denoising has been done using the sparse denoising method introduced in the previous section.

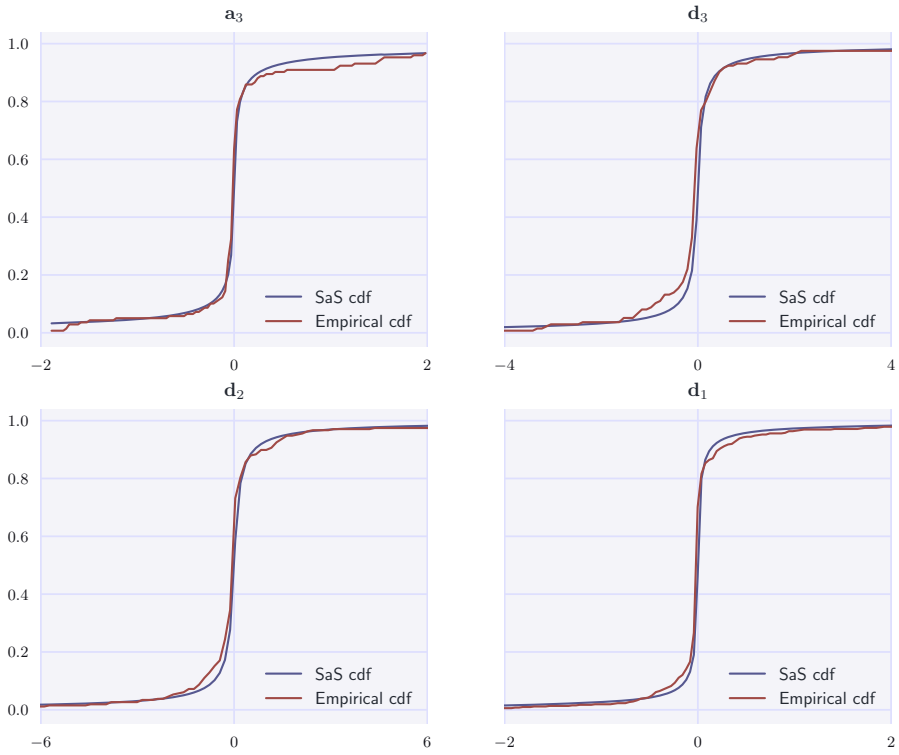


Figure 5.5: Modelling the cumulative density functions of the four wavelet sub-band (one approximation and three details) of the three-level decomposition of the SW06 experiment's recorded signal.

due to an implosion of a lightbulb at the water and it has been recorded during the Shallow Water 2006 (SW06) experiment. This characterization will be the bases for the retrieval of the parameters of the sea-bed by introducing three different parameterized models of the experimental environment as we will present in Chapter 6. We perform the feature extraction process using a 3-level multiresolution analysis and statistical modeling of the wavelet coefficients using the symmetric alpha-stable distribution. The characterization scheme gives the following feature vector that contains the SaS parameters of the four sub-bands.

$$\mathbf{d} = [0.5929, 0.0314, 0.7539, 0.0855, 0.8013, 0.1447, 0.6799, 0.0215]^T. \tag{5.15}$$

Figure 5.5 compares the SaS cumulative distribution function $p(X < x)$ with the empirical ones. This figure shows that the SaS distribution and the corresponding empirical distribu-

tions fit quite well, which is an indication that SaS distribution represents well the statistics of the wavelet sub-band coefficients of an real underwater acoustic signal.

5.2.2 Similarity measurements

In order for a classification scheme to be usable for some inversion procedure, a measure of the difference between two signals characterized by the statistics of the wavelet coefficients should be defined. As shown in [1] an appropriate similarity measure is defined by the Kullback-Leibler Divergence (KLD) [66] already presented in which expresses the distance D_s between two acoustic signals $\mathbf{s}_1$ and $\mathbf{s}_2$, represented by their feature vectors $\mathbf{d}_1$ and $\mathbf{d}_2$ respectively, and has the following form:

$$D_s(S_1, S_2) = D_s(\mathbf{d}_1, \mathbf{d}_2) = \sum_{k=0}^{L} \mathrm{KLD}(\Phi_1^k || \Phi_2^k). \tag{5.16}$$

Here, $\mathrm{KLD}(\Phi_1^k \| \Phi_2^k)$ is the Kullback-Leibler divergence between the Φ_1^k and Φ_2^k distributions, which are evaluated using the estimated *SaS* parameters (α_2^k, γ_2^k) and (α_1^k, γ_1^k), respectively and it is expressed by the closed form

$$D(\Phi_1^k \| \Phi_2^k) = \ln\left(\frac{c_2^k}{c_1^k}\right) - \frac{1}{\alpha_1^k} + \left(\frac{\gamma_2^k}{\gamma_1^k}\right)^{\alpha_2^k} \frac{\Gamma(\frac{\alpha_2^k+1}{\alpha_1^k})}{\Gamma(\frac{1}{\alpha_1^k})}, \tag{5.17}$$

where $\Gamma(\cdot)$ is the Gamma function and

$$c_i^k = \frac{2\Gamma(\frac{1}{\alpha_i^k})}{\alpha_i^k \gamma_i^k} \ , \ i = 1, 2.$$

Formula 5.16 is based on the assumption that the statistical character of the wavelet coefficients at each level is independent to that of another level.

5.3 The Probabilistic Signal Characterization Scheme (PSCS)

In this section, the new proposed approach for characterizing acoustic signals is presented. It is based on a wavelet packet analysis of the waveforms, followed by an adaptation of a HMM to a transformed version of selected wavelet packet coefficients.

The SSCS was introduced as a feature extraction technique of acoustic signals that uses no physical observables of them. As it is known obtaining physical observables consists a generally difficult procedure. The SSCS as it was presented in the previous section associates each wavelet subband coefficient vector with a suitable statistical distribution. In order to

determine for each subband coefficient vector a corresponding $S\alpha S$ distribution, we have assumed that the elements of any coefficient vector are independent and identically distributed (i.i.d). Although, this assumption might be sufficient for signals with no significant spectral variations with the time this is probably is a problematic assumption for propagated pulses through a dispersive waveguide where the measured signal consists of a sequence of various energy packet arrivals. Other issues of the SSCS are the assumption of statistical independency among the wavelet sub-band coefficients of the signal and the lack of the translation invariant property of the wavelet transform.

In the proposed signal characterization scheme, we have relaxed the (i.i.d) assumption by employing a HMM to characterize the extracted wavelet packet coefficients of the signal. In the new approach, the wavelet packet features of a signal form a time series of vectors, each one of which contains the spectra information of the signal for a (discrete) time. We will describe this time series by a group of multivariate Gaussian distributions. These distributions are chosen with full covariance matrix to take into account possible dependencies among coefficients belonging to different subbands of the signal decomposition. In addition, the scheme uses the stationary version of the wavelet packet transform (see Section 4.2.6) which offers a translation invariant characterization.

5.3.1 Feature extraction

— Stationary wavelet packet decomposition —

As it has been mentioned, the stationary wavelet decomposition provides a time invariant time-frequency decomposition, giving good resolution in both time and frequency domains. Therefore the wavelet packet coefficients keep the important time-frequency information carried by the original signal in the time domain.

Following the notation of Section 4.2, we perform time-frequency analysis of the signals using J decomposition levels. Then we keep only the nodes corresponding to the J-th (last) level as a signature of the signal.

$$\mathbf{s}[n] \Rightarrow \left\{ \boldsymbol{\nu}_J^0, \boldsymbol{\nu}_J^1, \ldots, \boldsymbol{\nu}_J^{J-1} \right\} \quad \text{(coefficients in frequency order).} \tag{5.18}$$

(See Section 4.2.4 for the definition of the $\boldsymbol{\nu}_J^j$).

Note that, according the properties of the wavelet packet transform, we can reconstruct the coefficients of any k-th level where $k < J$, so there is no loss of information. However, a

more sophisticated approach will be applied, by finding the best decomposition by keeping coefficients at various decomposition levels, as will be presented in Section 7.2.1. This is one of the goals of our future plans.

Each sub-band vector $\mathcal{V}_j^i$ can be associated with an energy signature, by means of the Euclidean norm, $\mathcal{E}_j^i$ as follows:

$$\mathcal{E}_j^k = \frac{\|\mathcal{V}_J^k\|_2^2}{\sum_{i=0}^{2^J-1}\|\mathcal{V}_J^i\|_2^2},\ k = 0,\ldots,2^J-1. \tag{5.19}$$

— The time-frequency feature matrix —

In order to decrease the dimension of the feature space, we keep only the sub-bands with energy content above a certain threshold. Additionally, this could contribute to the reduction of the noise influence on the decomposition coefficients due to the noise. Therefore, we get the characterization

$$\begin{aligned} \text{signal} \ &\Rightarrow \ \left\{\mathcal{V}_J^k \mid \mathcal{E}_J^k > \text{threshold}\right\} \\ &= \ \left\{\mathcal{V}_J^{k_1}, \mathcal{V}_J^{k_2} \ldots, \mathcal{V}_J^{k_D}\right\}. \end{aligned} \tag{5.20}$$

In this work, we have considered a threshold $= 0.025$, corresponding to 2.5% of the total energy of the decomposition.

We then gather the vectors $\{\mathcal{V}_J^\ell\}_{\ell=1}^D$ to form the following matrix Y as

$$\mathbf{Y} = \begin{bmatrix} \longleftarrow & \mathcal{V}_j^{k_1} & \longrightarrow \\ \longleftarrow & \mathcal{V}_j^{k_2} & \longrightarrow \\ & \vdots & \\ \longleftarrow & \mathcal{V}_j^{k_D} & \longrightarrow \end{bmatrix} \in \mathbb{R}^{D\times N'} \tag{5.21}$$

At this point, we adopt the following notation:

(*Notation*) Consider a MxN matrix, generally denoted by $\mathbf{A}$. We shall denote by $\mathbf{a}_{ri}$ the i-th, $i \in \{1,2,\ldots,M\}$, row of this matrix. Furthermore, we shall denote by $\mathbf{a}_{cj}$ the j-th, $j \in \{1,2,\ldots,N\}$, column of the matrix $\mathbf{A}$.

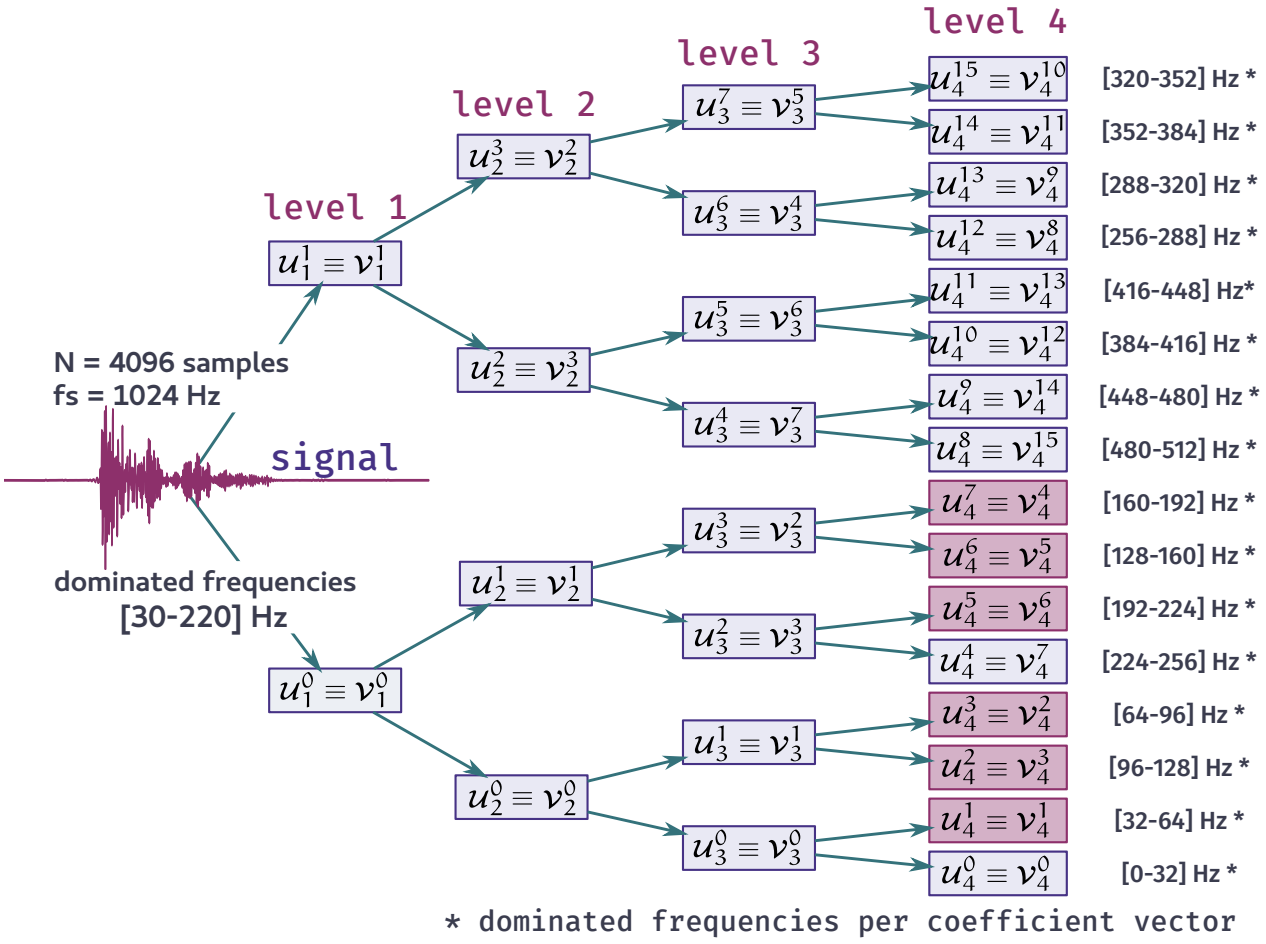


Figure 5.6: Wavelet packet decomposition, in both natural and frequency order, of a recorded acoustic signal obtained by the SW06 experiment, corresponding to a lightbulb implosion after a denoising procedure.

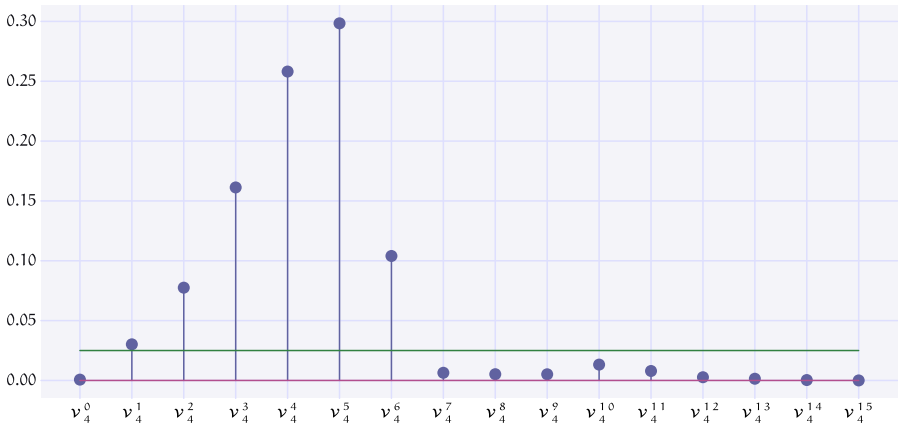


Figure 5.7: Energy contribution per wavelet packet coefficient sub-band. Threshold has picked to be 0.025 (green horizontal line).

Following the previous notation, we can express the $D \times N'$ matrix $\mathbf{Y}$ using both the row-wise and column-wise definitions as follows:

$$\mathbf{Y} = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \mathbf{y}_{c1} & \mathbf{y}_{c2} & \cdots & \mathbf{y}_{cN'} \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \text{ (column-wise notation)} \tag{5.22}$$

$$= \begin{bmatrix} \longleftarrow & \mathbf{y}_{r1} & \longrightarrow \\ \longleftarrow & \mathbf{y}_{r2} & \longrightarrow \\ & \vdots & \\ \longleftarrow & \mathbf{y}_{rD} & \longrightarrow \end{bmatrix} \text{ (row-wise notation).} \tag{5.23}$$

Then, we can now find two indexes n_1 and n_2 for which the columns $\mathbf{y}_{c1}, \ldots, \mathbf{y}_{c,n_1-1}$ and $\mathbf{y}_{c,n_2+1}, \ldots, \mathbf{y}_{cN'}$ have energies below certain thresholds that depend on the noise level of the considered signal. In brief, we take an envelope of the energy of the coefficients per column of $\mathbf{Y}$ and the indexes n_1 and n_2 corresponds to the columns enclosing the effective wavelet packet coefficients of the signal.

Hence, an acoustic signal can be characterized using a matrix $\mathbf{X}^*$ determined by the columns of the matrix $\mathbf{Y}$ between $\mathbf{y}_{cn_1}$ and $\mathbf{y}_{cn_2}$. Thus,

$$\mathbf{X}^* = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \mathbf{x}^*_{c1} & \mathbf{x}^*_{c2} & \cdots & \mathbf{x}^*_{cN} \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}, \tag{5.24}$$

where $\mathbf{x}^*_{cj} = \mathbf{y}_{c,n_1-1+j}$, $j = 1, \ldots, N$ and $N = n_2 - n_1 + 1$ is the number of the columns of $\mathbf{X}^*$.

Since each row of $\mathbf{X}^*$ corresponds to a different frequency sub-band, and each band carries different information about the source properties as well as the properties of the environment through which the signal has been transmitted, a normalization process is applied to ensure that the characterization of the signal is as sensitive as possible to the variation of environmental parameters, such as the densities of the sub-bottom layers which according to the literature are difficult to be retrieved from physical observables relative to the acoustic field.

Furthermore, in problems of pattern recognition a standard preprocessing technique is applied to the observations in order that all the feature components to be in a notionally common scale. In our case, is to be achieved by standardization of the magnitude of each row of $\mathbf{X}^*$ to have a zero mean value and unity standard deviation.

It should be noted, that the proposed characterization scheme focuses on the modeling of the energy motifs revealing the wavelet packet coefficients of the signals. The (temporal or whole) energy of the coefficients is evaluated by considering the magnitude of the coefficients. This gives as a side effect that the characterization of a signal with our approach will lack of phase information of wavelet packet coefficients.

We would now like to describe the magnitude of the Wavelet packet coefficients via a number of Normal distributions. To achieve a better adaptation of the coefficients with the normal distribution, we will also smooth these coefficients by projecting them onto the logarithmic scale. Such a smoothing transformation as a preprocessing mechanism in general decreases the representations misfit of a Gaussian distribution based modeling of the observed data [114, 115, 116, 117].

It should be noted that a subject of future research is associated with the potential use of a family of more general distributions such as mixtures of Gaussian distributions that would possibly give superior characterization properties.

Taking into account the above considerations, the matrix $\mathbf{X}^*$ is transformed to a new matrix $\mathbf{X}$ of the form

$$\mathbf{X} = \begin{bmatrix} \longleftarrow & \mathbf{x}_{r1} & \longrightarrow \\ \longleftarrow & \mathbf{x}_{r2} & \longrightarrow \\ & \vdots & \\ \longleftarrow & \mathbf{x}_{rD} & \longrightarrow \end{bmatrix}, \tag{5.25}$$

where its rows are given by:

$$\mathbf{x}_{ri} = \log\left(\frac{|\mathbf{x}^*_{ri} - \overline{\mathbf{x}}^*_{ri}|}{\sigma_{\mathbf{x}^*_{ri}}}\right),\ i = 1, \ldots, D, \tag{5.26}$$

where the bar denotes the mean value, and σ denotes the standard deviation of a set of values. This matrix consists of the final time-frequency features that we consider.

Note that, so far, the feature extraction procedure has been deterministic. Next, a signal is to be further characterized by a stochastic model over these extracted features.

— Associating the extracted Wavelet packet coefficients with a HMM —

In the previous subsection we saw that, a signal can be associated with a matrix $\mathbf{X}$ by transforming selected wavelet packet coefficients of the signal. Each column of this matrix provides information about the energy of the signal in several frequency subbands at a specific time interval. Therefore, we can consider the columns of the matrix $\mathbf{X}$ as a time series of realization of high-dimensional random variables which characterizes the original waveform. Let $\mathbf{x}_n$ be the n-th random variable corresponding to a measured signal. For each $\mathbf{x}_n$ we introduce a hidden (unobserved) discrete random variable $\mathbf{z}_n$, such that each $\mathbf{x}_n$, given any possible realization (state) of the corresponding $\mathbf{z}_n$ to obey a Gaussian distribution. In addition, we assume that $\mathbf{z}_n$ depends only on the very previous hidden random variable $\mathbf{z}_{n-1}$. Now, we can characterize the matrix $\mathbf{X}$ containing the time-frequency features of the signal with a single HMM with Gaussian emission distributions.

According to the definitions presented in Section 4.5.1, a HMM can be parameterized by the set $\boldsymbol{\lambda}$:

$$\boldsymbol{\lambda} = \{\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\Phi}\}, \tag{5.27}$$

where $\boldsymbol{\pi}$ is a vector which contains the initial state probabilities, $\mathbf{A}$ is the transition matrix with elements of the transition probabilities that express the probability for moving from one hidden state to another as the time passes, and finally $\boldsymbol{\Phi}$ contains the parameters for the Gaussian emission distributions that control the observed time series $\mathbf{X}$.

Considering that these time series are the result of the feature extraction of underwater acoustic signals transmitted in a dispersive waveguide, it is reasonable to assume that time-frequency features progress over time, transitioning from one state to an other one. Therefore, a restricted transition model is considered here instead of a fully-connected one in which the state of the hidden variable is allowed to be transitioned from each state at the current time step to every other at the next time step .

The considered HMM belongs to the so called *left-to-right* HMMs, due to the fact that the order of the hidden state can either stays the same or increases as time increases. The transition probabilities in such a case are modeled by an upper-triangular matrix. In our model we apply an additional constrain on the transition matrix in which no steps of more than one state a time is allowed (Bakis topology), which is extensively used in applications of digit or speech recognition [118, 119, 120, 121, 122, 123]. Therefore, in the general case when K hidden states are considered, the transition matrix takes the form:

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} & 0 & 0 & \dots & 0 \\ 0 & A_{22} & A_{23} & 0 & \dots & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & \dots & 0 & A_{K-2,K-2} & A_{K-2,K-1} & 0 \\ 0 & \dots & 0 & 0 & A_{K-1,K-1} & A_{K-1,K} \\ 0 & \dots & 0 & 0 & 0 & 1 \end{bmatrix} \tag{5.28}$$

Note that for the last state in a left-to-right model, we have $A_{KK} = 1$. Furthermore, the form of the transition matrix forces the initial state probabilities to have the following constant values

$$\boldsymbol{\pi} = [1.0,\ 0.0, \dots,\ 0.0]^T \in \mathbb{R}^K. \tag{5.29}$$

Since any predefined zero in the transition matrix or the initial state probabilities, remains zero as the EM algorithm is being performed, the application of the above constraints in the training procedure realized by the EM algorithm is straightforward.

To summarize, a signal can be characterized by its time-frequency feature matrix $\mathbf{X}$ and eventually through the HMM by means of the set $\boldsymbol{\lambda}$ (5.27). Thus, this generative model after being trained with respect to the extracted wavelet packet features of the signal, consists of a representative set of parameters of the signal. The joint probability distribution over both observed and hidden variables is given with respect to the set $\boldsymbol{\lambda}$ by

$$p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\lambda}) = p(\mathbf{z}_1|\boldsymbol{\pi}) \left(\prod_{n=2}^{N} p(\mathbf{z}_n|\mathbf{z}_{n-1}, \mathbf{A}) \right) \prod_{n=1}^{N} p(\mathbf{x}_n|\mathbf{z}_n, \boldsymbol{\Phi}). \tag{5.30}$$

Therefore, a signal can be associated with a single HMM by using the **Expectation-Maximization** algorithm as it had been described in Section 4.5.2, which is composed by the following steps:

1. Start with a initial model $\boldsymbol{\lambda}^{\text{old}}$ as described in Section 4.5.2.

2. **E step** : Evaluate the posterior distribution over the $\mathbf{Z}$, $p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\lambda}^{\text{old}})$.

3. **M step** : Maximize the expectation $\mathbb{E}_{\mathbf{Z}|\mathbf{X},\boldsymbol{\lambda}^{\text{old}}}\{\ln p(\mathbf{X},\mathbf{Z}|\boldsymbol{\lambda})\}$ with respect to $\boldsymbol{\lambda}$.

4. Evaluate the log likelihood $\ln p(\mathbf{X}|\boldsymbol{\lambda})$. If the convergence is not achieved, set $\boldsymbol{\lambda}^{\text{old}} = \boldsymbol{\lambda}$ and go to step 2.

Taking into account the stochastic nature of the training process, the resulted model may vary per realization and sometimes the algorithm might lead to a local instead of the global maximum, for the likelihood function. To address such issues we perform several realizations of the EM algorithm and we keep the model that gives the best performance with respect to the value of the likelihood.

The remaining issue that we have to address is the calculation of the number K of the model's hidden states, such that the observations $\mathbf{X}$ to be optimally characterized by a learned HMM. Let us denote by $\boldsymbol{\lambda}_K$ a generic k-state HMM over the observables. We introduce the posterior distribution of the observation given the assumption of a k-state modeling that is evaluated on the basis of the integration of $p(\mathbf{X}|K,\boldsymbol{\lambda}_K)$ over all possible k-th state HMMs. This posterior probability is written in the following form

$$p(\mathbf{X}|K) = \int p(\mathbf{X}|K,\boldsymbol{\lambda}_K)d\boldsymbol{\lambda}_K. \tag{5.31}$$

Note that, the calculation of the optimal number of the model's hidden states requires an optimization over the integral, therefore

$$K^{\text{opt}} = \arg\max_K p(\mathbf{X}|K). \tag{5.32}$$

Unfortunately, there is no simple way of evaluating the above integral, as a result the previous optimization problem is hard to solve. However, an estimation of K^{opt} can be done by means of the Bayesian Information Criterion (BIC) [124] which consists of an asymptotic approximation of the logarithm of the integrated likelihood function. Following the work by Celeux and Durand [125] this approximation for HMMs is given by

$$p(\mathbf{X}|K) \approx \text{BIC}(K) = \ln p(\mathbf{X}|K,\hat{\boldsymbol{\lambda}}_{\boldsymbol{K}}) - \frac{\nu_K}{2}\ln N, \tag{5.33}$$

where $\hat{\boldsymbol{\lambda}}_{\boldsymbol{K}}$ is the trained model after performing the EM when K hidden states are considered, N is the length of the sequence of the observables that it is equal to the number of the columns of $\mathbf{X}$, and with ν_K the model free parameters. The chosen structure of the HMMs leads to the following number of free model parameters:

$$\nu_K = K * \left(2 + \frac{D^2+3D}{2}\right) - 2, \tag{5.34}$$

where D as before denotes the dimensionality of the observations.

As a result, the number of the hidden states can be estimated by the following simple maximization problem:

$$K^{\text{opt}} = \arg\max_{K} \text{BIC}(K). \tag{5.35}$$

— A feature extraction example —

In this section, we present an example of the signal characterization technique described above, using the real acoustic signal presented in Figure 5.4.

After the feature extraction using the wavelet packet transform with 4 levels (Figure 5.6) and keeping the columns with energy above a certain threshold, we get the feature matrix $\mathbf{X} \in \mathbb{R}^{6,\,502}$ using 6 wavelet packet coefficient bands corresponding to relative energy above of 2.5% as illustrated in Figure 5.7. The cropped signal and the rows of the $\mathbf{X}$ matrix are presented in Figure 5.8.

The wavelet of choice is the Wavelet Daubechies 4 (db4) which has extensively used in applications of the statistical signal characterization.

By applying the BIC penalized likelihood criterion to the family of HMM with the previously mentioned left-to-right structure, we conclude that data are optimally fit using hidden variables of 6 states.

Figure 5.9 presents the quantity BIC for various number of hidden states. By considering totally 6 different hidden states the transition modelling can been described by the Figure 5.10. We perform the EM algorithm 20 times, choosing the best of the trained models. For our signal the best trained model has the following transition matrix

$$\mathbf{A} = \begin{bmatrix} 0.981848 & 0.018152 & 0 & 0 & 0 & 0 \\ 0 & 0.988036 & 0.011964 & 0 & 0 & 0 \\ 0 & 0 & 0.988485 & 0.011515 & 0 & 0 \\ 0 & 0 & 0 & 0.981751 & 0.018249 & 0 \\ 0 & 0 & 0 & 0 & 0.988978 & 0.011022 \\ 0 & 0 & 0 & 0 & 0 & 1.0 \end{bmatrix} \tag{5.36}$$

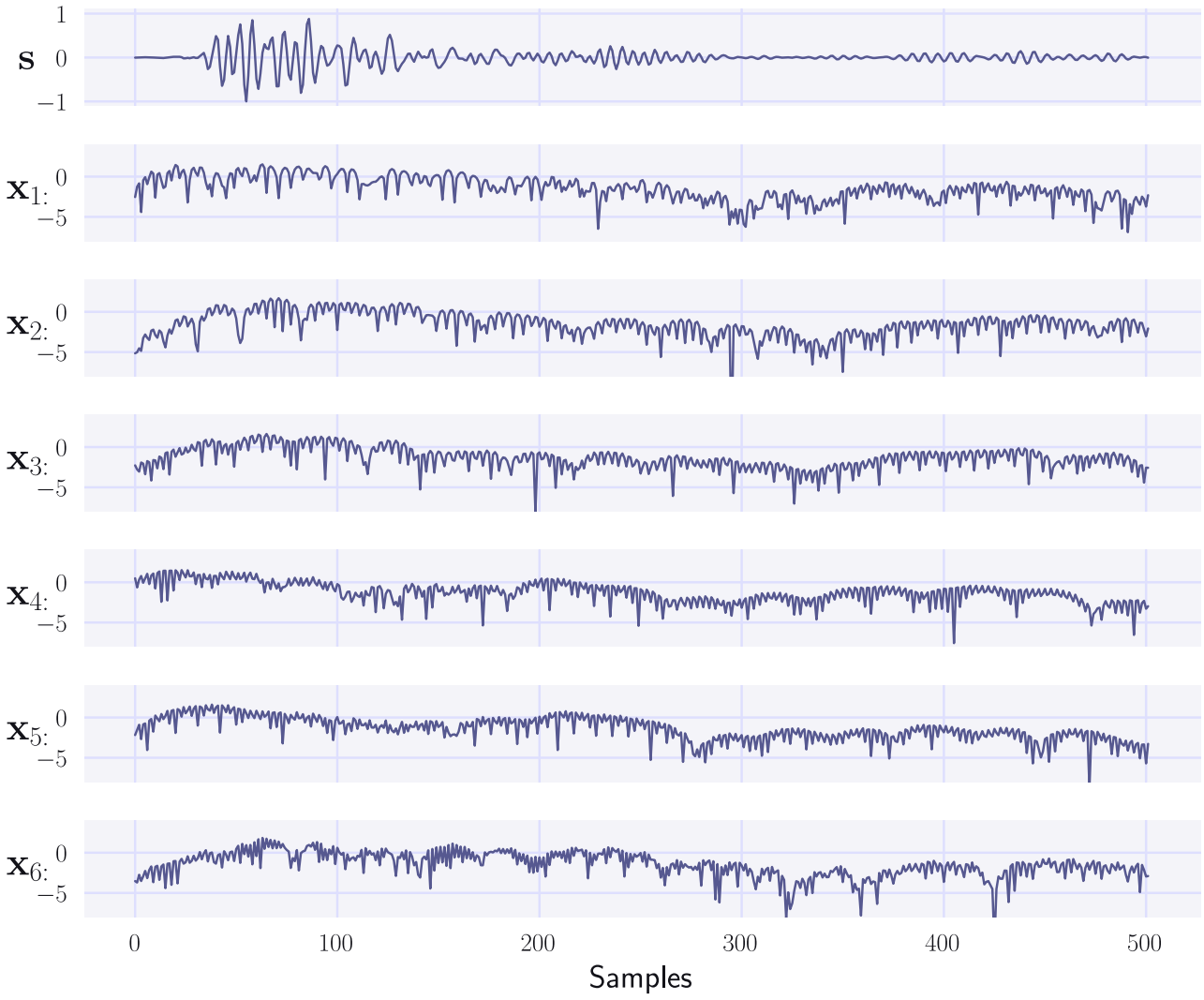


Figure 5.8: The cropped signal of the characterization example and the selected log-scale wavelet coefficients after standardization.

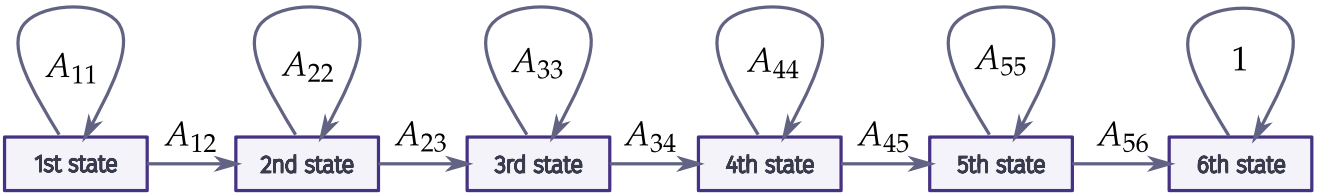


Figure 5.10: Transition modelling of the HMM of the signal's characterization example.

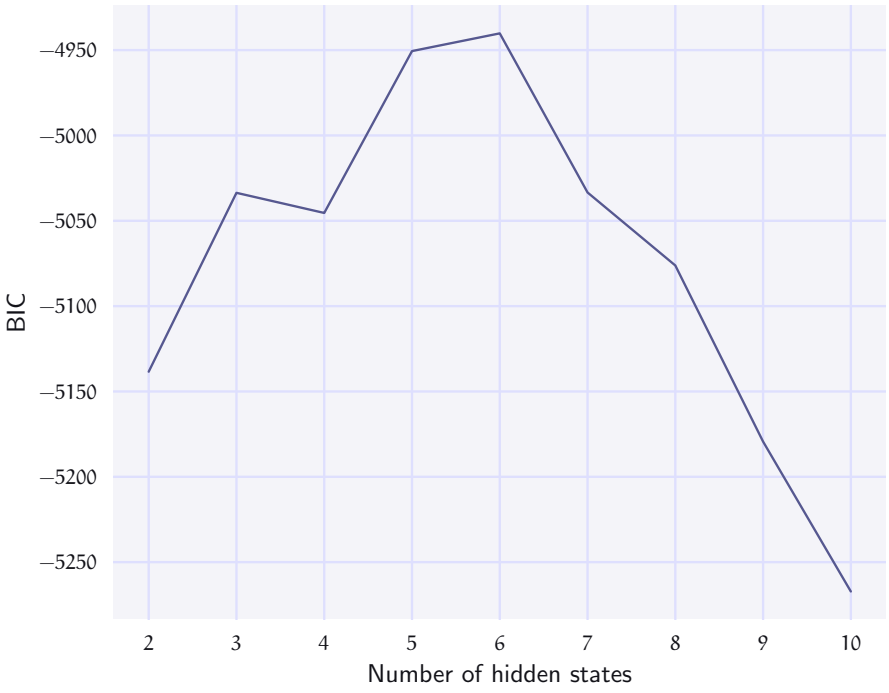


Figure 5.9: The BIC penalized likelihood criterion for selecting the number of the hidden states.

Furthermore, the mean values of the emission distributions are the following

$$\mu_1 = \begin{bmatrix} -0.289589 \\ -1.778478 \\ -0.810653 \\ 0.563156 \\ 0.198924 \\ -1.318611 \end{bmatrix}, \quad \mu_2 = \begin{bmatrix} 0.089679 \\ 0.274380 \\ 0.136299 \\ -0.782683 \\ -0.576287 \\ -0.089740 \end{bmatrix}, \quad \mu_3 = \begin{bmatrix} -0.820805 \\ -1.255032 \\ -1.655625 \\ -0.944006 \\ -0.761501 \\ -0.531937 \end{bmatrix},$$

$$\mu_4 = \begin{bmatrix} -1.898840 \\ -1.871181 \\ -1.792190 \\ -1.814542 \\ -1.434755 \\ -1.082241 \end{bmatrix}, \quad \mu_5 = \begin{bmatrix} -2.986100 \\ -3.393997 \\ -2.662216 \\ -2.078534 \\ -2.497090 \\ -3.137617 \end{bmatrix}, \quad \mu_6 = \begin{bmatrix} -2.206277 \\ -1.694613 \\ -1.423073 \\ -1.908468 \\ -2.471559 \\ -2.195566 \end{bmatrix} \tag{5.37}$$

Finally, in Figure 5.11 we present the covariance matrices of emission distributions of the fitted model. The next step is to extent our characterization scheme by providing a similarity measure between two trained models. Thereafter, by using the trained model with connection

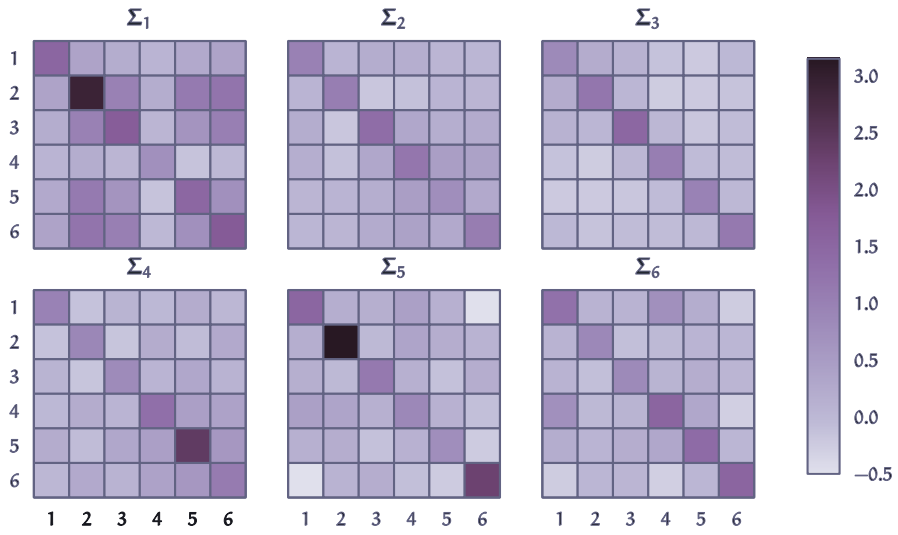


Figure 5.11: The covariance matrices.

with the observations $\mathbf{X}$, we apply the Viterbi algorithm (Section 4.5.2) to find the most probable values of the hidden variable sequence $\mathbf{z}_1, \ldots, \mathbf{z}_N$ which best explains the observations by means of the maximization of the conditional distribution $p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\lambda})$. Figure 5.12 indicates the retrieved state sequence of the example signal.

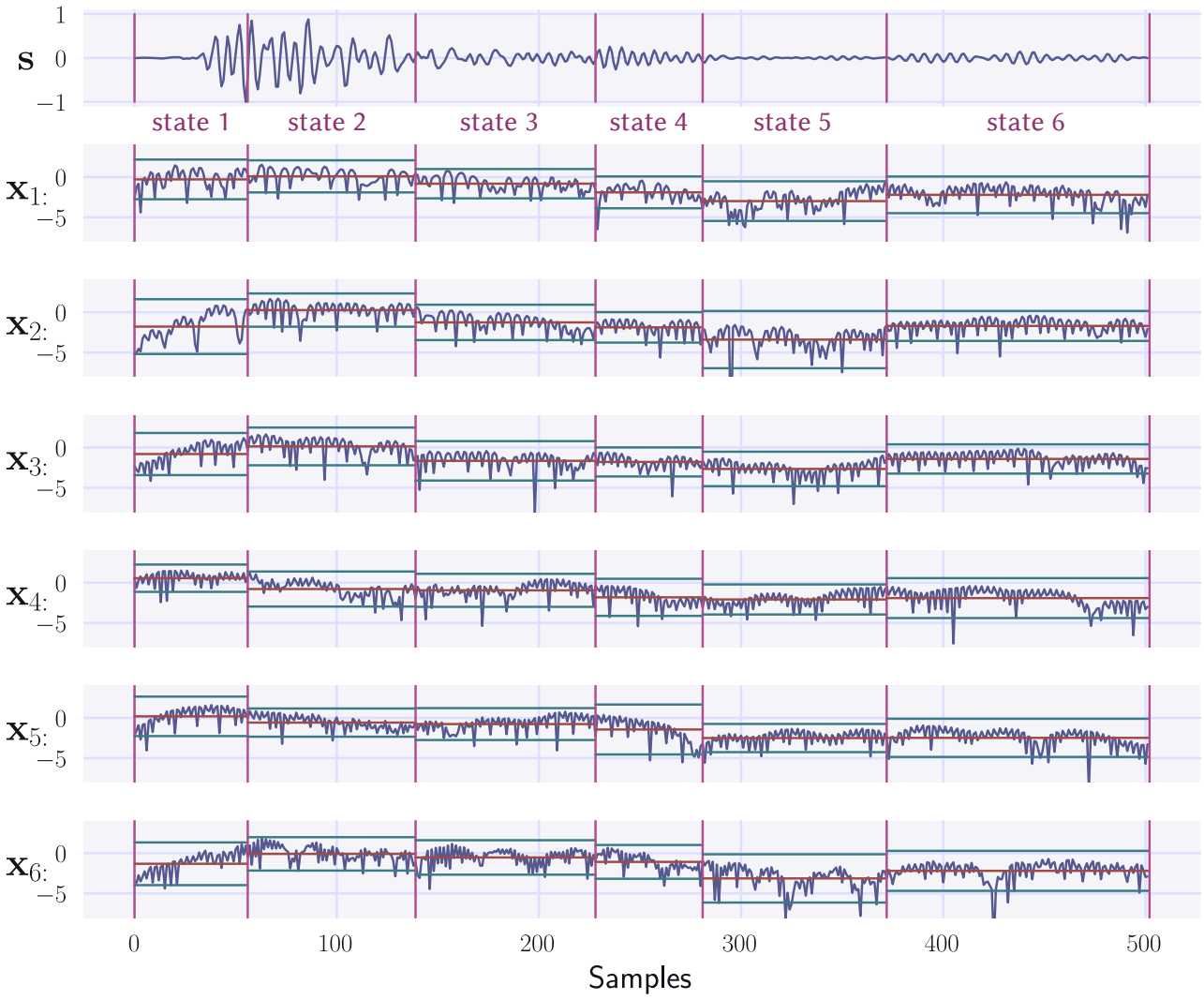


Figure 5.12: The most probable sequence (by using the Viterbi algorithm) of the hidden states for the signal of the classification example.

5.3.2 Similarity measurements

Given two acoustic signals $\mathbf{s}_1$ and $\mathbf{s}_2$ and their corresponding HMMs $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$, we wish to measure how similar these signals are by calculating the similarities measurements between their associated models, using the Kullback-Leibler .

$$\begin{aligned} D_s(\mathbf{s}_1, \mathbf{s}_2) \equiv \mathrm{KL}(\boldsymbol{\lambda}_1 \| \boldsymbol{\lambda}_2) &= \mathrm{KL}\Big(p(\mathbf{X}|\boldsymbol{\lambda}_1) \| p(\mathbf{X}|\boldsymbol{\lambda}_2) \Big) \\ &= \int p(\mathbf{X}|\boldsymbol{\lambda}_1) \log \frac{p(\mathbf{X}|\boldsymbol{\lambda}_1)}{p(\mathbf{X}|\boldsymbol{\lambda}_2)} d\mathbf{X} \\ &= \mathbb{E}_{\mathbf{X}|\lambda_1}\{\log p(\mathbf{X}|\boldsymbol{\lambda}_1) - \log p(\mathbf{X}|\boldsymbol{\lambda}_2)\}. \end{aligned} \tag{5.38}$$

Unfortunately, analytic evaluation of this integral is not possible. Therefore, we need a sufficient approximation of the KLD. An asymptotic approximation of the above integral is given by the following expression

$$D_s(\mathbf{s}_1, \mathbf{s}_2) \equiv \mathrm{KLD}(\boldsymbol{\lambda}_1 \| \boldsymbol{\lambda}_2) \approx \frac{1}{QN} \sum_{q=1}^{Q} [\log p(\mathbf{X}^{(q)} \| \boldsymbol{\lambda}_1) - \log p(\mathbf{X}^{(q)} \| \boldsymbol{\lambda}_2)]. \tag{5.39}$$

where $\mathbf{X}^{(q)}$, $q = 1, \ldots, Q$ form a sequence of observation, each one with N time steps, generated by model $\boldsymbol{\lambda}_1$. In this work we have fixed the number of generated sequences to be $Q = 2000$.

5.4 The Inversion Procedure

The applications presented in this work are associated with non-linear inverse problems of acoustical oceanography. The solution of these problems is formulated through optimization schemes aiming at the minimization of the difference between features of the measured signal and features of replica signals taken from an appropriate search space.

In this work we have applied statistical and probabilistic characterization schemes and the difference of the signals are obtained by means of the Kullback-Leibler Divergence, adopted to the appropriate characterization scheme.

5.4.1 Replica signals

Replica signals are obtained by applying a forward propagation model over a set $\mathcal{M}$ of model parameters describing the marine environment. A subset of these parameters consists the set $\mathcal{M}^{\mathrm{rec}}$ of the recoverable model parameters.

$$\mathcal{M} = \mathcal{M}^0 \cup \mathcal{M}^{\mathrm{rec}} \quad \text{with} \quad \mathcal{M}^0 \cap \mathcal{M}^{\mathrm{rec}} = \emptyset, \tag{5.40}$$

where $\mathcal{M}^0$ the set of the model parameters which are considered known.

Each model parameter is quantified by means of a real-valued variable m_j. Knowing the values of the set of the known model parameters $\mathcal{M}^0$, we can define different propagation models by considering various values for the recoverable model parameters. Let $\mathbf{m}$ the recoverable parameter vector containing‘ the M recoverable model parameters. Thus, this vector has the form:

RECOVERABLE PARAMETER VECTOR

$$\mathbf{m} = [m_1, m_2, \ldots, m_M]^T, \tag{5.41}$$

Throughout this work a normal mode propagation model is used to calculate the impulse response in a cylindrical coordinate system considering an axially symmetric environment,

$p_{\mathbf{m}}(r, z; z_s, \omega)$ for frequency ω at range r and depth z due to a point harmonic source at depth z_s. The acoustic signal at frequency ω when a source excitation function $S(\omega)$ is considered, is given by $p'_{\mathbf{m}}(r, z; z_s, \omega) = S(\omega) p_{\mathbf{m}}(r, z; z_s, \omega)$. By applying the inverse Fourier transform we obtain the replica signal $\mathbf{s}_r$ in the time domain.

The features of a replica signal are directly related to the recoverable parameters $\mathbf{m}$ using the two characterization methods considering in this thesis. Figure 5.13 shows the procedure for obtaining the features corresponding to a recoverable parameter vector $\mathbf{m}$, using the two signal characterization methods presented in this thesis. Nevertheless, in an actual inverse

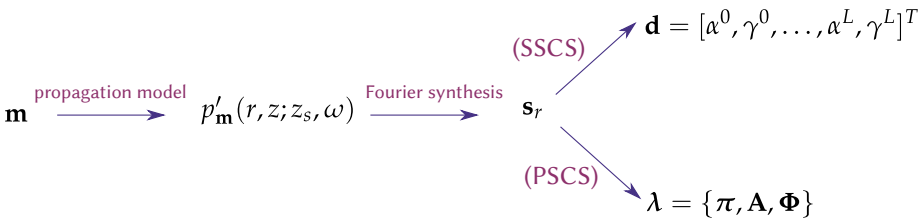


Figure 5.13: Flow diagram for obtaining the features (observables) of the replica signals in both characterization schemes.

problem, any prior information on the model parameters has to be used to reduce the number of the possible candidate combinations of the values of the recoverable parameters. Typically, each parameter is limited to some discrete values lying in a closed interval, forming a set of possible model vectors called the *search space* of the inverse problem.

5.4.2 Measured signals

In our inversion applications, we compare the features of the replica signals with the corresponding features of a measured signal. In real applications, these measured signals $\mathbf{s}_m$ are obtained by recording the acoustic field in hydrophones and they are contaminated with noise. Due to the fact that both the SSCS (Section 5.2) and PSCS (Section 5.3) characterization schemes require signals with a large signal-to-noise ratio (SNR), we apply the sparse denoising scheme (Section 5.1) to enhance the signal quality before a signal characterization procedure is applied. Note that the features of the measured signal are associated with same recoverable parameters of the replica signals. The clean signal is denoted as $\mathbf{s}_{mc}$

Some of the inversion test cases that we will present in the following chapter are associated with synthetic data. In these cases, in order to simulate the conditions of a real experiment, we form a vector $\mathbf{m}_t$ containing values of the recoverable model parameters which are considered as "true" of model parameters. By applying a forward propagation problem as in the case of replica signals, we obtain the "clear" signal corresponding to the true model parameters as $\mathbf{s}_{\mathbf{mt}}$.

In order to simulate actual conditions, we add Gaussian white noise to get the simulated raw signal $\tilde{\mathbf{s}}_{mt}$. Then we apply again the sparse denoising scheme to obtain a good approximation $\tilde{\mathbf{s}}_{mc}$ of the noise-free simulated signal $\mathbf{s}_{\mathbf{m}_t}$. This clean signal will be considered as the actual signal for the simulated experiment. Figure 5.14 shows the procedure for obtaining the signal $\mathbf{s}$ which will be used for calculating the similarities with any replica one.

Figure 5.14: Flow diagram for obtaining both the denoised actual the simulated signals.

(a) Actual data

$$\mathbf{s}_m \xrightarrow{\text{denoising}} \mathbf{s}_{mc}$$

(b) Synthetic data

$$\mathbf{m}_t \xrightarrow{\text{propagation model}} \mathbf{s}_{\mathbf{m}_t} \xrightarrow{\text{adding noise}} \tilde{\mathbf{s}}_{mt} \xrightarrow{\text{denoising}} \mathbf{s}_{mc}$$

5.4.3 Search space

The search space is defined according to the number and the type of the recoverable parameters in connection with any a-priori information of possible limited values of the recoverable parameters. We restrict each single parameter $m_j,\ j = 1, \ldots, M$ to a number of certain values in an interval $[\alpha_j, \beta_j] \subset \mathbb{R}$, where α_j and β_j are determined either by the physical limitations or the prior knowledge of the environment under consideration. In this dissertation, we consider for each recoverable parameter m_j, a power of two (2^{Q_j}) uniformly distributed values. The exponent Q_j is determined with respect to the desirable accuracy of m_j. Thus, the possible values of m_j are given as follows:

$$m_j^q = \alpha_j + q\,\frac{\beta_j - \alpha_j}{2^{Q_j} - 1},\ q = 0, \ldots, 2^{Q_j} - 1. \tag{5.42}$$

Furthermore, we use the above notation in order to refer to a specific candidate solution as

$$\mathbf{m}^{q_1, q_2, \ldots, q_M} \equiv [m_1^{q_1}, m_2^{q_2}, \ldots, m_M^{q_M}]^T. \tag{5.43}$$

The search space can now be described by the following set

SEARCH SPACE

$$\mathcal{A} = \{\mathbf{m}^{q_1, q_2, \ldots, q_M} \ :\ q_j = 0, \ldots, Q_j - 1\}. \tag{5.44}$$

It should be noted that the search space consists of totally $\#\mathcal{A}$ model vectors, where

$$\#\mathcal{A} = 2^{Q_1} 2^{Q_2} \cdots 2^{Q_M} = 2^{\sum_{j=1}^{M} Q_j}. \tag{5.45}$$

Finally, the "best" estimated solution of the inverse problem in respect to the search space $\mathcal{A}$ is simply given by solving the optimization problem

ESTIMATED SOLUTION

$$\mathbf{m}^{\text{est}} = \arg\min_{\mathbf{m} \in \mathcal{A}} D_s(\mathbf{s}, \mathbf{s}_{\mathbf{m}}). \tag{5.46}$$

where D_s a proper similarity measure of acoustic signals.

5.5 Optimization Method

5.5.1 Exhaustive search

The exhaustive search is the simplest optimization method for solving optimization problems over discrete search spaces. This method takes into account all possible candidate solutions defined by the search space and evaluates the corresponding values of the objective function chosen as the similarity measure. Adopting this optimization technique for our inversions requires the calculation of all replica signals that belong to a search space.

However, as the number of the recoverable environmental parameters is increased, the exhaustive search approach presented above ceases to be applicable. For example, if we have to recover 6 parameters among $2^7 = 128$ possible values for each parameter, we will have to evaluate a total of 2^{42} combinations. It is clear that in such a case, we would have to compute the synthetic acoustic field for 2^{42} simulated environments. The computation time would be last about **140000 T/C years**, where T is the expected time of calculating the acoustic field of one arbitrarily instance of the parameter vector $\mathbf{m}$ expressed in seconds, and C is the number of the available CPU cores.

Fortunately, a GA of 3 parallel populations, of 256 individual each, performing for 16 generations, can give a final set of candidate solutions ($3 * 256$ individuals) in about **3.4 T/C hours**. A statistical analysis of these candidate solutions will give the final estimation of the inverse problem.

5.5.2 Optimization using Genetic Algorithm (GA)

A GA is a means to perform the optimization scheme at a reasonable time. According to the procedure presented in Section 4.7 each parameter m_j is transformed to a binary vector $\mathbf{z}_j$:

CODING OF m_j

$$m_j \rightarrow \mathbf{z}_j, \tag{5.47}$$

The GA starts with a randomly chosen initial population (set) of K individuals (candidate solutions) each one consisting of a candidate vector $\mathbf{m}^{(k)}$. Each individual $\mathbf{m}^{(k)}$ is expressed by a expanded binary vector by appending the coding of $m_j^{(k)}$ as:

$$\mathbf{m}^{(k)} \rightarrow \mathbf{Z}^{(k)} = [\leftarrow \mathbf{z}_1^{(k)} \rightarrow \leftarrow \mathbf{z}_2^{(k)} \rightarrow \cdots \leftarrow \mathbf{z}_M^{(k)} \rightarrow]^T. \tag{5.48}$$

To avoid possible trapping to a local solution, a number of parallel populations $\{\mathrm{POP}_p^0\}_{p=1}^{\mathcal{P}}$ are considered in random and each one is evolving independently, by mimicking the mechanisms of natural selection where the fitness of individuals is quantified based on the similarity measure among the observables. Remember that in our case the KLD has been used as an appropriate measure between individuals characterized through statistical or probabilistic features. This evolutionary process is driven by a successive application of the following operation, for a predefined number of generations $\mathcal{G}$:

1. Reproduction (using tournament selection)
2. Crossover
3. Mutation
4. Hall of fame (Elite and Innovative individuals)

The top index in the notation of the population denotes the generation, therefore after the evolutionary process, the set $\{\mathrm{POP}_p^{\mathcal{G}}\}_{p=1}^{\mathcal{P}}$ will include estimations of the recoverable parameters.

The parameters of the GA algorithms that we have used in this thesis are presented in Table 6.3.

5.5.3 Conditional distributions over the final population of GA

The GA algorithm is terminated when the predefined number of generation has been reached. Due to the parallel evolution of independent populations after the GA termination we have $\mathcal{P}$ sets of candidate solutions. Each generation consists of a certain number of individuals , each one of which corresponding to a vector of recoverable parameters.

We will denote by POP the superset of the individuals belonging to each realization of GA, such as:

$$\mathrm{POP} = \{POP_p^{\mathcal{G}}\}_{p=1}^{\mathcal{P}}. \tag{5.49}$$

We denote the set POP by the term *final population* of the GA.

A simple choice to present the solution to the inverse problem is to pick up the vector nm corresponding to the best individual in the final population,. Although this choice has been used in the past, it suffers from the lack of confirmed convergence of the GA to this particular set. Instead we choose to present a qualitative analysis of the final population involving the whole set of the individuals.

Gerstoft [126] who has extensively used GA to similar inverse problems in oceanography has adopted the so called "a-posteriori distribution" as primitive form of statistical analysis of the final population. In our applications we adopted a different approach using GMM (Section 4.4.2) to present the qualitative information containing in the final population of the GA. We will refer to these distributions as *population distributions*.

All the Gaussian distributions are assumed with full covariance matrices to illustrate possible dependencies among the recoverable parameters.

$$p(\mathbf{m}|\mathrm{POP}) = \sum_{q=1}^{Q} \pi_q \mathcal{N}\Big(\mathbf{m}|\boldsymbol{\mu}_q(POP), \boldsymbol{\Sigma}_q(POP)\Big). \tag{5.50}$$

The marginal population distributions are given by evaluating the following integral

$$p(m_i|\mathrm{POP}) = \int_{\mathbb{R}^M} \delta(m_i' - m_i) p(\mathbf{m}'|\mathrm{POP}) d\mathbf{m}'. \tag{5.51}$$

Then, the considered "best" estimation of each parameter is chosen to be the one that maximizes its marginal population distribution:

$$m_i^{\mathrm{est}} = \arg\max_{m_i} p(m_i|\mathrm{POP}). \tag{5.52}$$

Moreover, it is useful to calculate the joint population densities

$$p(m_i, m_j|\mathrm{POP}) = \int_{\mathbb{R}^M} \delta(m_i' - m_i)\delta(m_j' - m_j) p(\mathbf{m}'|\mathrm{POP}) d\mathbf{m}'. \tag{5.53}$$

It should be noted that although the proposed statistical analysis which is applied to the GA results can describe quite well the statistics of the actual solution within the final population of the GA it can not be considered as an estimation of the actual posterior probability distributions (PPDs) of the model parameters.

In the next chapter we will present test cases of inverse problems to be solved either by using exhaustive search when we have only two recoverable parameters (one case) or by using optimization via GA and solution representation using GMMs, in more realistic inversion scenarios with three or more recoverable parameters.

chapter 6

APPLICATIONS

Abstract

In this chapter, we will present a few characteristic applications of inverse problems of acoustical oceanography in order to study the performance of the proposed schemes. The first three of the test cases are based on synthetic dated simulating measurements obtained in real environments. The last one uses real data obtained during the Shallow Water experiment 2006 (SW06). In all the cases, a Genetic Algorithm (GA) is responsible for solving the optimization problem associated with the inverse problem, and a Gaussian Mixture Model (GMM) is used for the statistical interpretation of the candidate solution produce the GA.

6.1 Hyper-Parameters of the Characterization and Inversion Schemes

All the applications presented here are controlled by the same set of hyper-parameters for the characterization and inversion schemes. Specifically, the selected SSCS parameters are given in Table 6.1 and those used in PSCS are given in Table 6.2.

Finally, the parameters which control the optimization process were selected following an initial analysis of the performance of the GA for the class of inverse problems considered in this thesis. The selected GA parameters appear in Figure 6.3 and they are the same for all the cases where a GA algorithm was used.

Parallel processing has been applied for the calculation of the replica field as well as the implementation of the HMM and the optimization procedure using GA. We used 40-50 CPU cores of the **Octopus** distributed computer cluster of FORTH. The typical execution time for the inversions was 2-5 hours for the SSCS scheme and 3-8 hours for the PSCS scheme.

Table 6.1: Parameters which control the SSCS characterization scheme.

SSCS Parameters	Values
Time-Frequency decomposition	Wavelet multiresolution analysis
Chosen wavelet / Decomposition levels	Daubechies 4 (db4) / 3
Statistical distributions	$S\alpha S$ distributions

Table 6.2: Parameters which control the PSCS characterization scheme.

PSCS Parameters	Values
Time-Frequency decomposition	Stationary wavelet packet transform
Chosen wavelet / Decomposition levels	Daubechies 4 (db4) / 4
Energy threshold	0.025
Transition modelling	Left-to-Right (Bakis)
Emission distributions	Gaussian distributions
Realization of the EM algorithm	20

Table 6.3: Parameters which control the optimization process.

GA parameters	Values
Parallel populations	3
Individuals per population	128
Generations	16
Tournament size	3
Crossover probability	0.8
Mutation probability	0.02
Elite population	4
Innovative population	8
Similarity parameter	0.1

6.2 Applications with Synthetic Data

6.2.1 Test case 1 : Inversions in a range independent swallow water environment using exhaustive search

In this section, the efficiency of the proposed probabilistic signal characterization scheme (PSCS) will be evaluated by applying the method in a simulated experiment associated with the range-independent environment described by Figure 6.1 and Table 6.4. The simulated environment is modelled as a water column with sound speed profile varying linearly with depth over a single semi-infinity substrate of constant properties. The sound speed profile in water is parameterized by the sound speed at depths $z = 0\ m$ and $z = h = 200\ m$. We consider as the "actual" sound speed profile a constant one, specifically $c_0 = c_h = 1500\ m/s$. Also, we assumed $c_b = 1620\ m/s$ and $\rho_b = 1650\ kg/m^3$ to be the sound speed and the density of the bottom, respectively. Sound source of modelled with a Gaussian excitation function is considered at 80 m and a single hydrophone is considered at a depth of 80 m and range 8 km. We wish to use the simulated (measured) acoustic signal for the retrieval of the environmental parameters.

We will present two inversion subcases. The first subcase (Case 1.a) is referred to the retrieval of the sound speed profile in the water, whereas the second subcase (Case 1.b) addresses the problem of estimating the properties of the substrate (Sound speed, density).

For comparison reasons, in both cases the corresponding inversion results from the statistical signal characterization scheme (SSCS) will also be presented.

Furthermore, in order to investigate the effect that the quality of the recorded signals has to the inversion results, we consider the following four alternative versions of the simulated recorded signal:

a) Noise free signal.

b) Noisy signal with signal to noise ratio (SNR) equals to 10 dB.

c) Noisy signal with signal to noise ratio (SNR) equals to 5 dB.

d) Denoised signal using the sparse denoising scheme to the noisy signal with SNR $= 5$ dB.

Figure 6.2 shows the signals under consideration. They have been produced using the procedure described in Figure 5.14. In particular, the normal mode program MODE1 developed

at FORTH has been used to obtain the signal in the frequency domain for the whole bandwidth of the source excitation function followed by an inverse fast Fourier transform to get the signals in the time domain.

Applying wavelet packet decomposition in four levels (see Figure 5.6) we get the energy contribution per wavelet coefficient vector of the fourth decomposition level as in Figure 6.3 , corresponding to the denoised version of the signal (Figure 6.2 (d)).

Furthermore, by applying the procedure described in Figure 6.4 we get the expected optimal number of the hidden states for the same signal using the BIC criterion.

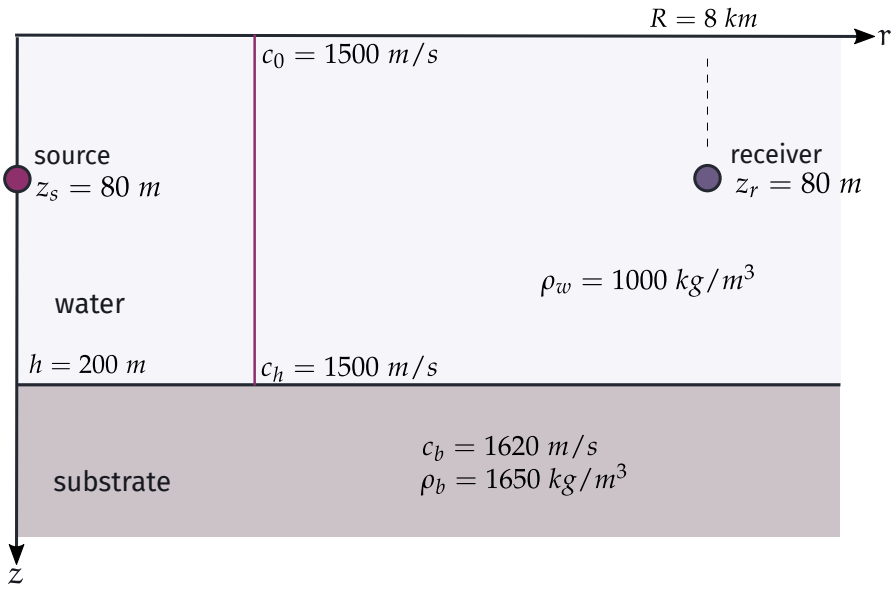


Figure 6.1: The simulated shallow water environment used in the inversion study.

Table 6.4: The parameters which describe the simulated shallow water environment

Parameters	Values
Water depth (h)	200 m
Range (R)	8 km
Central frequency (f_c)	100 Hz
Bandwidth (Δf)	40 Hz
Source/Receiver depth (z_s, z_r)	80 m
Sound speed in the water (c_w)	1500 m/s
Water density (ρ_w)	1000 kg/m^3
Sound speed in the substrate (c_b)	1620 m/s
Substrate density (ρ_b)	1650 kg/m^3

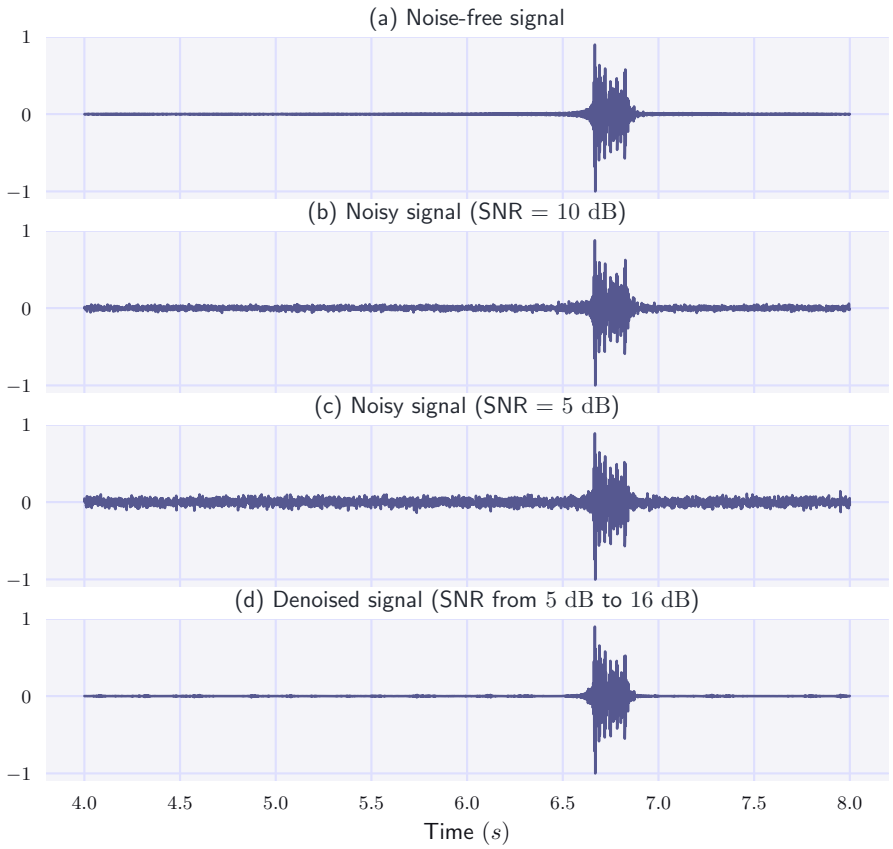


Figure 6.2: (a) The acoustic signal simulated in the shallow water environment described by Table 6.4, (b) A noisy version of the simulated signal with SNR = 10 dB, (c) A noisy version of the simulated signal with SNR = 5 dB, (d) A denoised version (SNR = 5 dB) of the signal with SNR = 5 dB.

— Case 1.a : Ocean acoustic tomography (single slice). —

The first subcase corresponds to a problem of ocean acoustic tomography at a single slice. We have assumed that the sound speed profile in the water is given by the following form:

$$c_w(z) = \frac{c_h - c_0}{h} z + c_0,\ z \in [0, h] \tag{6.1}$$

where $c_0 = c_w(0)$ and $c_h = c_w(h)$. We will assume that

$$c_0,\ c_h \in \{1485, 1486, \ldots, 1515\}\ m/s. \tag{6.2}$$

By considering all possible pairs of these two parameters, a dataset with a total of $31^2 = 961$ acoustic signals in the frequency domain is obtained. This exhaustive search was possible due to the fact that two only model parameters were considered in the inversion.

The replica fields at a single frequency are calculated using the normal mode program MODE1 developed at FORTH. Data obtained by this program is input to the inverse fast Fourier transform to get the signals in the time domain when a source with a Gaussian excitation function of central frequency 100 Hz and bandwidth 40 Hz.

Figure 6.3: Energy contribution per wavelet packet coefficient band. Threshold chosen: 0.025.

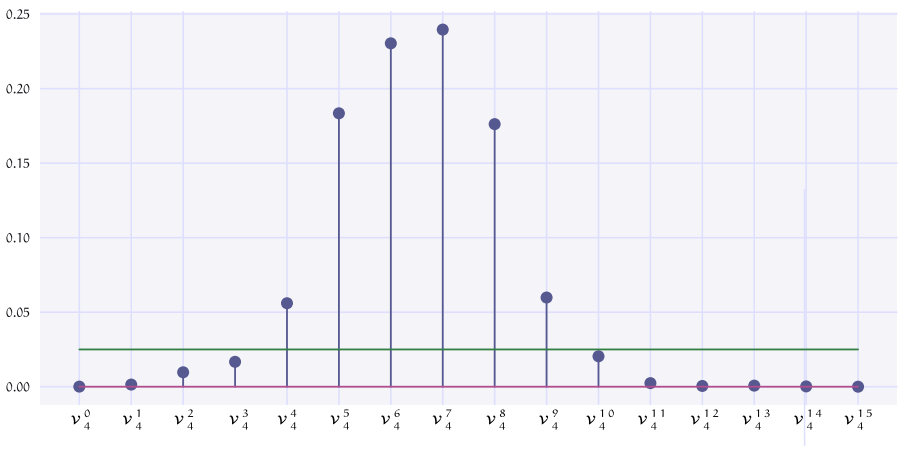


Figure 6.4: The BIC penalized likelihood criterion for selecting the number of the hidden states.

Table 6.5 gives the best solution obtained by both characterization schemes for each signal. The best solution is considered the one giving the minimum Kullback-Leibler divergence (KLD) between the "actual" and the "replica" signals both characterized by PSCS and SSCS methods. More information is given by Figure 6.5 giving the Kullback-Leibler divergence between the characteristics of the measured signal and the possible combinations of the recoverable parameters within the search space.

Analyzing the results as presented in Figure 6.5, we observe the following

- When noise-free signal is consider both methods return the exact values for the sound speed profile.

	SSCS		PSCS	
Test case	$c_0\ (m/s)$	$c_h\ (m/s)$	$c_0\ (m/s)$	$c_h\ (m/s)$
Noise-free signal	1500	1500	1500	1500
Noisy signal ($SNR = 10\ dB$)	1488	1511	1500	1499
Noisy signal ($SNR = 5\ dB$)	1486	1493	1499	1497
Denoised signal	1495	1490	1500	1500

Table 6.5: Inversion results for the ocean tomography case.

- The PSCS method gives exact results for the sound speed profile in the denoised signal case.
- The estimations of the sound speed profile with noisy signals are, as expected, inferior to those of the denoised signals in both cases.
- Overall the new proposed PSCS scheme gives in this case a better estimation of the sound speed profile independently of the noise level.

Also, it is interested to note that even in this simple example the KLD divergence between the simulated signal and the replica signals presents too many local minima, making the procedure to solve the inverse problem, complicated.

— Case 1.b : Sea-bed inversion —

The next test case is referred to the estimation of the sound speed and density of the sea-bed using a similar approach as before.

Again, We calculate the acoustic field using all the possible combination of the sound speed and density of the bottom subject to the set bellow:

$$c_b \in \{1550, 1555, \ldots, 1750\}, \tag{6.3}$$

$$\rho_b \in \{1400, 1410, \ldots, 1800\}. \tag{6.4}$$

These sets form 1681 candidate solutions. The best solution is considered the one giving the minimum Kullback-Leibler divergence (KLD) between the "actual" and the "replica" signals both characterized by PSCS and SSCS methods. Figure 6.6 presents the KLD between the characteristics of "simulated" and "replica" signals. Additionally, the best estimated solutions are presented in Table 6.6.

- When noise-free signal is considered both methods return the exact values of the bottom properties.
- The PSCS method gives exact results also for the denoised signal case.
- In accordance with the previous sub-case the new proposed PSCS scheme gives in general better estimations of the sea-bed parameters with comparison with SSCS.

In the subsequent subsections we will present more complicated cases.

	SSCS		PSCS	
Test case	$c_b\ (m/s)$	$\rho_b\ (kg/m^3)$	$c_b\ (m/s)$	$\rho_b\ (kg/m^3)$
Noise-free signal	1620	1650	1620	1650
Noisy signal ($SNR = 10\ dB$)	1575	1760	1620	1570
Noisy signal ($SNR = 5\ dB$)	1630	1610	1620	1650
Denoised signal	1625	1660	1620	1650

Table 6.6: Inversion results for the case of seabed recovery.

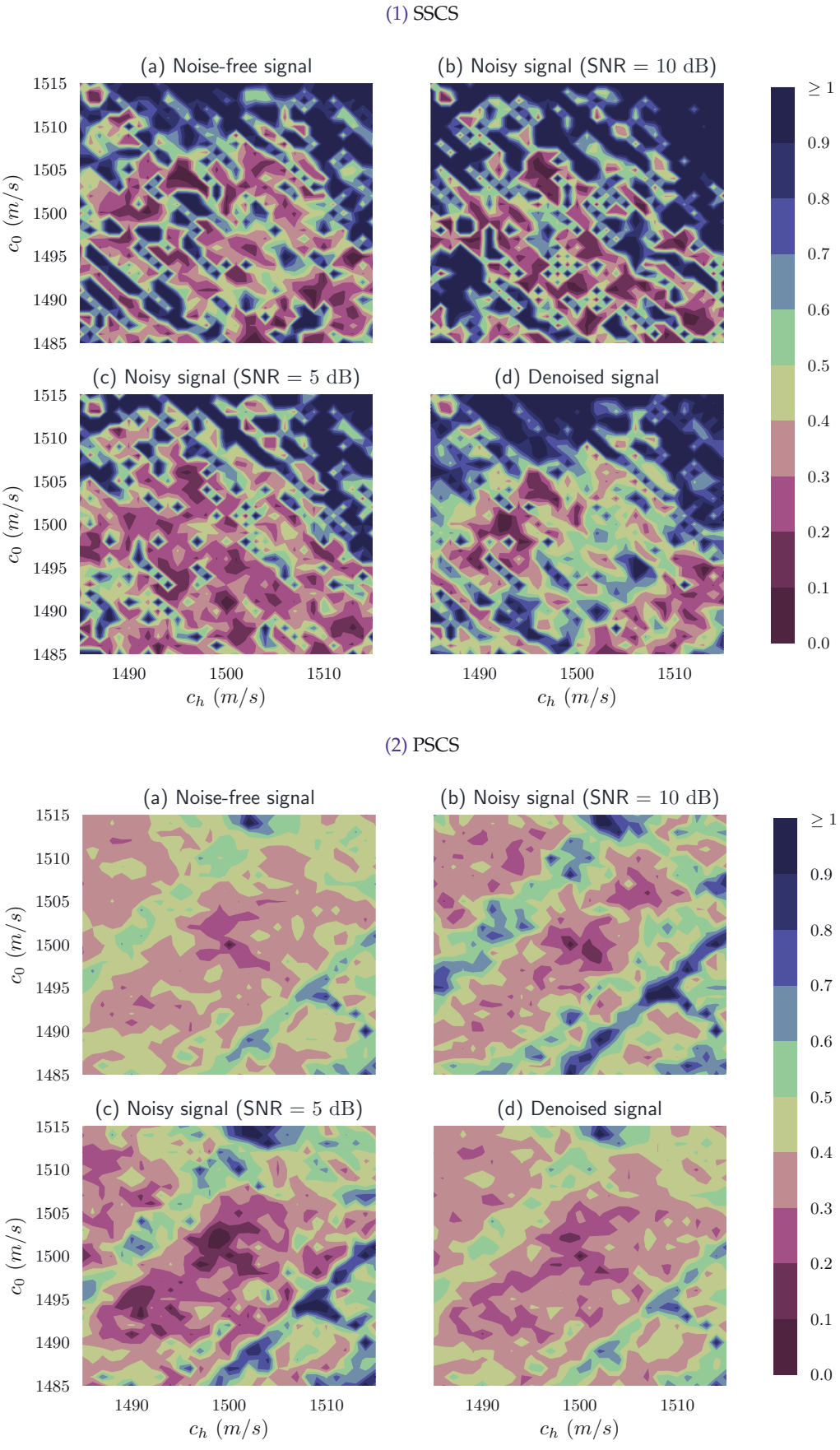


Figure 6.5: KLD measurements for the water column inversion example. Each contour was normalized independently.

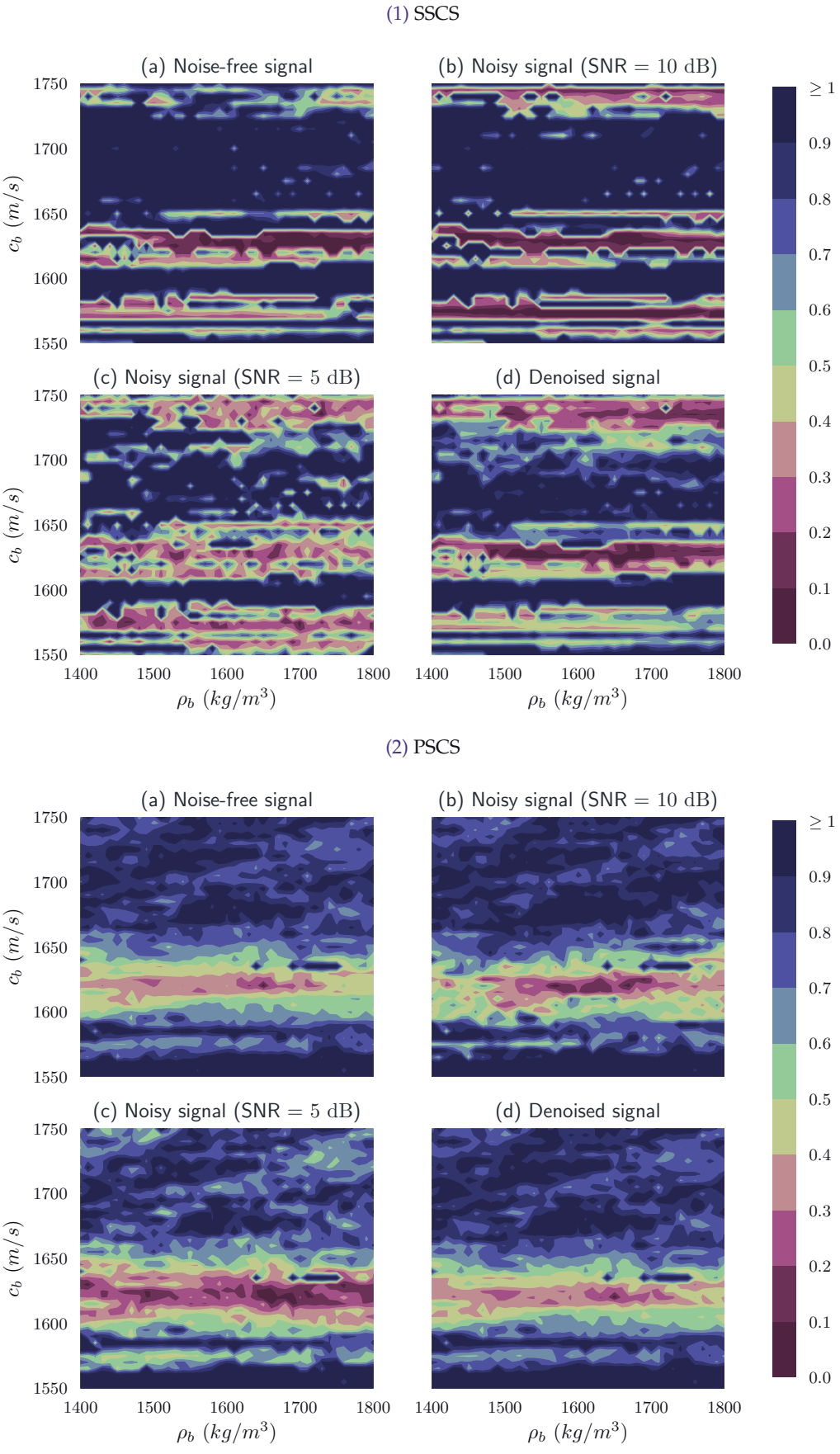


Figure 6.6: KLD measurements for the bottom inversion example. Each contour was normalized independently.

6.2.2 Test case 2 : Inversions in a range independent swallow water environment using GA

The second simulated environment corresponds to a range independent ocean channel described by Figure 6.7. Table 6.7 presents the environmental and operational parameters of the simulated experiment. In this environment a single receiver is considered to be 8 km away from a source with a Gaussian excitation function of central frequency 150 Hz and bandwidth 40 Hz.

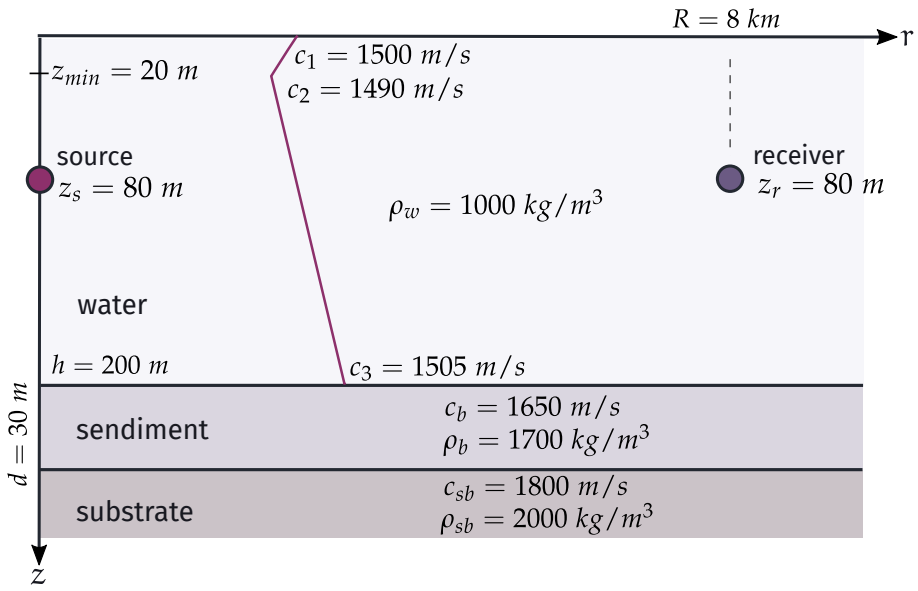


Figure 6.7: The simulated sea environment of the second test case.

Three versions of the simulated acoustic signal appear in Figure 6.8. The first one corresponds to the signal s_{mt}. The second signal corresponds to the $\tilde{\mathbf{s}}_{mt}$ when we have added noise with $SNR = 5$dB and the last one corresponds to the clean (denoised) signal $\mathbf{s}_{mc}$ where the sparse denoising scheme has been applied. We shall use the last, denoised signal to invert the recoverable parameters described bellow.

In this case, the sound speed profile is considered to be bi-linear modelled by the following relation:

$$c_w(z) = \begin{cases} c_1 + \frac{c_2 - c_1}{z_{min}} z, & 0 \leq z \leq z_{min}, \\ c_2 + \frac{c_3 - c_2}{h - z_{min}} (z - z_{min}), & z_{min} < z \leq h. \end{cases} \tag{6.5}$$

We will consider the following subcases

Table 6.7: Environmental and operational parameters describing the simulated shallow water environment of the second test case. The search space of the recoverable parameters are provided as well.

Parameters	Values	Search space
Water depth (h)	200 m	-
Range (R)	8000 m	$[7800, 8300]$ (m)
Central frequency (f_c)	150 Hz	-
Bandwidth (Δf)	40 Hz	-
Source/Receiver depth $(z_s,\ z_r)$	80 m	$[70, 90]$ (m)
Sound speed at $z = 0$ (c_1)	1500 m/s	$[1490, 1530]$ (m/s)
Depth of minimum sound speed (z_{min})	20 m	$[5, 30]$ (m)
Minimum sound speed (c_2)	1490 m/s	$[1470, 1510]$ (m/s)
Sound speed at depth h (c_3)	1505 m/s	$[1490, 1530]$ (m/s)
Water density (ρ_w)	1000 kg/m^3	-
Sediment thickness (d)	30 m	$[5, 30]$ (m)
Sound speed in the sediment (c_b)	1650 m/s	$[1500, 1800]$ (m/s)
Sediment density (ρ_b)	1700 kg/m^3	$[1400, 2000]$ (kg/m^3)
Sound speed in the substrate (c_{sb})	1800 m/s	$[1600, 2100]$ (m/s)
Substrate density (ρ_{sb})	1650 kg/m^3	$[1500, 2200]$ (kg/m^3)

- Estimation of source and receiver locations. The recoverable parameters are the source and receiver depths as well as the range. All the other model parameters are considered known.
- Recovery of the sound speed profile in water column. All the other model parameters are considered known.
- Recovery of the sea-bed parameters. Again, all the other model parameters are considered known.
- Recovery of all the recoverable parameters. In this case we intent to test the methods in a multiparametric inverse problem.

We use the same procedure as the previous case to extract the time-frequency characteristics based on the stationary wavelet decomposition of the signal, keeping the energy threshold 0.025 (Figure 6.9) and eventually to estimate the optimal number of hidden states (Figure 6.10).

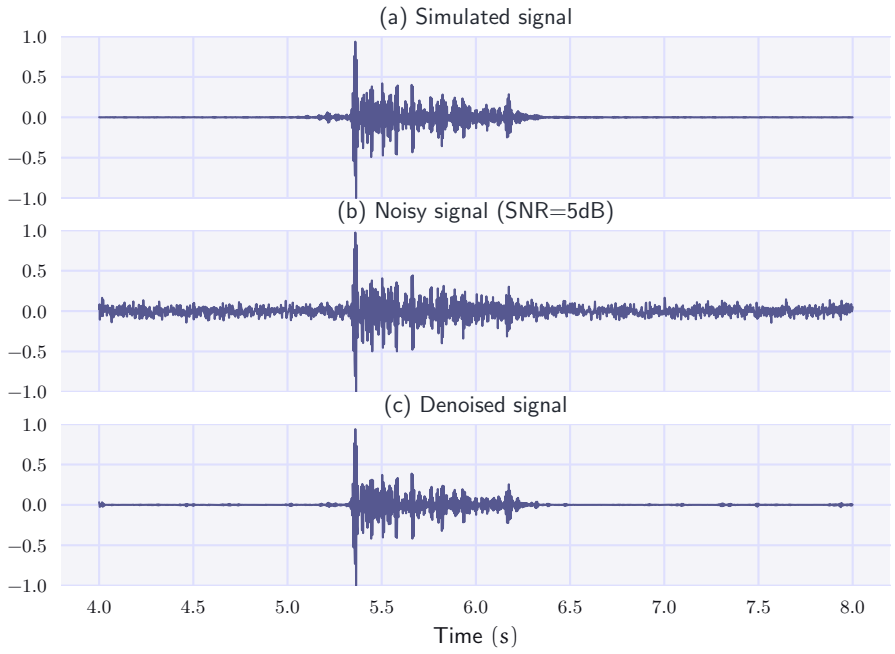


Figure 6.8: Three versions of the simulated signal corresponding to the second inversion test case.

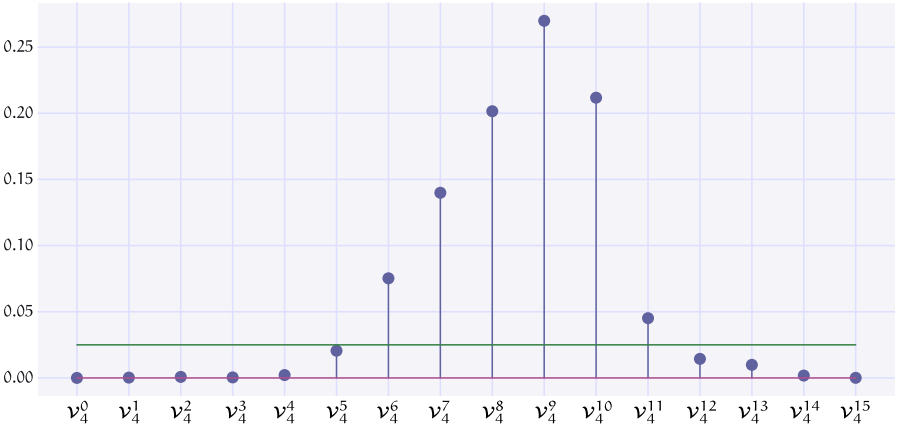


Figure 6.9: Energy contribution per wavelet packet coefficient band of the signal of the second test case. Threshold has been picked 0.025.

The inverse problem was solved considering three parallel populations of a Genetic Algorithm (GA). Details relative to the selected parameters of this algorithm have been provided in Table 6.3.

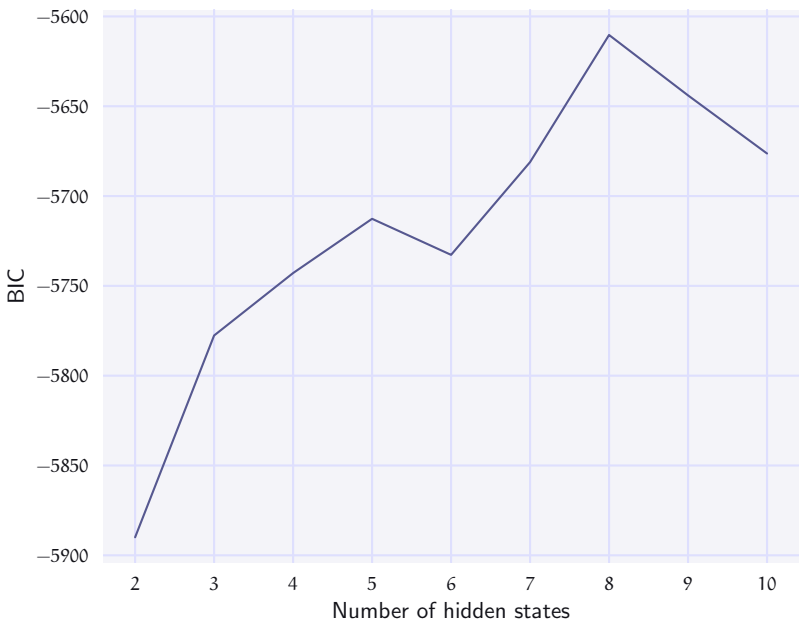


Figure 6.10: The BIC penalized likelihood criterion for selecting the number of the hidden states corresponding to the second test case.

— Case 2.a : Estimation of source and receiver locations —

In this test case, we invert only for the parameters which define the position of source and receiver. We define the model parameter vector $\mathbf{m}$ as:

$$\mathbf{m} = [R, z_s, z_r]^T. \tag{6.6}$$

Note that, any other parameter is taken at its actual value. Again, both the SSCS and PSCS schemes are applied. Figures 6.11 and 6.12 present the marginal population density functions of the final population POP of the GA, produced using the SSCS and PSCS schemes. Moreover, Figures 6.13 and 6.14 illustrate the corresponding joint distributions for each pair of the recoverable parameters. Finally, Table 6.8 summarizes the best estimations using both schemes defined as the values of the maximum of the marginal population distributions.

Table 6.8: Recovered model parameters corresponding to case 2.a. In parentheses the mismatch between the estimated and actual value of each parameter is provided.

Parameters	SSCS	PSCS
Range R (m)	8003.7 $(+3.7)$	7999.2 (-0.8)
Source depth z_s (m)	79.9 (-0.1)	79.7 (-0.3)
Receiver depth z_r (m)	79.6 (-0.4)	80.1 $(+0.1)$

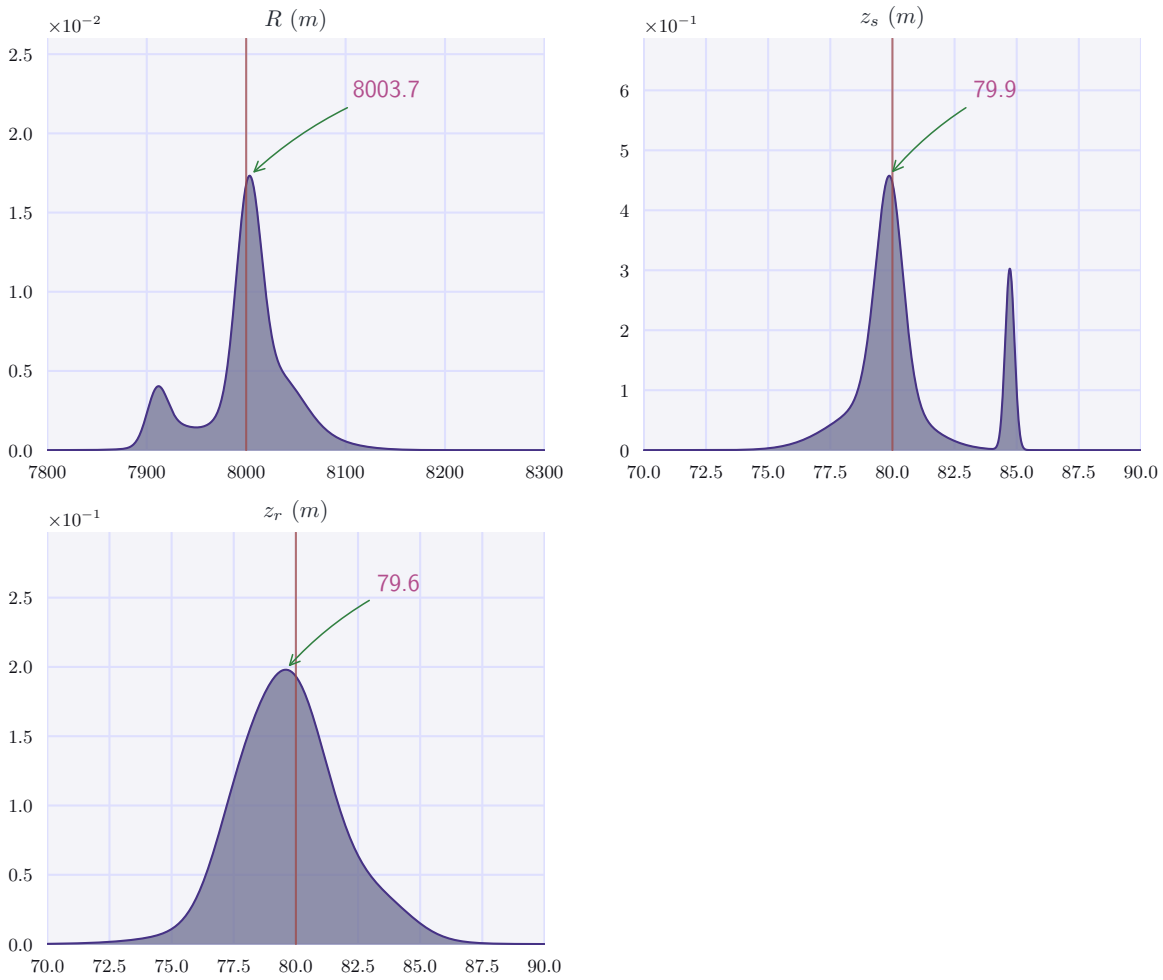


Figure 6.11: Marginal population distributions of the recoverable environmental and operational parameters of the case 2.a using **SSCS**. The distributions show statistics of the final population of GA. The vertical lines state for the actual values, whereas the arrows point to the maxima of the marginal densities.

While both schemes gave similar good estimations of the recoverable parameters, the marginal population distributions corresponding to the PSCS scheme gave narrower confidence intervals. The biggest difference of the variance occurred in the recovered depth of the receiver z_r. In addition, both schemes illustrate a false peak for the population distributions for the depth of the source z_s at about 85 m. However, the magnitude of this false peak is significantly lower in PSCS.

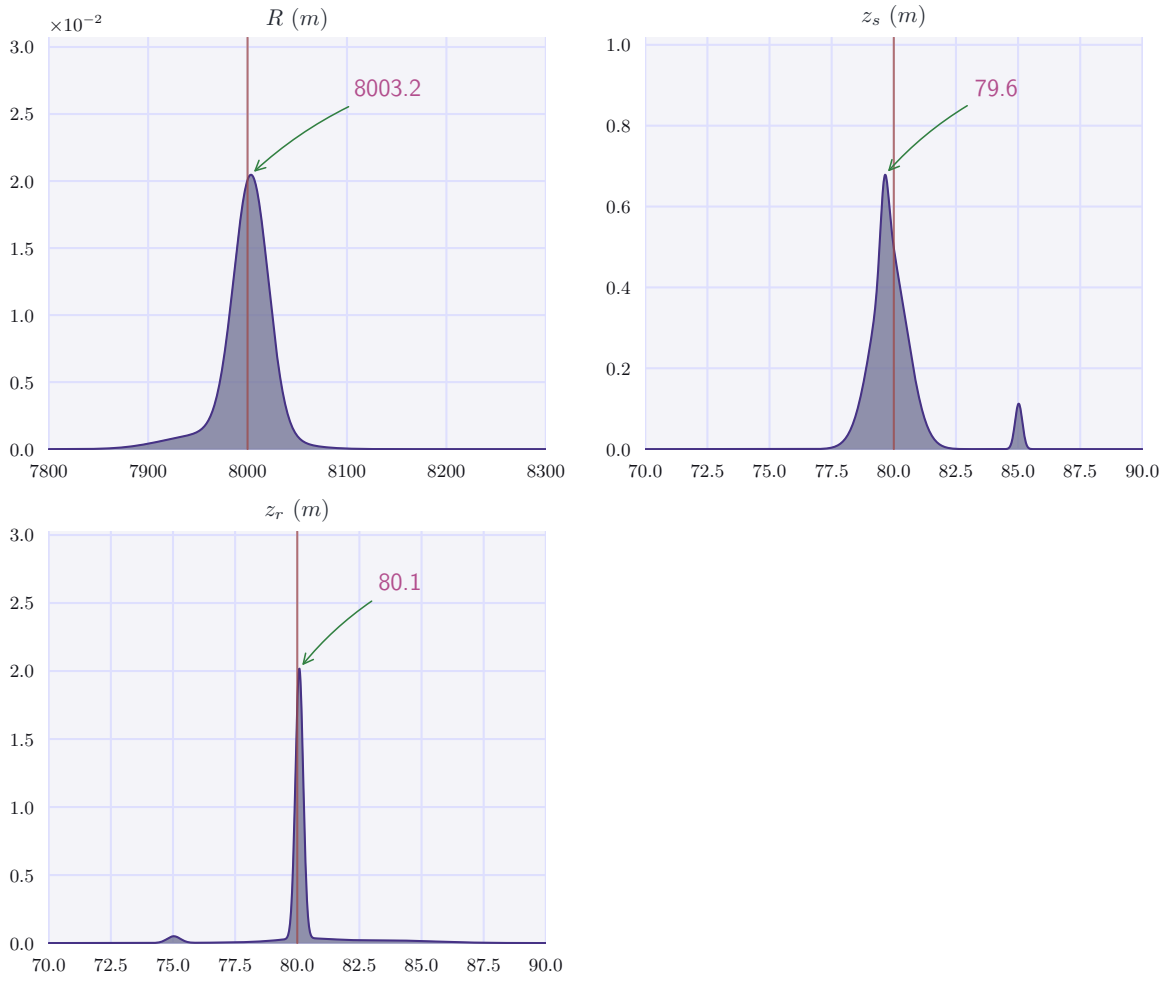


Figure 6.12: Marginal population distributions of the recoverable environmental and operational parameters of the case 2.a using **PSCS**. The distributions show statistics of the final population of GA. The vertical lines state for the actual values, whereas the arrows point to the maxima of the marginal densities.

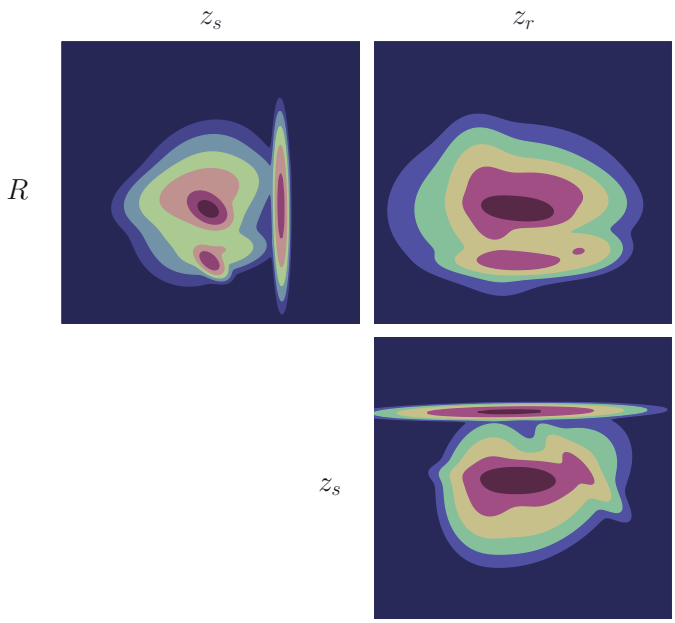


Figure 6.13: Joint population distributions corresponding to the inversion results of the test case 2.a, using **SSCS**. The distributions show statistics of the final population of GA. Each distribution has been normalized independently.

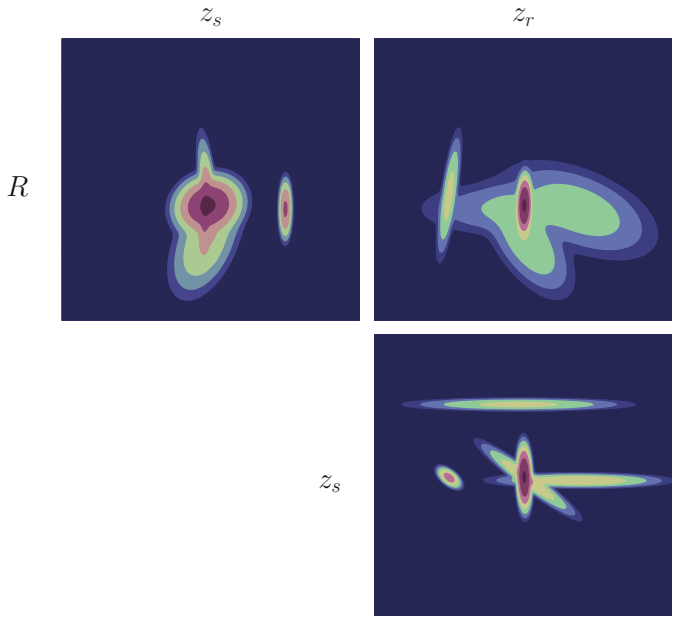


Figure 6.14: Joint population distributions corresponding to the inversion results of the test case 2.a, using **PSCS**. The distributions show statistics of the final population of GA. Each distribution has been normalized independently.

— Case 2.b : Recovering the sound speed profile in water column —

To simulate the ocean tomography problem at a single slice, we will invert for the values which define the sound speed profile in the water column. Likewise the previous case, we have to preset all the other model parameters. Therefore, we consider that the geoacoustic parameter and operational parameters are fixed to their real values.

The recoverable parameters are expressed by the following model vector:

$$\mathbf{m} = [z_{\text{min}}, c_1, c_2, c_3]^T, \tag{6.7}$$

where c_1 is the sound speed at the top of the water column, c_2 is the water sound speed at depth z_{min}, while c_3 denotes the water sound speed at the water-sediment surface. Search space for these parameters is given in Table 6.7.

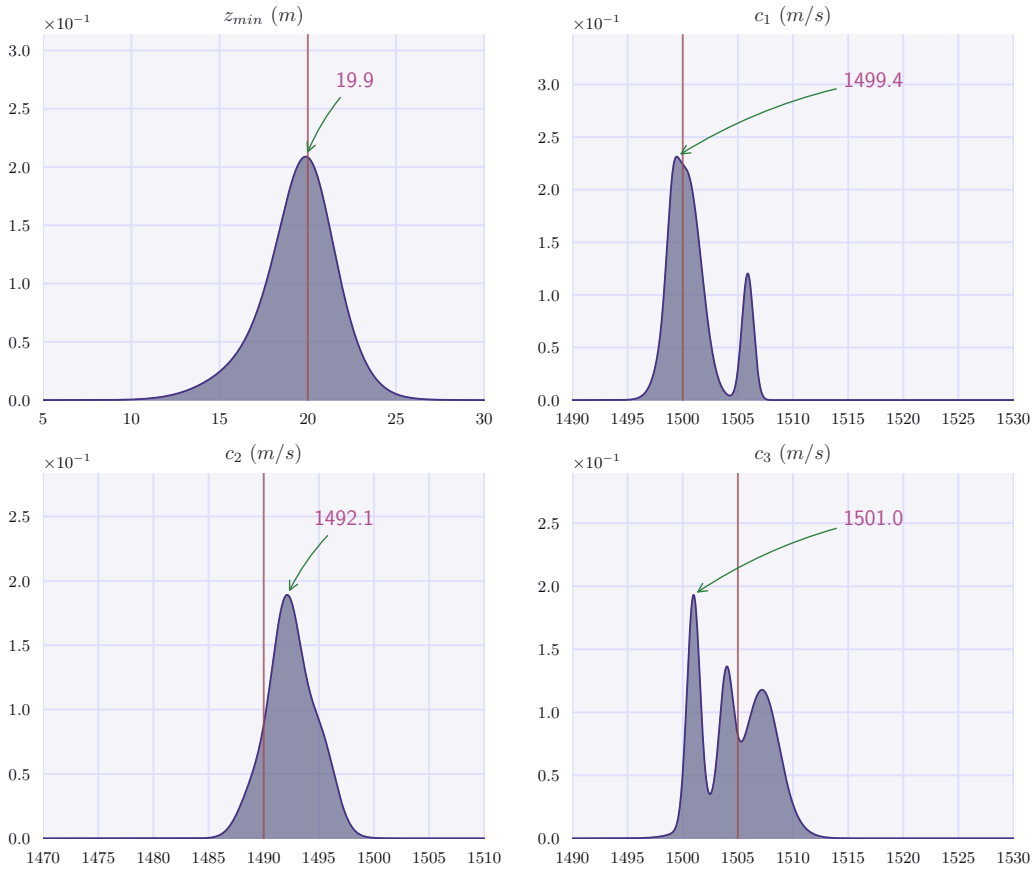


Figure 6.15: Marginal population distributions of the recoverable environmental and operational parameters of the test case 2.b using **SSCS**. The distribution show statistics of the final population of GA. The vertical lines state for the actual values, whereas the arrows point to the maxima of the marginal densities.

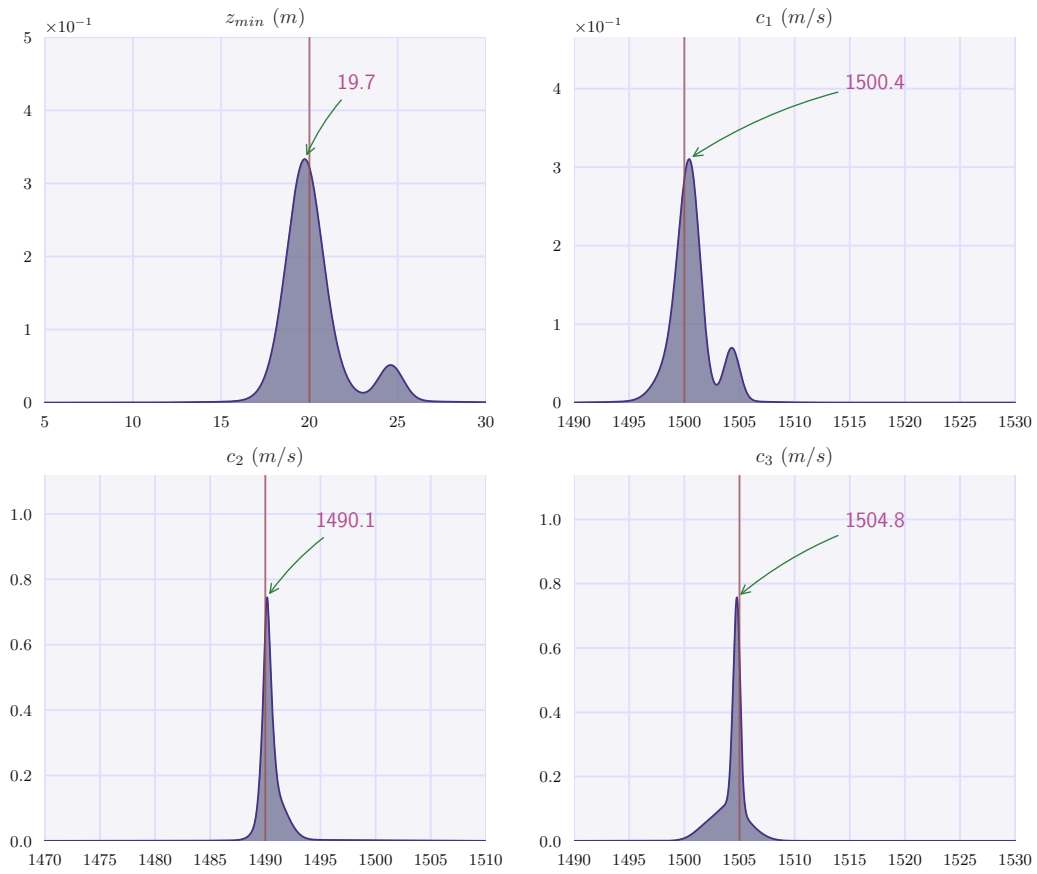


Figure 6.16: Marginal population distributions of the recoverable environmental and operational parameters of the test case 2.b using **PSCS**. The distributions show statistics of the final population of GA. The vertical lines state for the actual values, whereas the arrows point to the maxima of the marginal densities.

The inversion results for two schemes, are provided by means of marginal and joint probabilities densities Figures 6.15 to 6.18. Also, the "best" values appear in Table 6.9. While Figure 6.19 presents the actual and the recovered sound speed profiles. Both characterized led to a reconstructed sound speed profile close enough to the actual one. The Again the PSCS seems to be superior to as regards both the maxima of marginal densities end are spread. The spread of the densities of the PSCS scheme are much narrower of those of SSCS, which is a indication that the confidence intervals of PSCS are lower than SSCS.

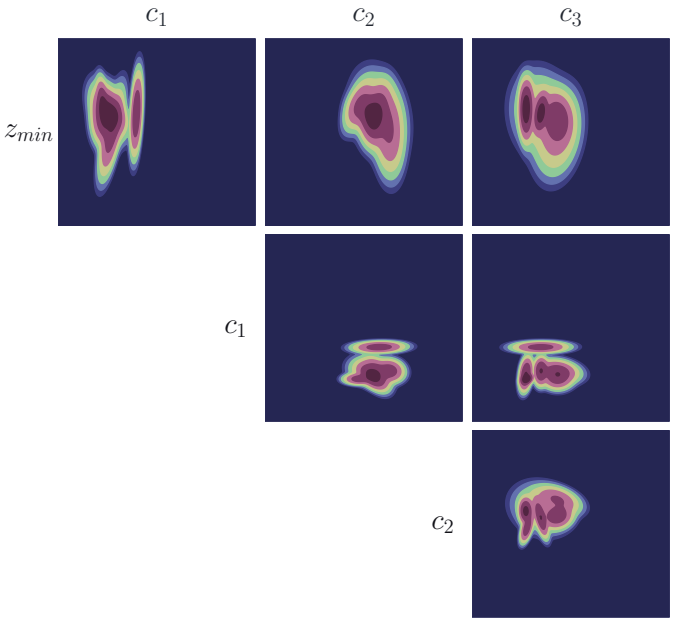


Figure 6.17: Joint population distributions corresponding to the inversion results of the test case 2.b using **SSCS**. The distributions show statistics of the final population of GA. Each distribution has been normalized independently.

Parameters	Units	SSCS	PSCS
Depth of minimum sound speed	z_{min} (m)	19.9 (-0.1)	19.7 (-0.3)
Sound speed at $z = 0$	c_1 (m/s)	1499.4 (-0.6)	1500.4 $(+0.4)$
Minimum sound speed	c_2 (m/s)	1492.1 $(+2.1)$	1490.1 $(+0.1)$
Sound speed at depth h	c_3 (m/s)	1501.0 (-4.0)	1504.8 (-0.2)

Table 6.9: Recovered model parameters corresponding to case 2.b. In parentheses the mismatch between the estimated and actual value of each parameter is given.

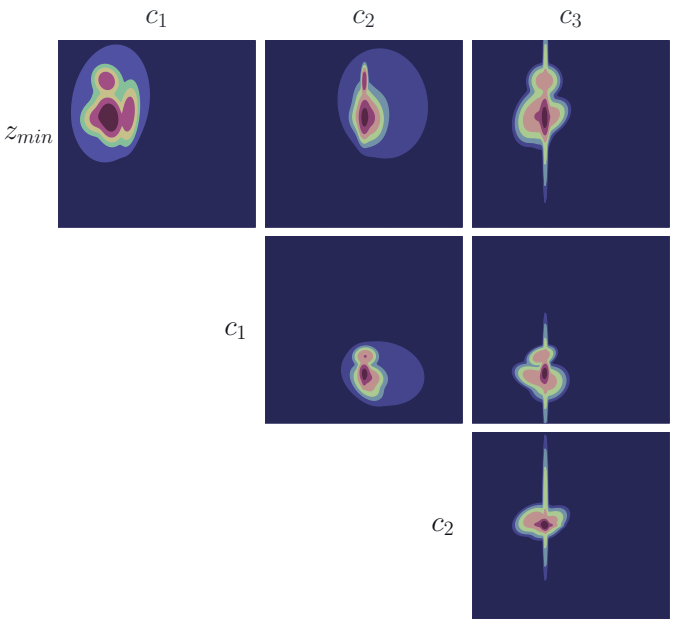


Figure 6.18: Joint population distributions corresponding to the inversion results of the test case 2.b using **PSCS**. The distributions show statistics of the final population of GA. Each distribution has been normalized independently.

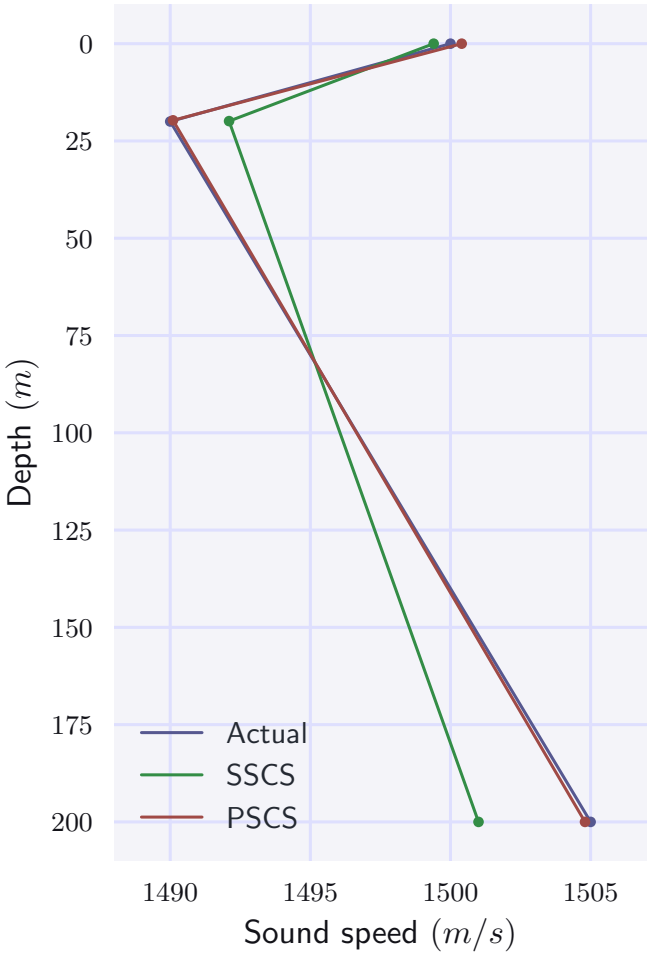


Figure 6.19: Actual and recovered sound speed profiles in water column using both SSCS and PSCS characterization methods.

— Case 2.c : Recovering the sea-bed parameters —

In this case, we will recover the properties of the seabed. So, the recoverable parameters are described by the following model vector:

$$\mathbf{m} = [d, c_b, \rho_b, c_{sb}, \rho_{sb}]^T. \tag{6.8}$$

After performing signal characterization by the two inversion schemes followed by inversion procedure using the GA we get the results appearing in Figures 6.20-6.23. Furthermore, Table 6.10 presents the inversion results using the two alternative inversion approaches.

Analyzing the results we observe the following:

- The PSCS characterization method gives results which indicate a beter estimation of the actual recoverable values with respect to SSCS methods.
- The spread of the marginal densities for the sediment thickness and the sound speed for the two layers, using the PSCS is narrow, much narrower with respect to the SSCS case.
- The spread of the marginal densities for the densities of two sea-bed layers based on the two characterization schemes is wide. This was expected, as it is well known, the densities have less influence on the acoustic field in the water column in relation to the influence of the sound speed in these layers.
- Overall, the inversion results using both signal characterization schemes are considered satisfactory for the problem in hand.

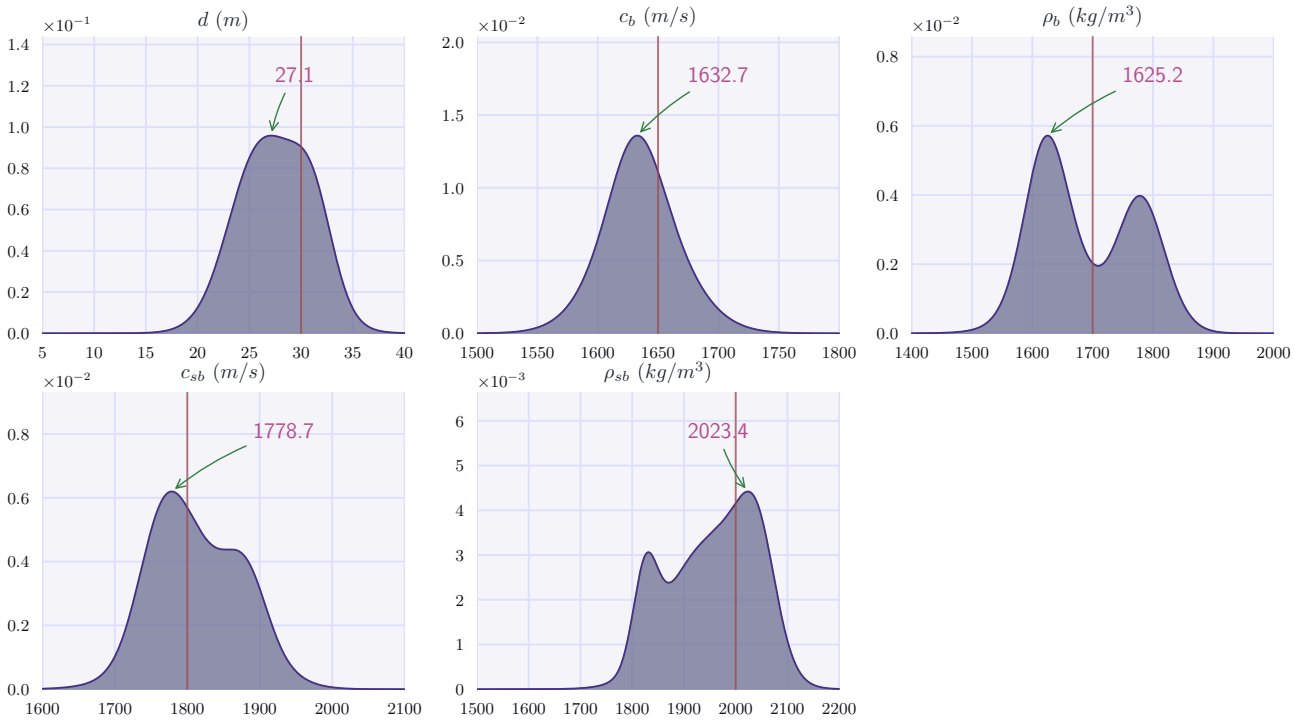


Figure 6.20: Marginal population distributions of the recoverable environmental and operational parameters of the case 2.c using **SSCS**. The distributions show statistics of the final population of GA. The vertical lines state for the actual values, whereas the arrows point to the maxima of the marginal densities.

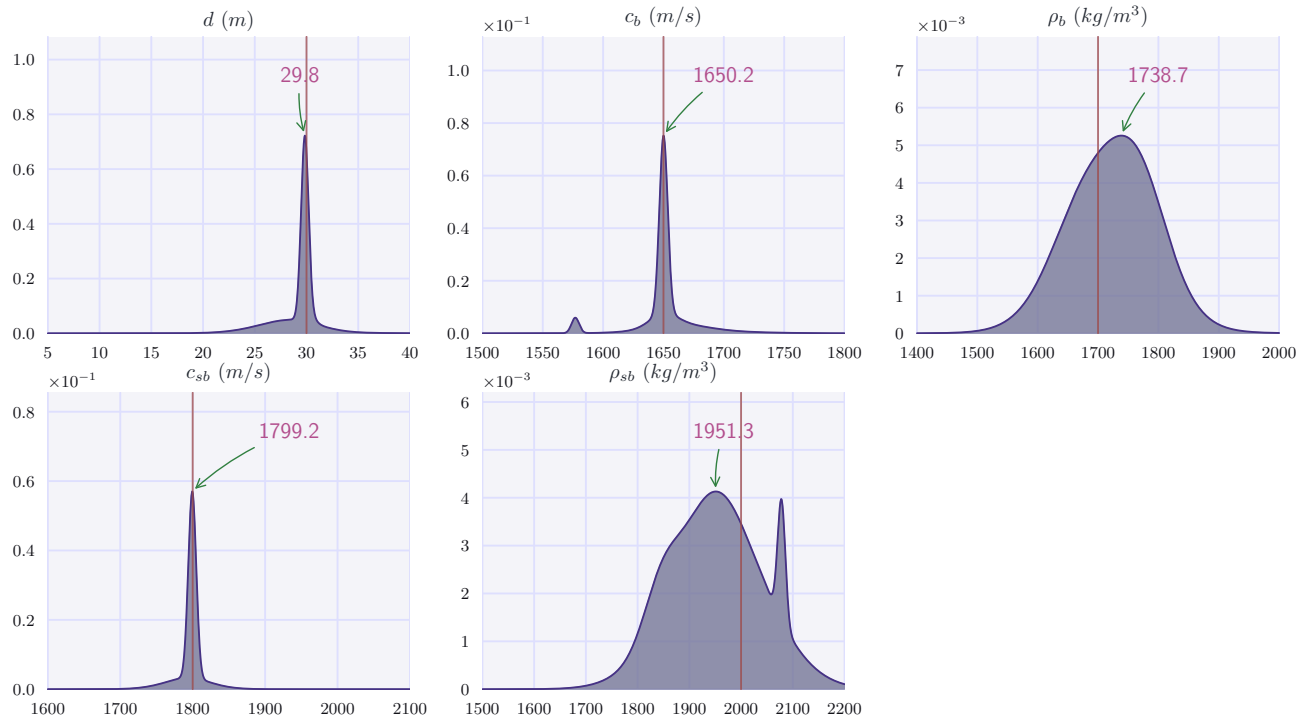


Figure 6.21: Marginal population distributions of the recoverable environmental and operational parameters of the case 2.c using **PSCS**. The distributions show statistics of the final population of GA. The vertical lines state for the actual values, whereas the arrows point to the maxima of the marginal densities.

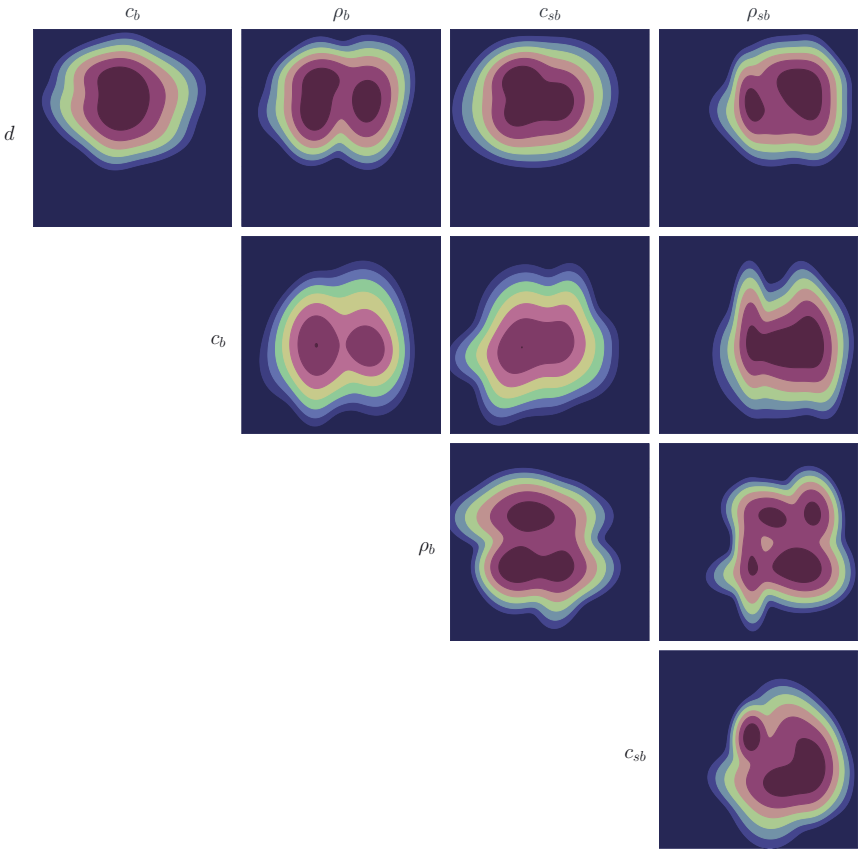


Figure 6.22: Joint population distributions corresponding to the inversion results of the case 2.c using **SSCS**. The distributions show statistics of the final population of GA. Each distribution has been normalized independently.

Table 6.10: Recovered model parameters corresponding to case 2.c. In parentheses the mismatch between the estimated and actual value of each parameter is provided.

Parameters	Units	SSCS	PSCS
Sediment thickness	$d\ (m)$	27.1 (-2.9)	29.8 (-0.2)
Sound speed in the sediment	$c_b\ (m/s)$	1632.7 (-17.3)	1650.2 $(+0.2)$
Sediment density	$\rho_b\ (kg/m^3)$	1625.2 (-74.8)	1738.7 $(+38.7)$
Sound speed in the substrate	$c_{sb}\ (m/s)$	1778.7 (-21.3)	1799.2 (-0.8)
Substrate density	$\rho_{sb}\ (kg/m^3)$	2023.4 $(+23.4)$	1951.3 (-48.7)

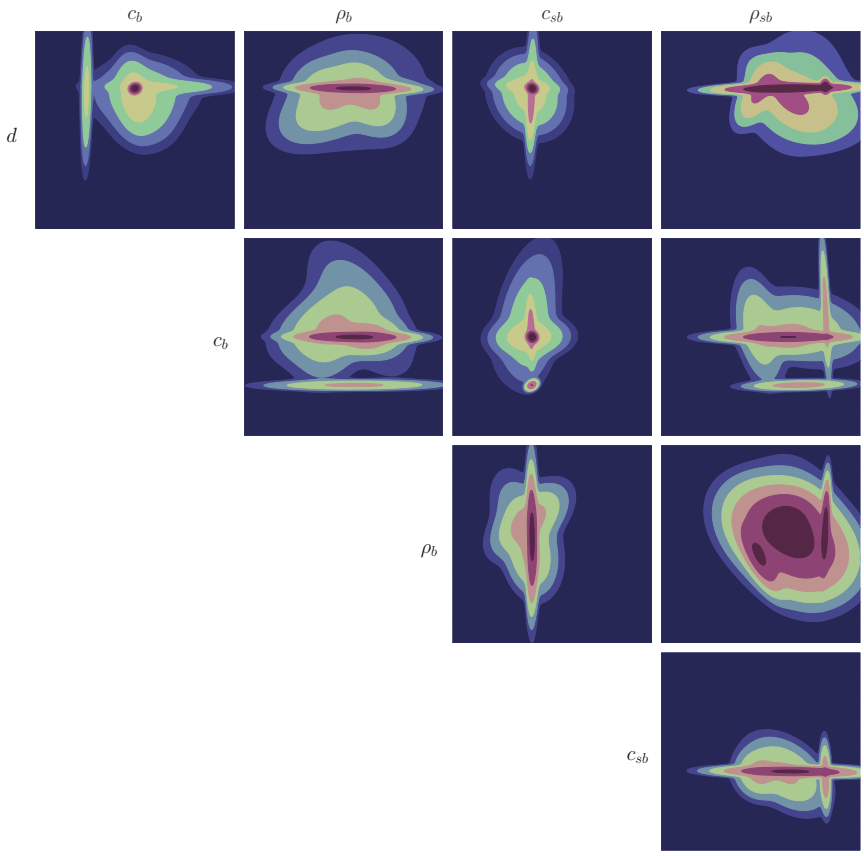


Figure 6.23: Joint population distributions corresponding to the inversion results of the case 2.c using **PSCS**. The distributions show statistics of the final population of GA. Each distribution has been normalized independently.

— Case 2.d : Recovering all the unknown parameters —

The actual potential of the inversion techniques presenting in this work is to be demonstrated, when we are dealing with multiparametric inversion scenarios. Although these scenarios are not usual in practical applications, it is interesting to test the methods in this complicated inversion problem.

The necessity of simultaneously inversion with many parameters, is be notably in many real applications where there is a significant lack of knowledge of the characteristics of the marine environment.

The recoverable parameters of this inverse problem corresponding to this test case is the union of the subsets considered in the previous subcases and it is defined by the following vector:

$$\mathbf{m} = [z_{min},\ c_1,\ c_2,\ c_3,\ d,\ c_b,\ \rho_b,\ c_{sb},\ \rho_{sb},\ R,\ z_s,\ z_r]^T \tag{6.9}$$

The post-processing of the set of the GA candidate solutions obtained by both methods gives the marginal and the joint density functions appearing in Figures 6.24 to 6.27. The inversion results in both cases were very well. Furthermore, the PSCS results again gave narrower pop-

ulation distributions compared to the SSCS results for most of the recoverable parameters. The results of this example is a strong indication that the PSCS and secondary SSCS methods are very good choices for multiparametric inverse problems. The inversion results and their mismatches are given in Table 6.11.

Table 6.11: Recovered model parameters corresponding to case 2.d. In parentheses the mismatch between the estimated and actual value of each parameter is provided.

Parameters	Units	SSCS	PSCS
Depth of minimum sound speed	$z_{min}\ (m)$	19.1 (-0.9)	19.8 (-0.2)
Sound speed at $z = 0$	$c_1\ (m/s)$	1500.5 $(+0.5)$	1499.7 (-0.3)
Minimum sound speed	$c_2\ (m/s)$	1494.2 $(+4.2)$	1495.0 $(+5.0)$
Sound speed at depth h	$c_3\ (m/s)$	1507.1 $(+2.1)$	1503.8 (-1.2)
Sediment thickness	$d\ (m)$	32.9 $(+2.9)$	29.8 (-0.2)
Sound speed in the sediment	$c_b\ (m/s)$	1625.5 (-24.5)	1649.8 (-0.2)
Sediment density	$\rho_b\ (kg/m^3)$	1643.2 (-56.8)	1662.5 (-37.5)
Sound speed in the substrate	$c_{sb}\ (m/s)$	1839.2 $(+39.2)$	1805.2 $(+5.2)$
Substrate density	$\rho_{sb}\ (kg/m^3)$	1866.5 (-133.5)	2031.1 $(+31.1)$
Range	$R\ (m)$	8002.7 $(+2.7)$	8000.2 $(+0.2)$
Source depth	$z_s\ (m)$	80.7 $(+1.7)$	80.1 $(+0.1)$
Receiver depth	$z_r\ (m)$	81.8 $(+1.8)$	80.0 (0.0)

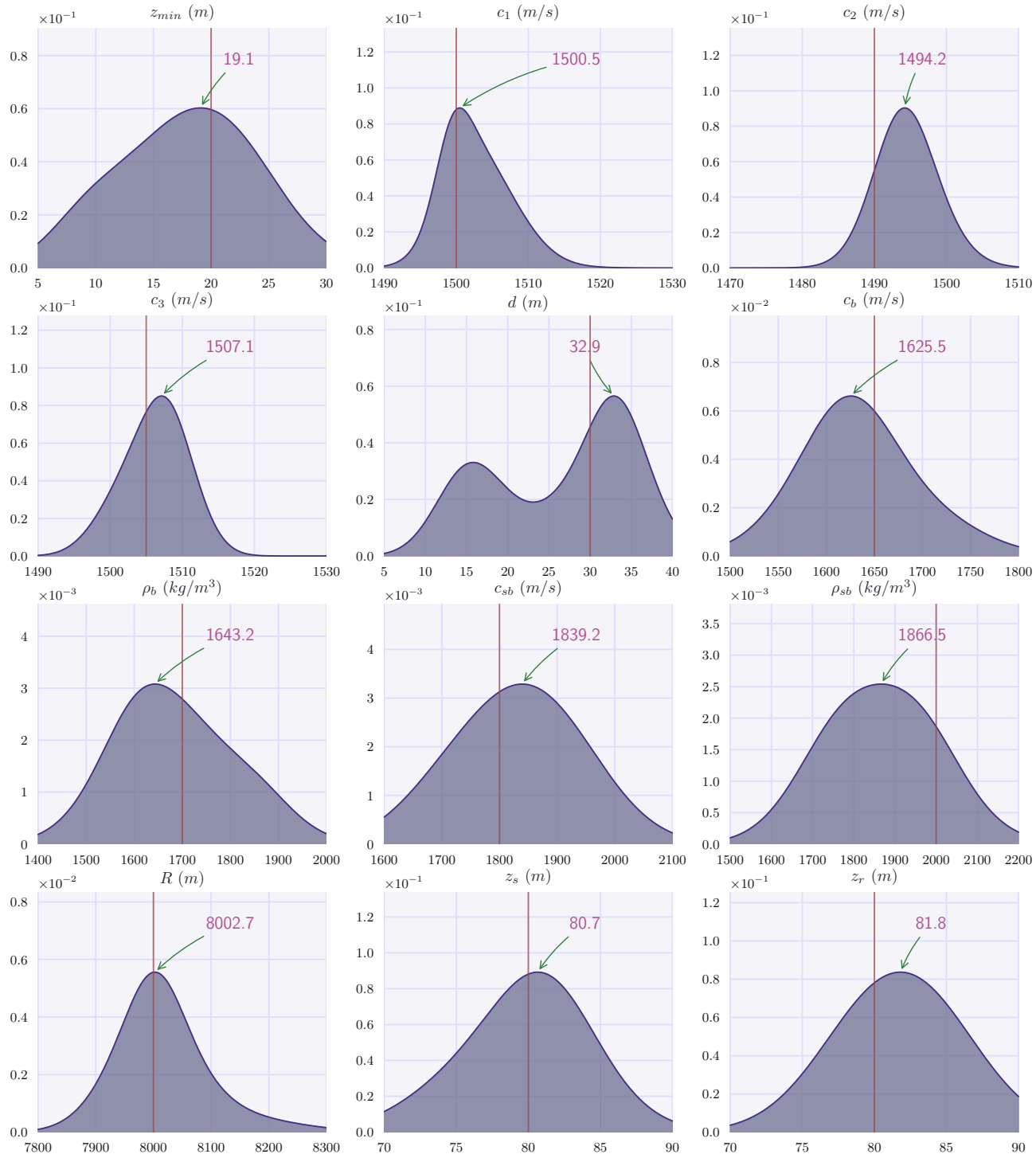


Figure 6.24: Marginal population distributions of the recoverable environmental and operational parameters of the case 2.d using **SSCS**. The distributions show statistics of the final population of GA. The vertical lines state for the actual values, whereas the arrows point to the maxima of the marginal densities.

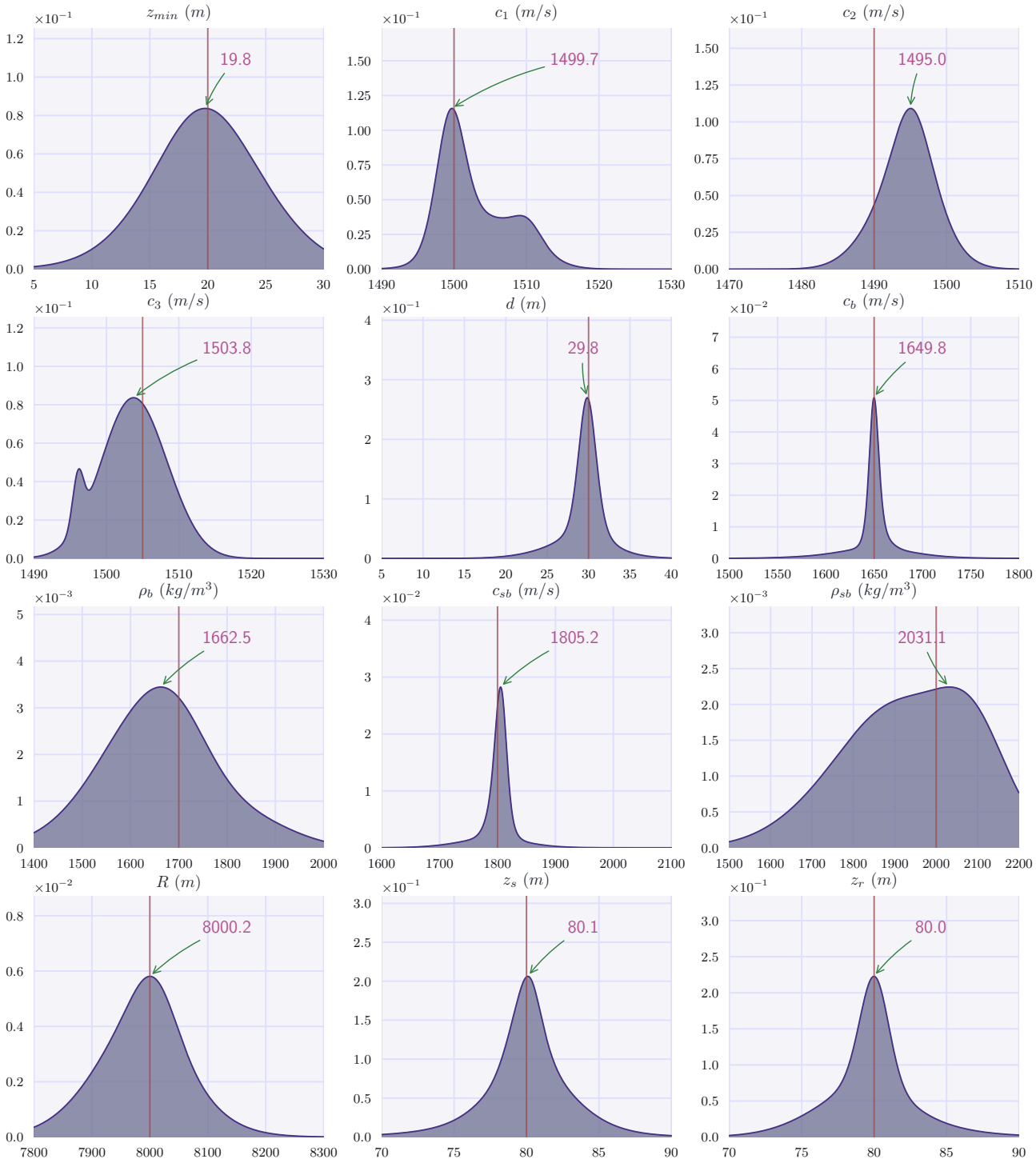


Figure 6.25: Marginal population distributions of the recoverable environmental and operational parameters of the case 2.d using **PSCS**. The distributions show statistics of the final population of GA. The vertical lines state for the actual values, whereas the arrows point to the maxima of the marginal densities.

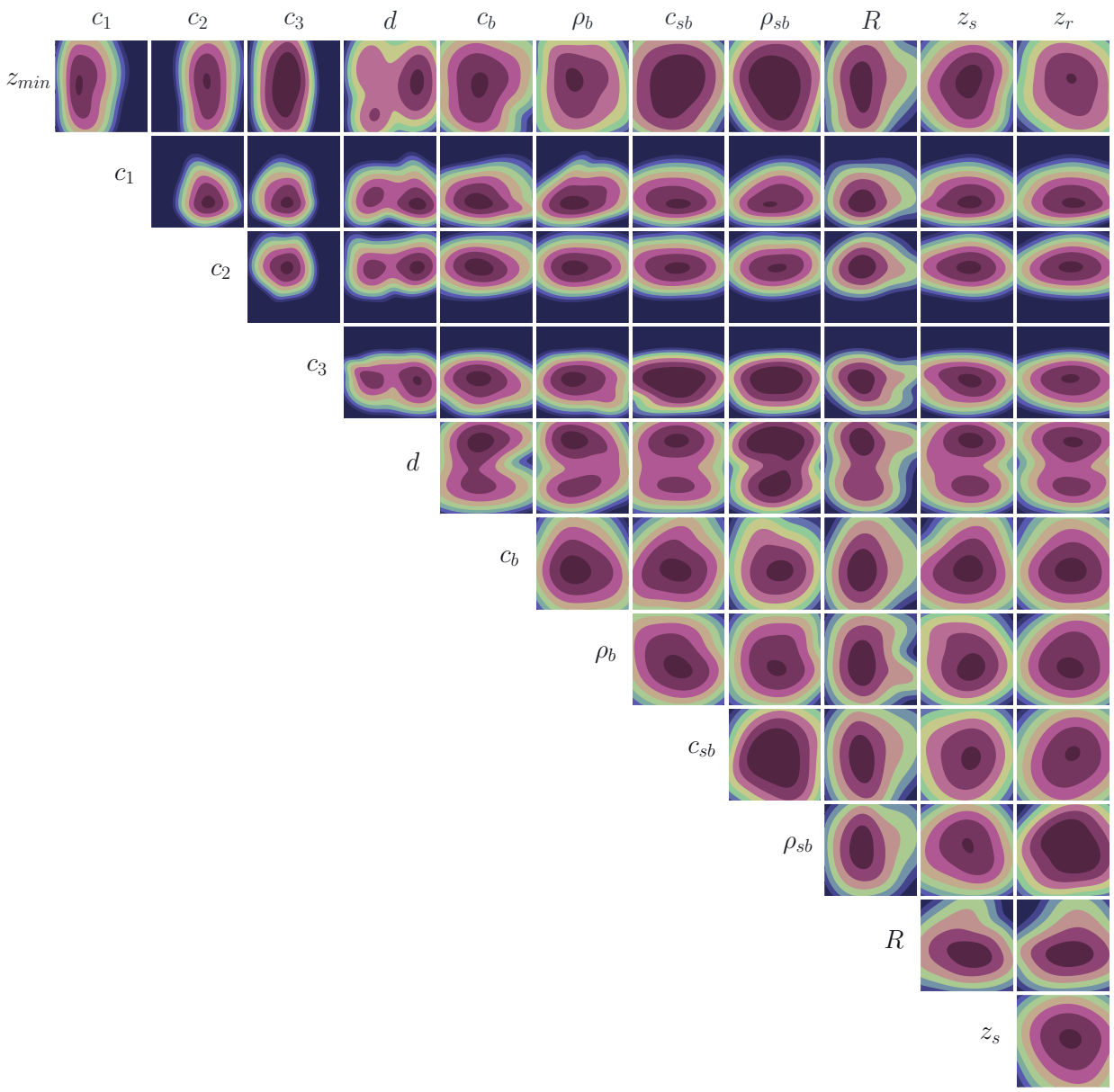


Figure 6.26: Joint population distributions corresponding to the inversion results of the case 2.d using **SSCS**. The distributions show statistics of the final population of GA. Each distribution has been normalized independently.

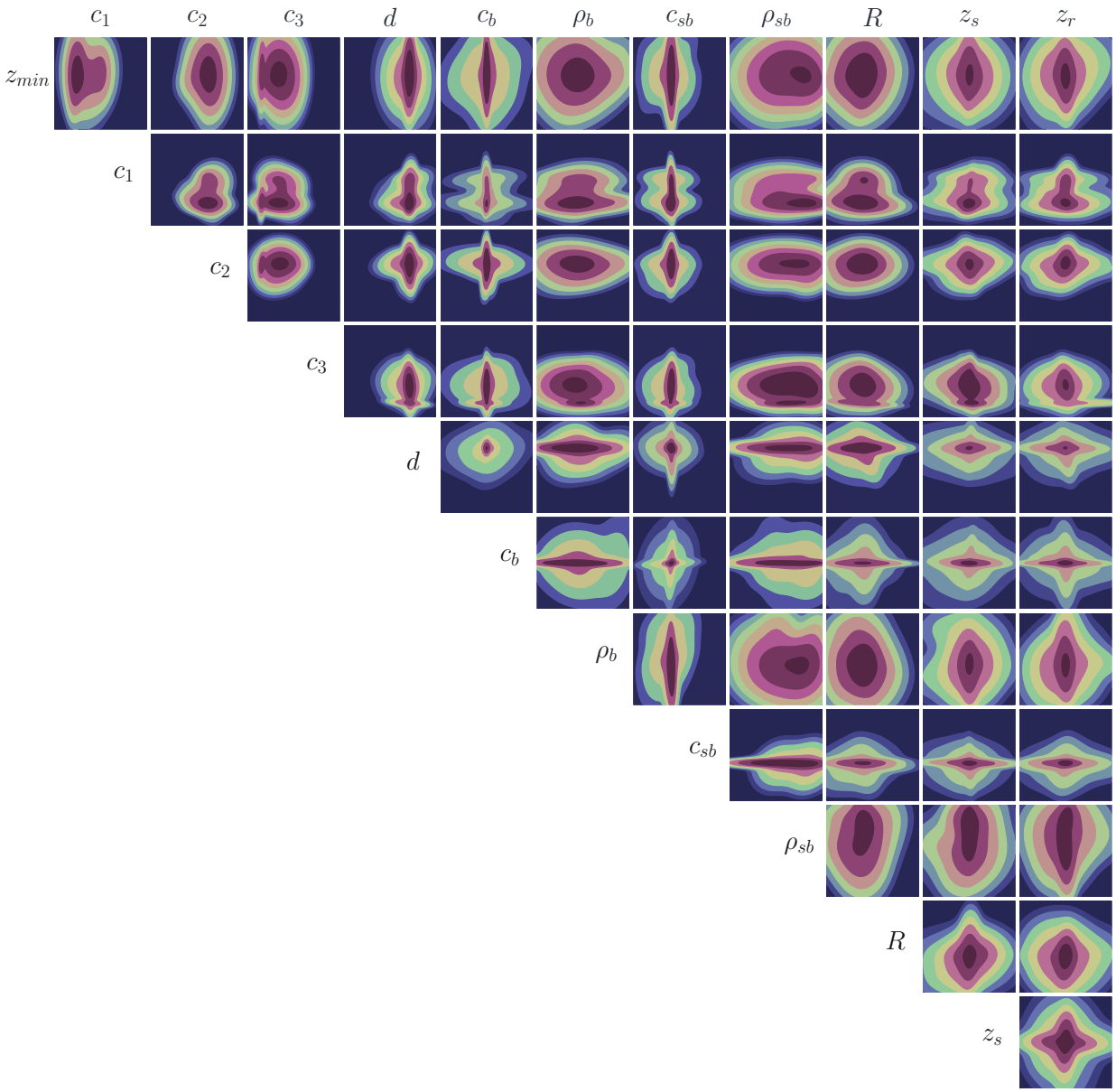


Figure 6.27: Joint population distributions corresponding to the inversion results of the case 2.d using **PSCS**. The distributions show statistics of the final population of GA. Each distribution has been normalized independently.

6.2.3 Test case 3 : A cold eddy in shallow water

The third test case is referred to a range dependent environment with the presence of a cold eddy. Table 6.12 presents the environmental and operational parameters of the simulated experiment. In this environment a single receiver is considered to be 8 km away from a source with a Gaussian excitation function of central frequency 50 Hz and bandwidth 20 Hz.

Three versions of the simulated acoustic signal appear in Figure 6.30.

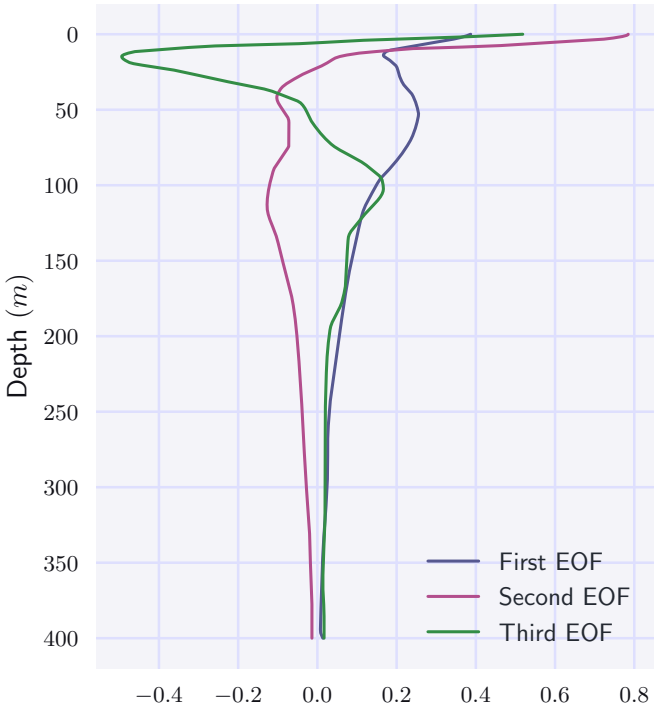


Figure 6.28: The three EOFs representation the sound speed anomaly for the cold eddy of the second test case.

In our case, the cold eddy which is consider a compact support anomaly, is represented in terms of three orders of Empirical Orthogonal Functions (EOFs) appearing in Figure 6.28 consisting the basis functions by witch the fluctuations of a reference range-independent sound speed profile $c_0(z)$ is represented as follows:

$$c(r,z) = c_0(z) + \sum_{n=1}^{3} \alpha_n(r) f_n(z). \tag{6.10}$$

In the case under consideration the reference sound speed profile in the water column is picewise linear between the values $c_0(0) = 1500\ m/s$, $c_0(100) = 1495\ m/s$, and $c_0(400) = 1509\ m/s$.

Following the setup introduced in [127], we consider five segments of equal width in witch of these the sound speed profile is given by the following formula:

$$c(r,z) = c_0(z) + \sum_{n=1}^{3} \alpha_n^i f_n(z), \quad i = 1,\ldots,5. \tag{6.11}$$

The starting and ending range of the segments is considered known.

Therefore, the retrieval of the 15 parameters $\alpha_n^i,\ i = 1,\ldots,5,\ n = 1,2,3$ leads to an estimation of the sound speed profile in the water column. Note that the bottom properties are considered known.

Table 6.12 presents the environmental and operational parameters controlling this test case, while Table 6.13 contains the actual coefficients of the EOFs representing the sound speed variations along with their search spaces which are exactly ones used in previous works [127, 128, 129].

The model parameter vector contains totally 15 EOF coefficients, three for each segment. Therefore $\mathbf{m}$ has the form:

$$\mathbf{m} = [a_1^1, a_1^2, a_1^3, a_1^4, a_1^5, a_2^1, a_2^2, a_2^3, a_2^4, a_2^5, a_3^1, a_3^2, a_3^3, a_3^4, a_3^5]^T. \tag{6.12}$$

Table 6.12: Environmental, source and receiver parameters of the test case 3 (Cold Eddy).

Parameters	Value
Water Depth (m)	400
Density of the Water (kg/m^3)	1000
Starting Range of the Eddy (m)	2000
Ending Range of the Eddy (m)	3200
Sound Velocity at the Bottom (m/s)	1600
Density of the Bottom (kg/m^3)	1500
Source Depth (m)	50
Receiver Depth (m)	50
Receiver Range (m)	5000
Central Frequency/Bandwidth (Hz)	50/20

The statistical characterization scheme again is driven by the 3-level wavelet multiresolution analysis of the synthetic signal using the Daubechies 4 (db4) wavelet, while the probabilistic one is performed based on the 4-level stationary wavelet packet decomposition using the same wavelet. The selection of the coefficient vectors for characterizing the signal was chosen

Segments						Search Space	
Order	1st	2nd	3rd	4th	5th	L. Bound	U. Bound
1	−19.21	−33.007	−44.713	−25.657	−8.726	−50.0	0.0
2	27.845	34.352	44.438	32.824	22.876	0.0	50.0
3	−11.105	−11.008	−14.892	−13.073	−12.001	−30	0

Table 6.13: EOF coefficients and search space for the environment of test case 3.

considering the energy contributions of each one of the final level of the decomposition as illustrated by Figure 6.32.

Then, the inverse problem was solved through a Genetic Algorithm (GA).

The post-processing of the final population of the GA was carried out by employing a Mixture Model that consists of 8 normal distributions with full covariance matrices.

For this case the coupled normal mode program MODE4 developed at FORTH, capable of treating range-depended problems of are used to obtain the replica fields.

Figures 6.34 and 6.33 show the marginal probabilities densities of the problem's recoverable parameters. Furthermore, the joint distributions are presented in Figures 6.35 and 6.36. Note that the inversion results for the recoverable parameters are very close to their actual values.

Figures 6.37 illustrates the real and the reconstructed structures of the eddy. Moreover, in Figure 6.38 we can see the actual and the recovered water sound speed profiles for each one of the five segments of the simulated environment.

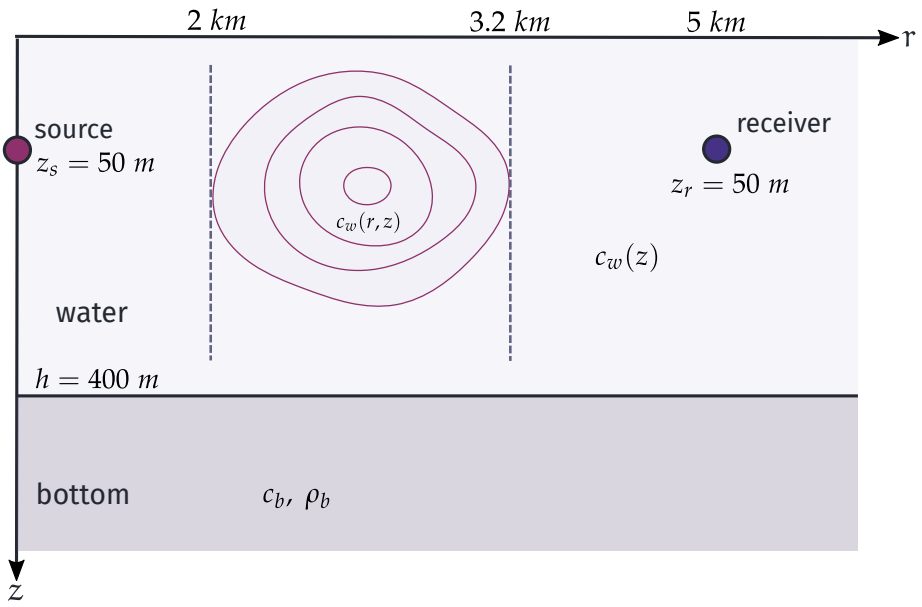


Figure 6.29: The simulated cold eddy sea environment of the third inversion test case.

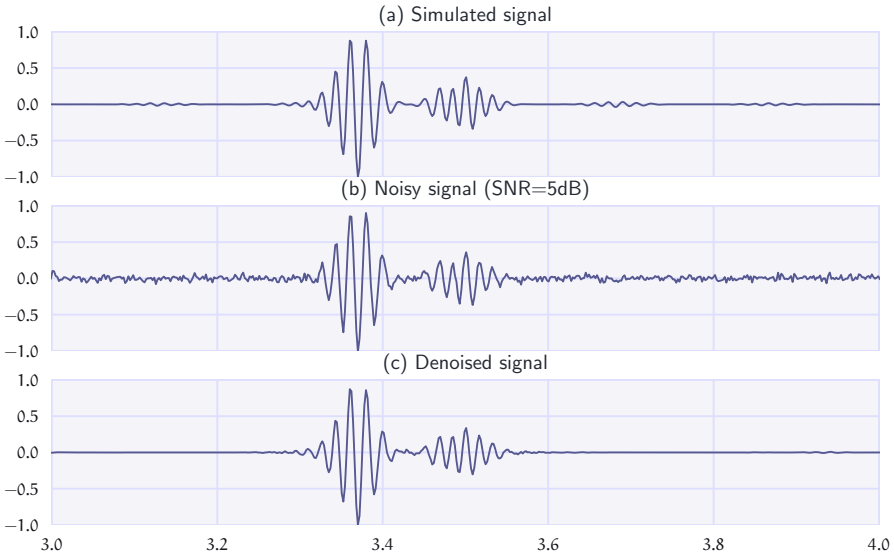


Figure 6.30: Three versions of the simulated signal that corresponds to the propagation of the simulated sound through the environment with the present of a cold eddy.

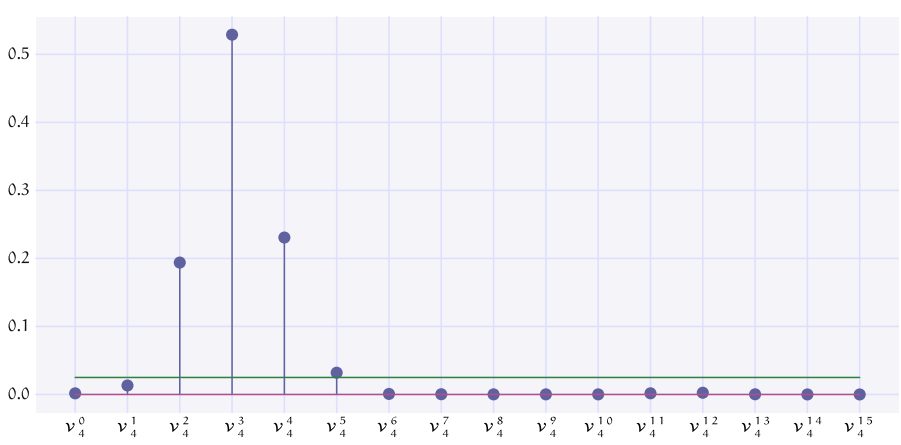


Figure 6.31: Energy contribution per wavelet packet coefficient band corresponding to the eddies' signal. Threshold has been picked 0.025.

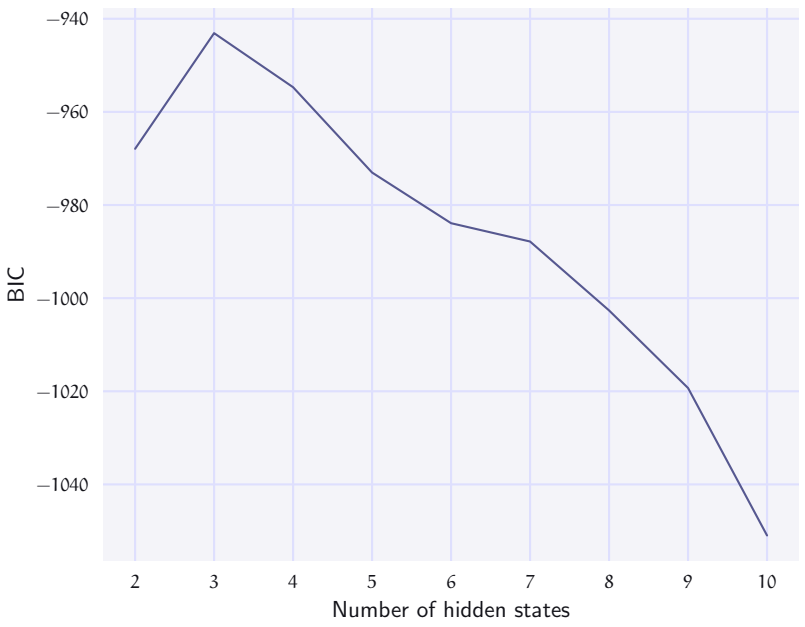


Figure 6.32: The BIC penalized likelihood criterion for selecting the number of the hidden states.

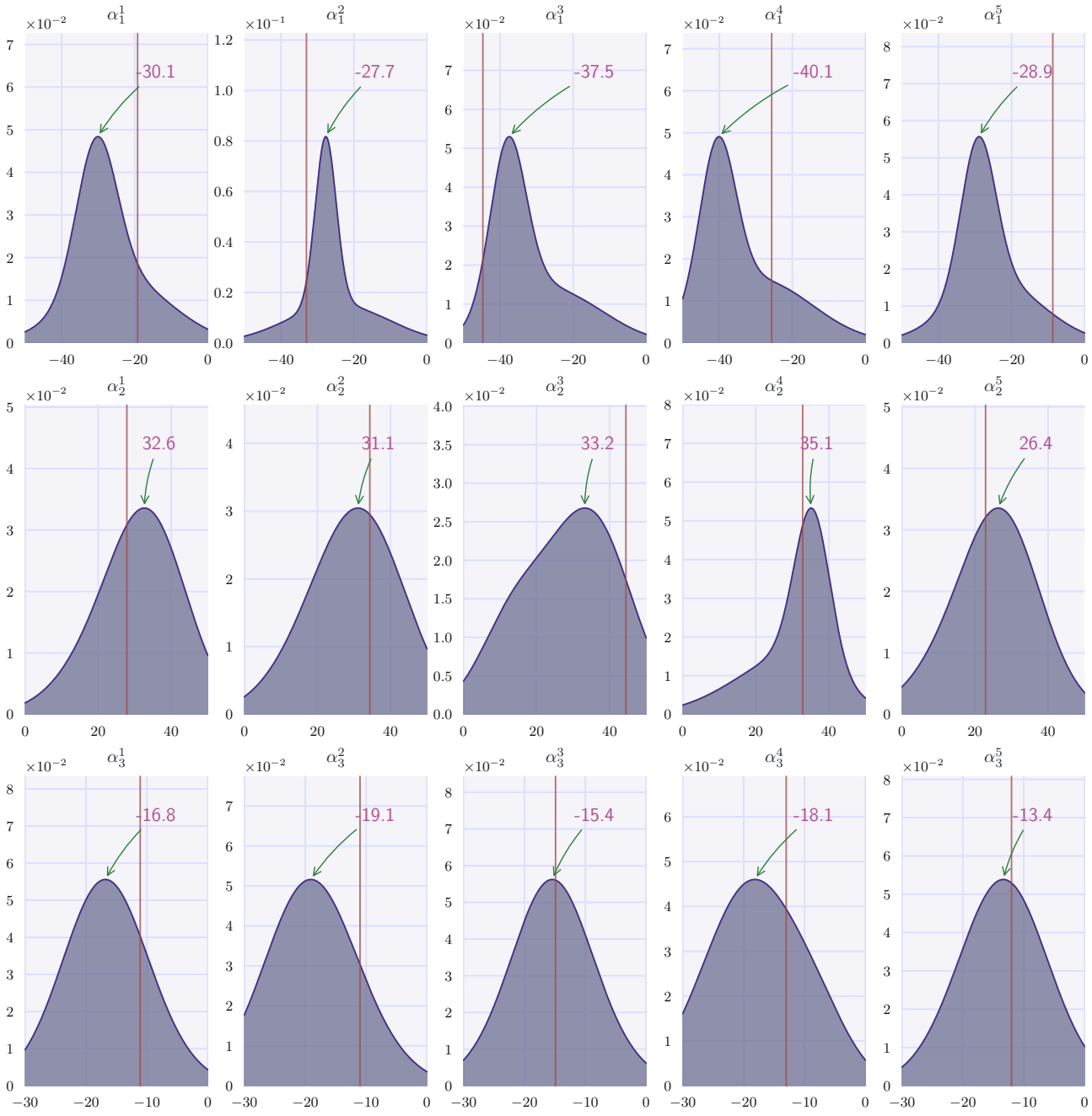


Figure 6.33: Marginal population distributions of the EOFs coefficients describing the eddy of the third test case using **SSCS**. The distributions show statistics of the final population of GA. The vertical lines state for the actual values, whereas the arrows point to the maxima of the marginal densities.

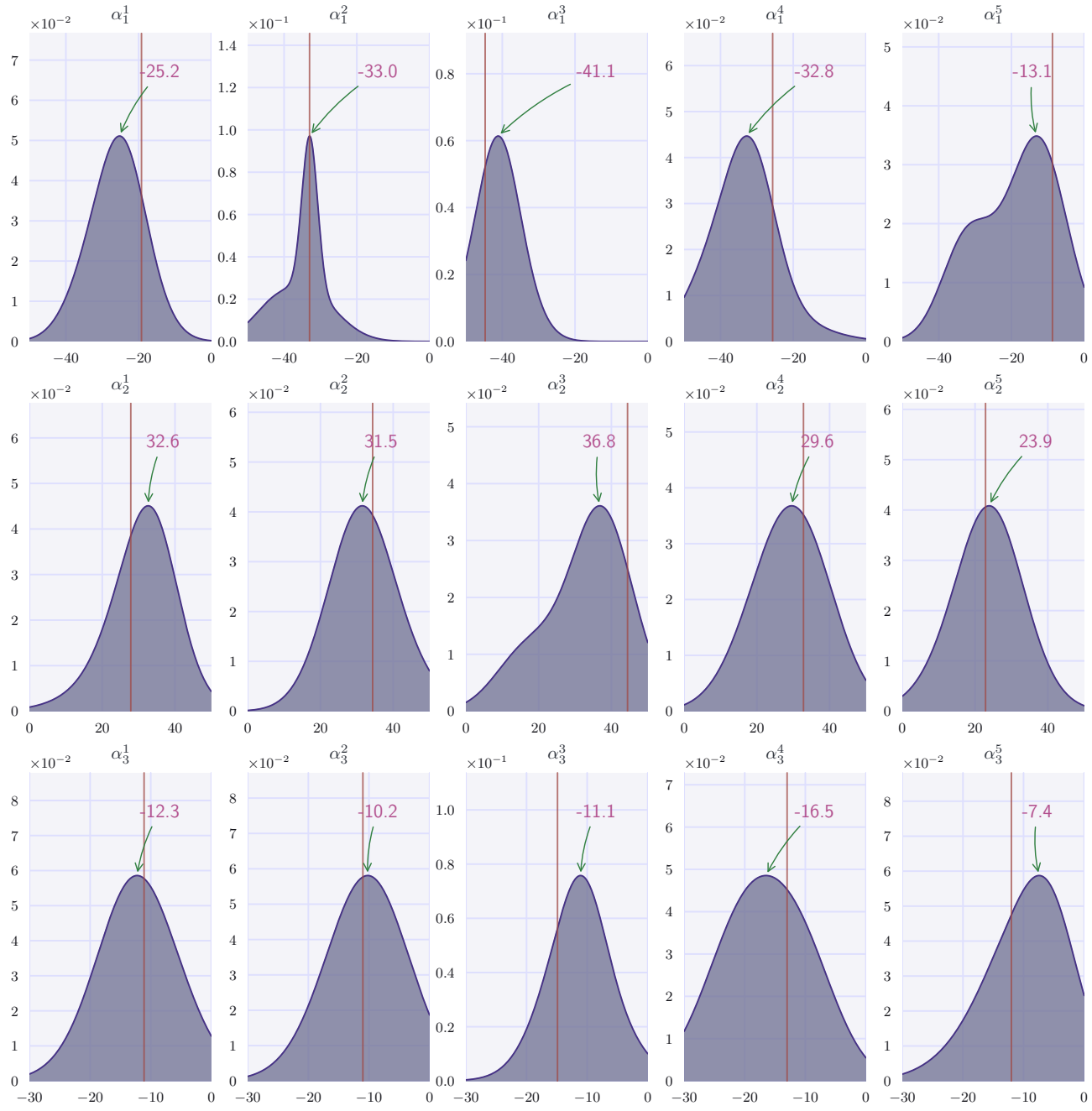


Figure 6.34: Marginal population distributions of the EOFs coefficients describing the eddy of the third test case using **PSCS**. The distributions show statistics of the final population of GA. The vertical lines state for the actual values, whereas the arrows point to the maxima of the marginal densities.

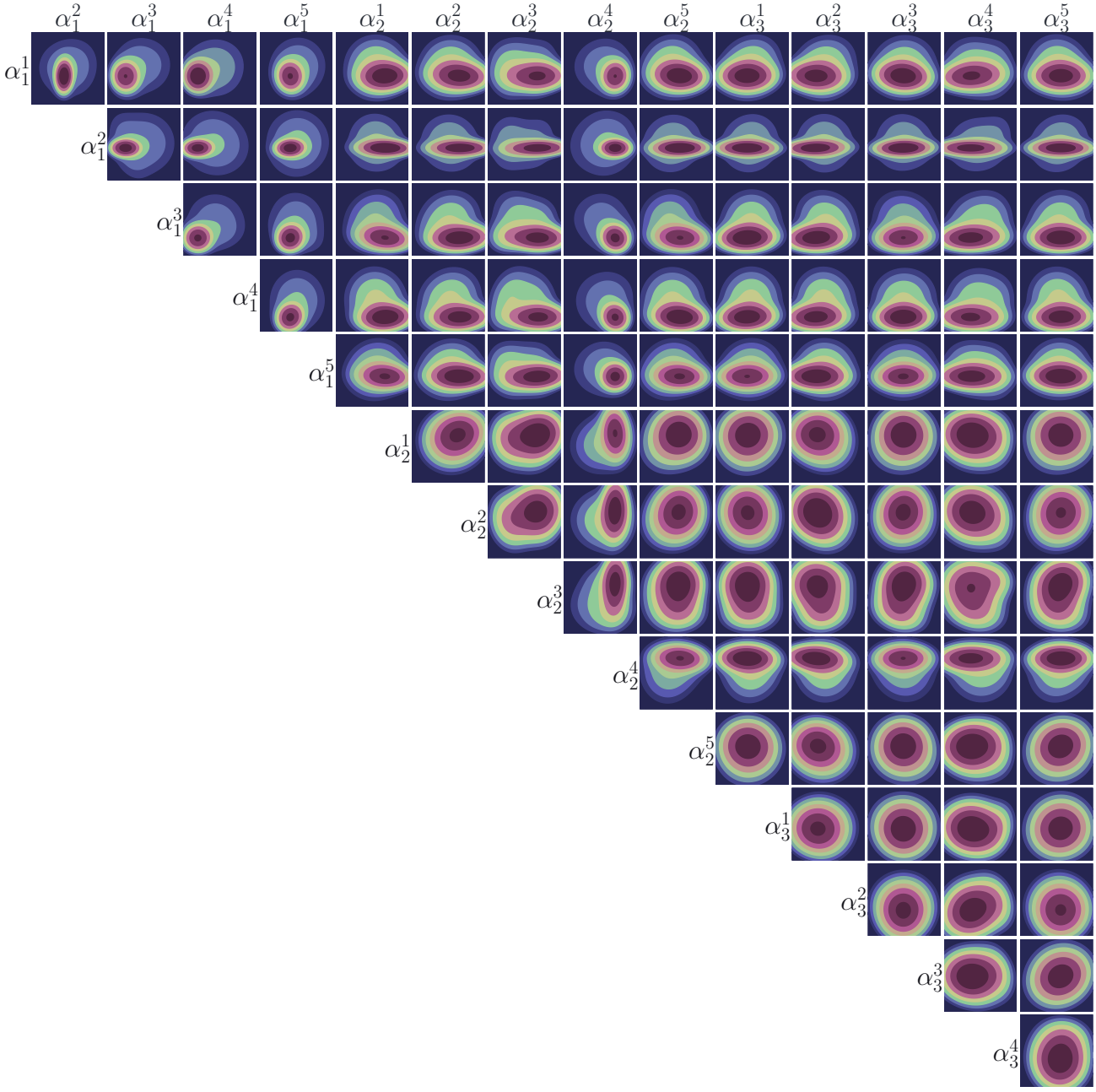


Figure 6.35: Joint population distributions corresponding to the inversion results of the third test case using **SSCS**. The distributions show statistics of the final population of GA. Each distribution has been normalized independently.

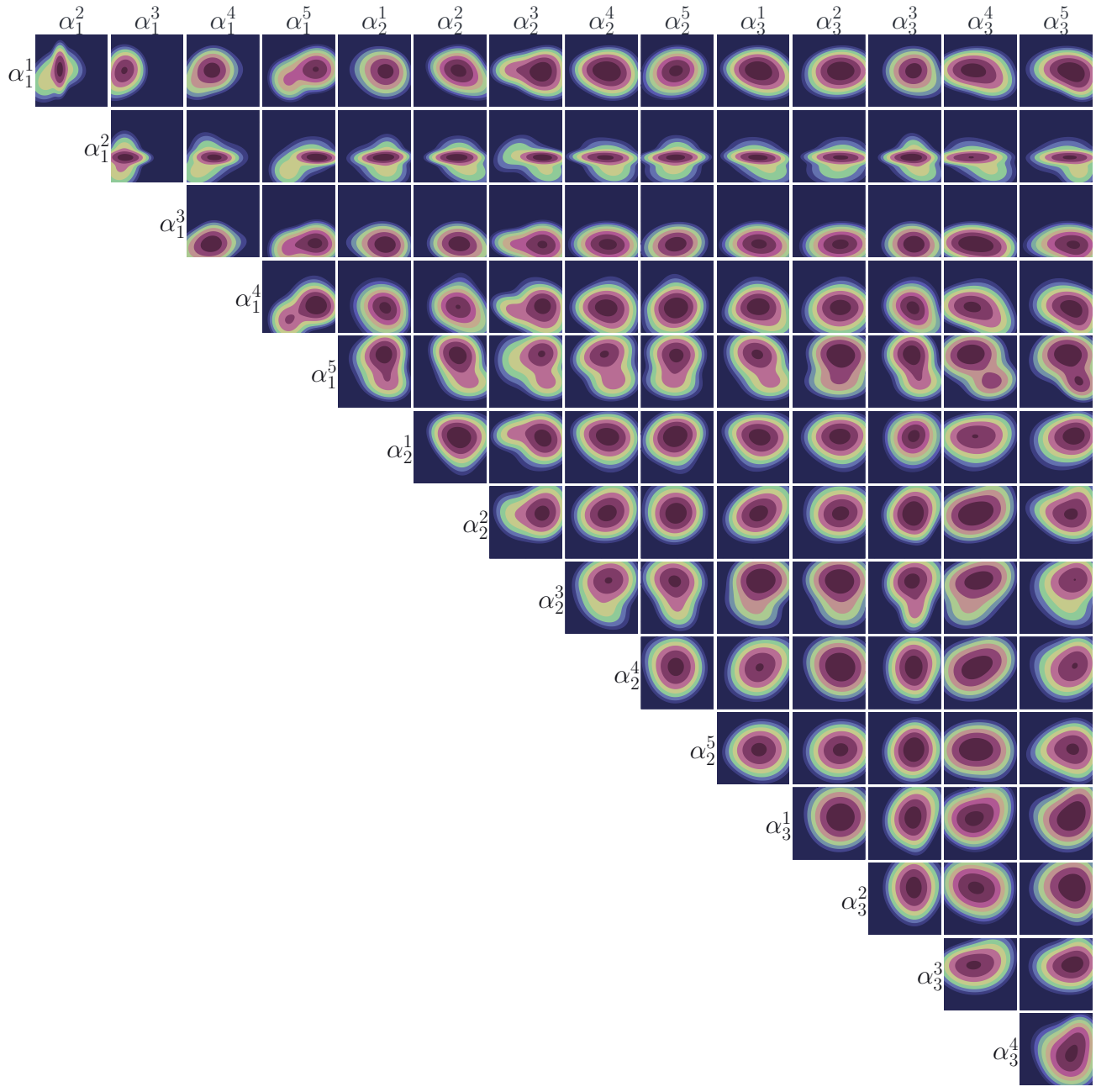


Figure 6.36: Joint population distributions corresponding to the inversion results of the third test case using **PSCS**. The distributions show statistics of the final population of GA. Each distribution has been normalized independently.

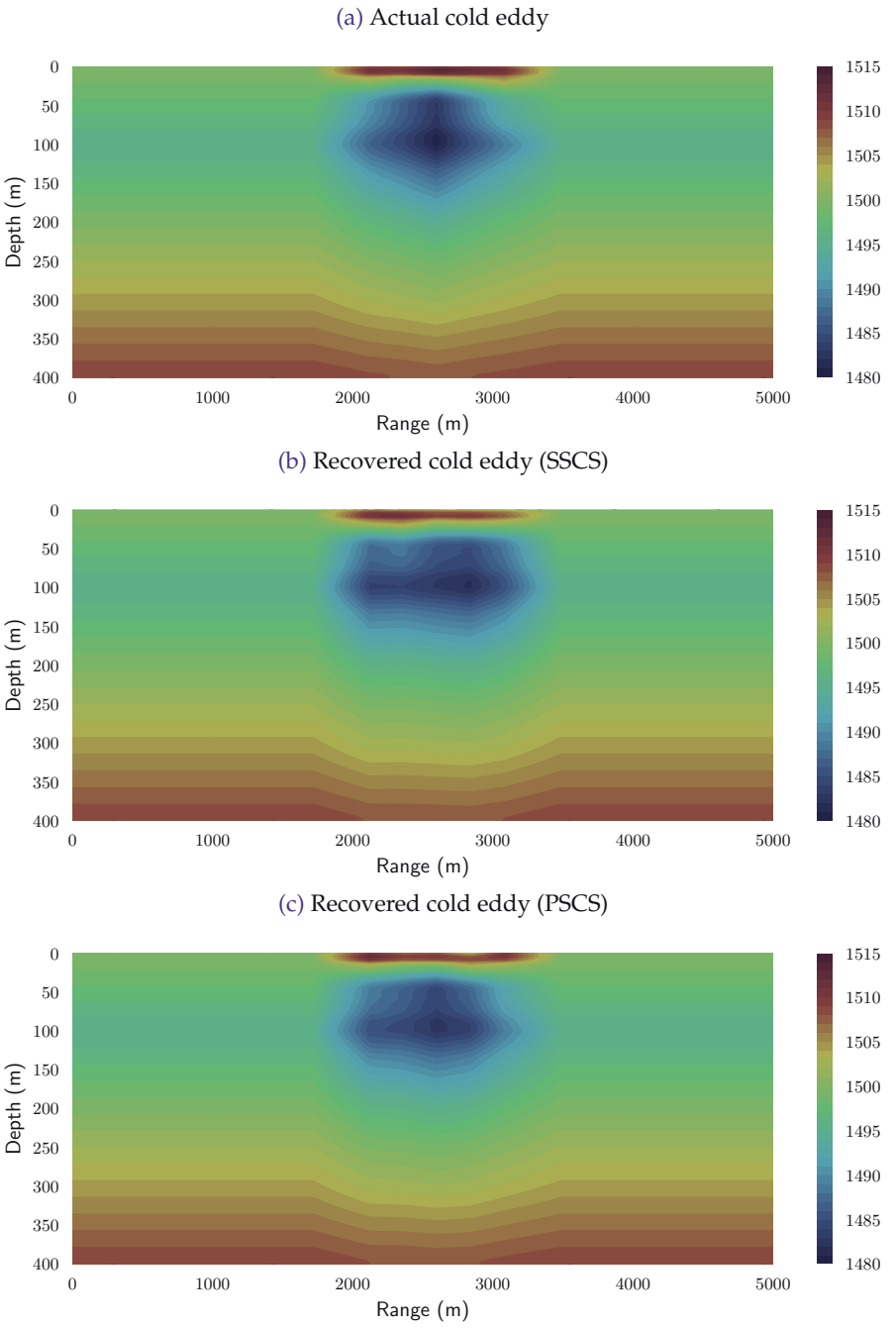


Figure 6.37: The real and the recovered cold eddies using both the statistical and the probabilistic schemes.

It is interested to note that both methods have been able to recover the eddy with very good qualitative results. Of course the prior information on the extend of the eddy was essential in confining the range-dependent irregularity at a certain area. However, it is noticeably that the recovery of the sound speed profiles in each segment was very good despite the fact that the search space in each segment was the same. This test case as an additional evidence of the potential of both SSCS and PSCS methods for signal characterization and subsequent inversion in problems of acoustical oceanography even in range-dependent cases.

Table 6.14: Recovered model parameters corresponding to case of the cold eddy. In parentheses the mismatch between the estimated and actual value of each parameter is provided.

Parameters	SSCS	PSCS
α_1^1	-30.1 (-10.9)	-25.2 (-6.0)
α_1^2	-27.7 $(+5.3)$	-33.0 (0.0)
α_1^3	-37.5 $(+7.2)$	-41.1 $(+3.6)$
α_1^4	-40.1 (-14.4)	-32.8 (-7.1)
α_1^5	-28.9 (-20.2)	-13.1 (-4.4)
α_2^1	32.6 $(+4.8)$	32.6 $(+4.8)$
α_2^2	31.1 (-3.3)	31.5 (-2.9)
α_2^3	33.2 (-11.2)	36.8 (-7.6)
α_2^4	35.1 $(+2.3)$	29.6 (-3.2)
α_2^5	26.4 $(+3.5)$	23.9 $(+1.1)$
α_3^1	-16.8 (-5.7)	-12.3 (-1.2)
α_3^2	-19.1 (-8.1)	-10.2 $(+0.8)$
α_3^3	-15.4 (-0.5)	-11.2 $(+4.7)$
α_3^4	-18.1 (-5.0)	-16.4 (-3.4)
α_3^5	-13.4 (-1.4)	-7.4 $(+4.6)$

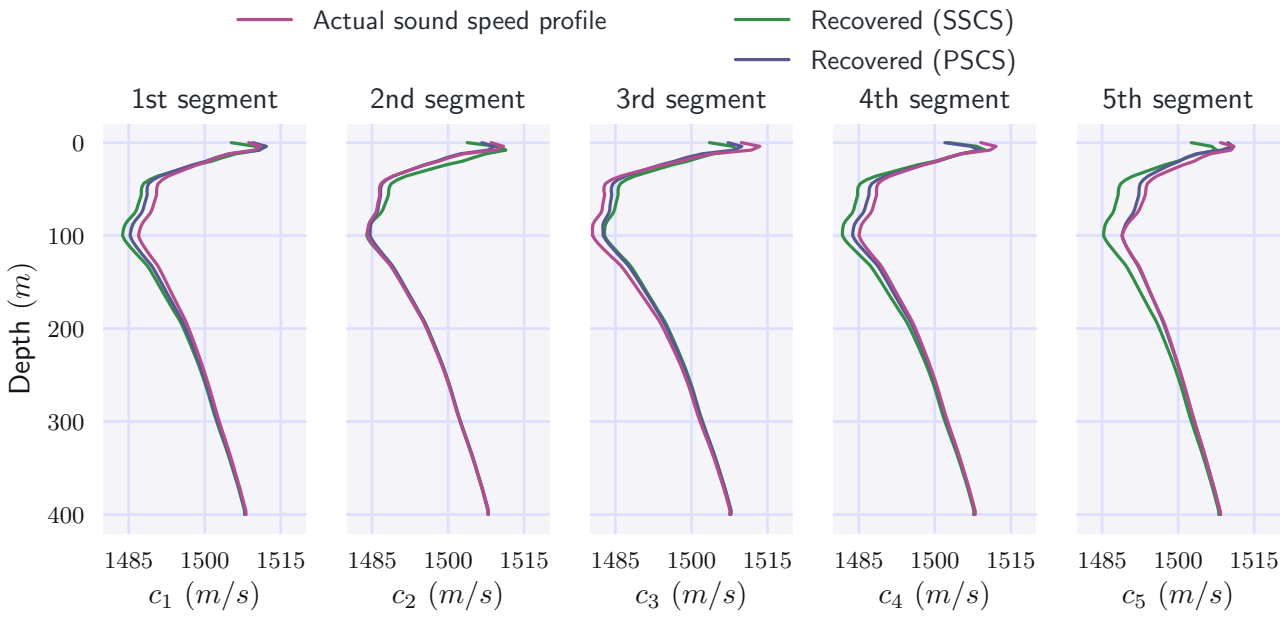


Figure 6.38: Sound speed profiles (actual and recovered) for the five discrete segments of the simulated cold eddy (test case 3). The recovered profiles are formed using the maxima of the marginal densities of the EOFs coefficients.

6.3 Applications with Real Data from the SW06 Experiment

In 2006, a multipurpose experiment (SW06) was carried out off the coast of New Jersey. For geoacoustic inversion purposes, evacuated glass light bulbs were lowered into the water and imploded at a depth of approximately 22 m, producing an impulsive source recorded at distance of about 7 km. The propagation channel was a shallow water waveguide with almost range-independent characteristics. The water depth was about 79.1 m. The sound speed profile in the water column is considered known based on the measurements by a conductivity-temperature probe (CTD41) (Figure 6.40) during the light bulb implosion experiment [130, 131].

The sea-bed is described as a two-layer medium with a homogeneous semi-consolidated layer of approximately $20 - 25$ m thickness overlying a harder substrate. For the purposes of our study, both layers will be considered as fluids.

Three different formulations of the track of the experiment are to be used for producing the replica acoustic fields. The similarity measurements of these simulated data with the recorded signal (after denoising) will lead to estimations of the recoverable model parameters. Figure 6.57 shows the recorded signal together with its clean version after applying the sparse denoising scheme [129].

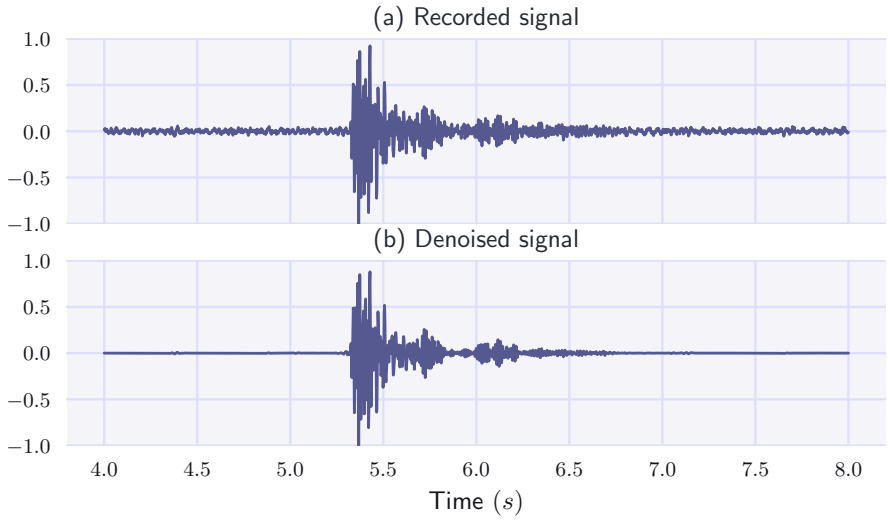


Figure 6.39: The raw and the enhanced signals from the SW06 experiment.

Since the denoised signal in Figure 6.57 was the one that we had employed in Sections 5.2 and 5.3 as a representative example in order to take efficient descriptions of the procedures of ex-

tracting the statistical and probabilistic features (observables) corresponding to the inversion schemes which are addressed by this thesis, we will not reproduce the same analysis here.

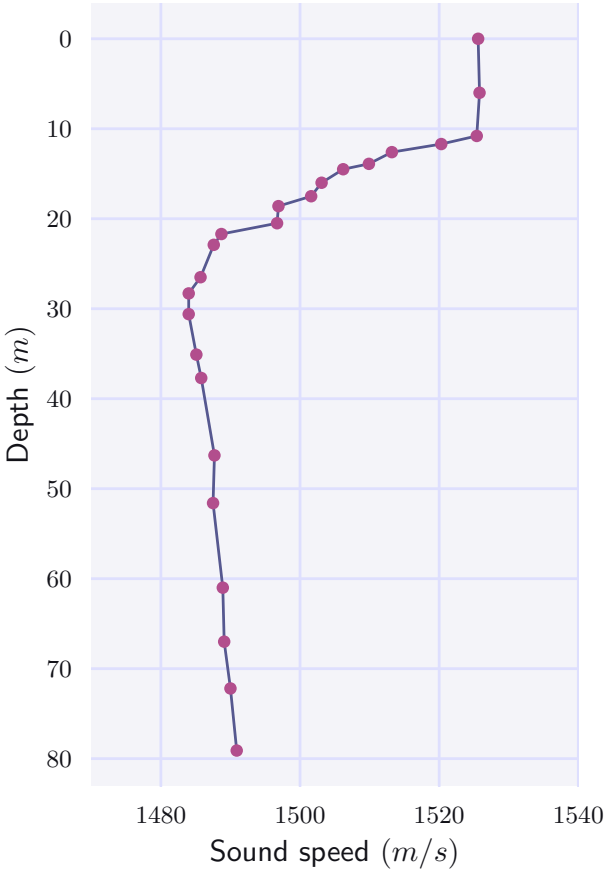


Figure 6.40: An approximation of the sound speed profile in the water column based on the measurements by CTD41 during the SW06 experiment.

6.3.1 First formulation

In the first and simpler formulation, we are following the model we have also used in previous works [106, 129]. Specifically, the environment is assumed to be range independent with a sediment layer overlying a semi-infinity substrate. Both seabed layers are supposed to be associated with constant sound velocities, densities and attenuations. Note that the attenuations introduced here is considered to be the imaginary part of the horizontal wave number following Jensen et. al to describe loss mechanism of the environment. For inversion cases can be considered as an additional parameter to be recovered.

The recoverable parameters are consisted of the source-receiver horizontal range, the sediment thickness and the sound velocities, the densities and attenuations for the seabed layers. These model parameter are expressed mathematically by the following vector

$$\mathbf{m} = [R, d, c_b, c_{sb}, \rho_b, \rho_{sb}, \alpha_b, \alpha_{sb}]^T. \tag{6.13}$$

Applying both the statistical and probabilistic inversion schemes as presented in the previous chapter, we end up with the marginal distributions for the recoverable parameters as presented by Figures 6.33 and 6.34, respectively. Moreover, Figures 6.44 and 6.45 show the joint

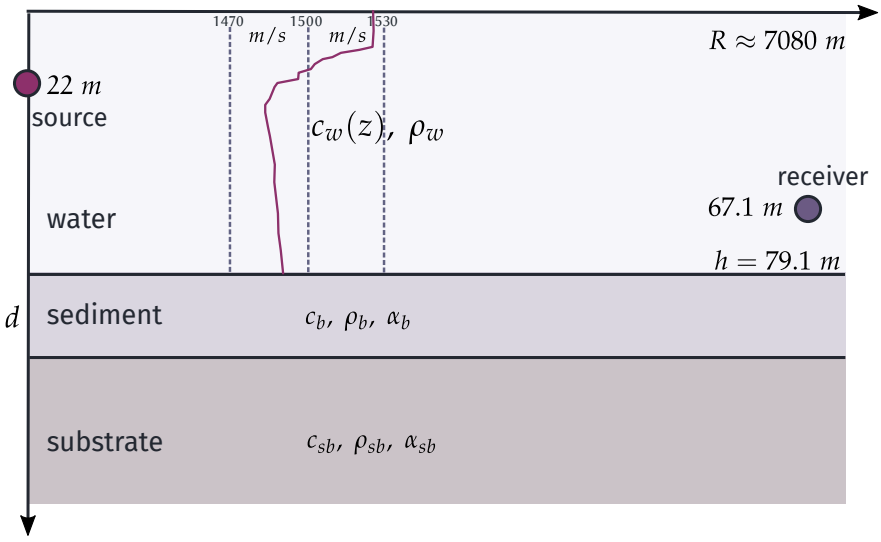


Figure 6.41: The range-independent environment that is considered quite close to the track of the Shallow Water 06 experiment corresponding to the first simpler formulation.

distributions associated to each inversion scheme. The most probable values of the model parameters are summarized in Table 6.15. The same table indicates also the selected search space. Comments in the inversion results will be presented in Section 6.3.4.

Parameters	Units	Search Space	Estimated value (PSCS)	Estimated value (SSCS)
Range	$R\ (m)$	[6750, 7250]	7081.8	6916.2
Sediment thickness	$d\ (m)$	[0, 30]	24.4	17.2
Sediment sound speed	$c_b\ (m/s)$	[1500, 1800]	1606.0	1636.3
Substrate sound speed	$c_{sb}\ (m/s)$	[1600, 2200]	1735.7	2019.8
Sediment density	$\rho_b\ (kg/m^3)$	[1400, 2200]	2005.4	1783.6
Substrate density	$\rho_{sb}\ (kg/m^3)$	[1400, 2200]	2044.6	1663.5
Sediment attenuation	$\alpha_b\ (nep/mkHz)$	[0, 0.03]	0.005375	0.004384
Substrate attenuation	$\alpha_{sb}\ (nep/mkHz)$	[0, 0.03]	0.01955	0.02069

Table 6.15: Estimated values of the recoverable parameters using the first formulation.

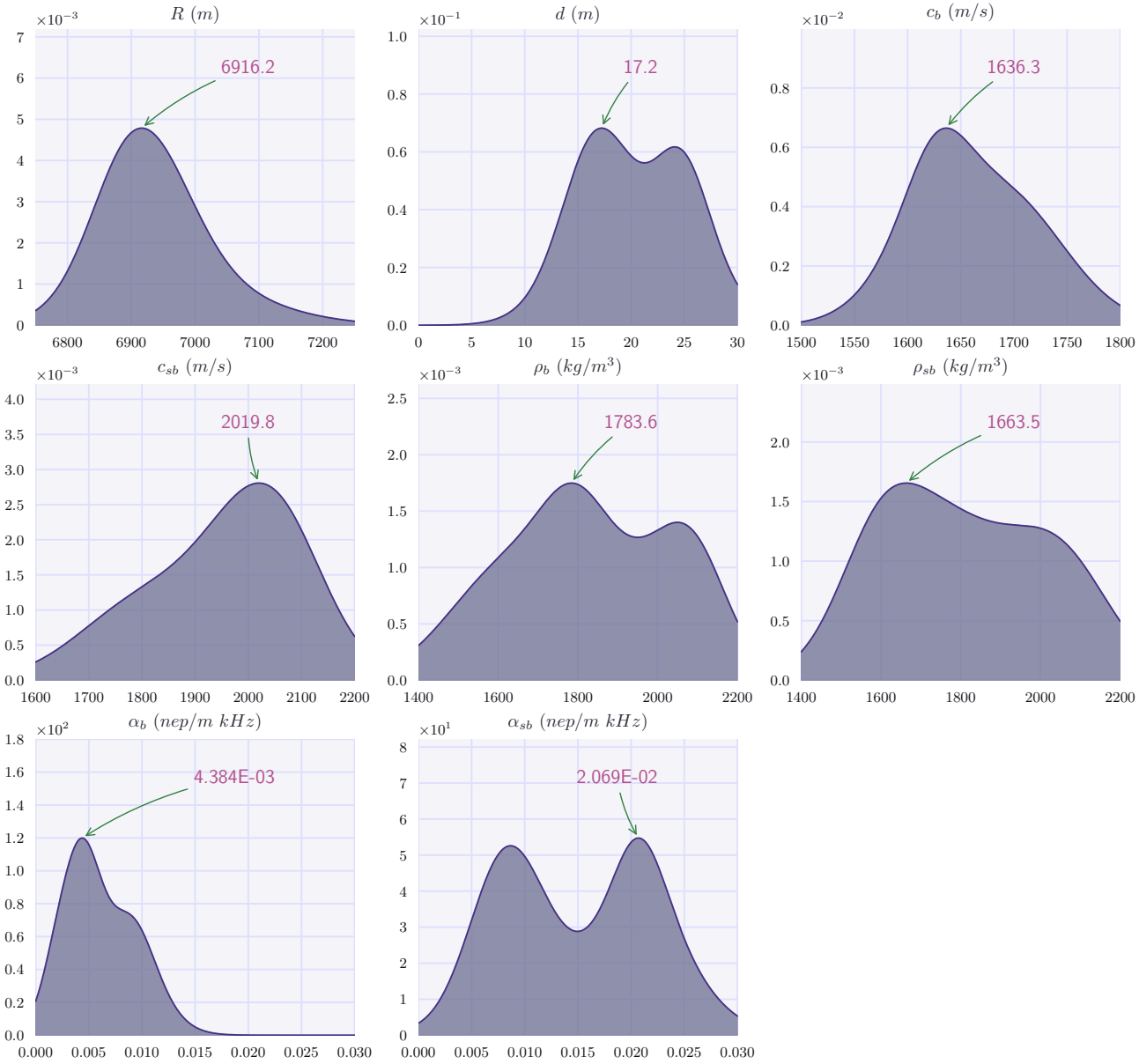


Figure 6.42: Marginal population distributions of the unknown parameters of the sea environment of the SW06 experiment using the **SSCS**. The distributions show statistics of the final population of GA.

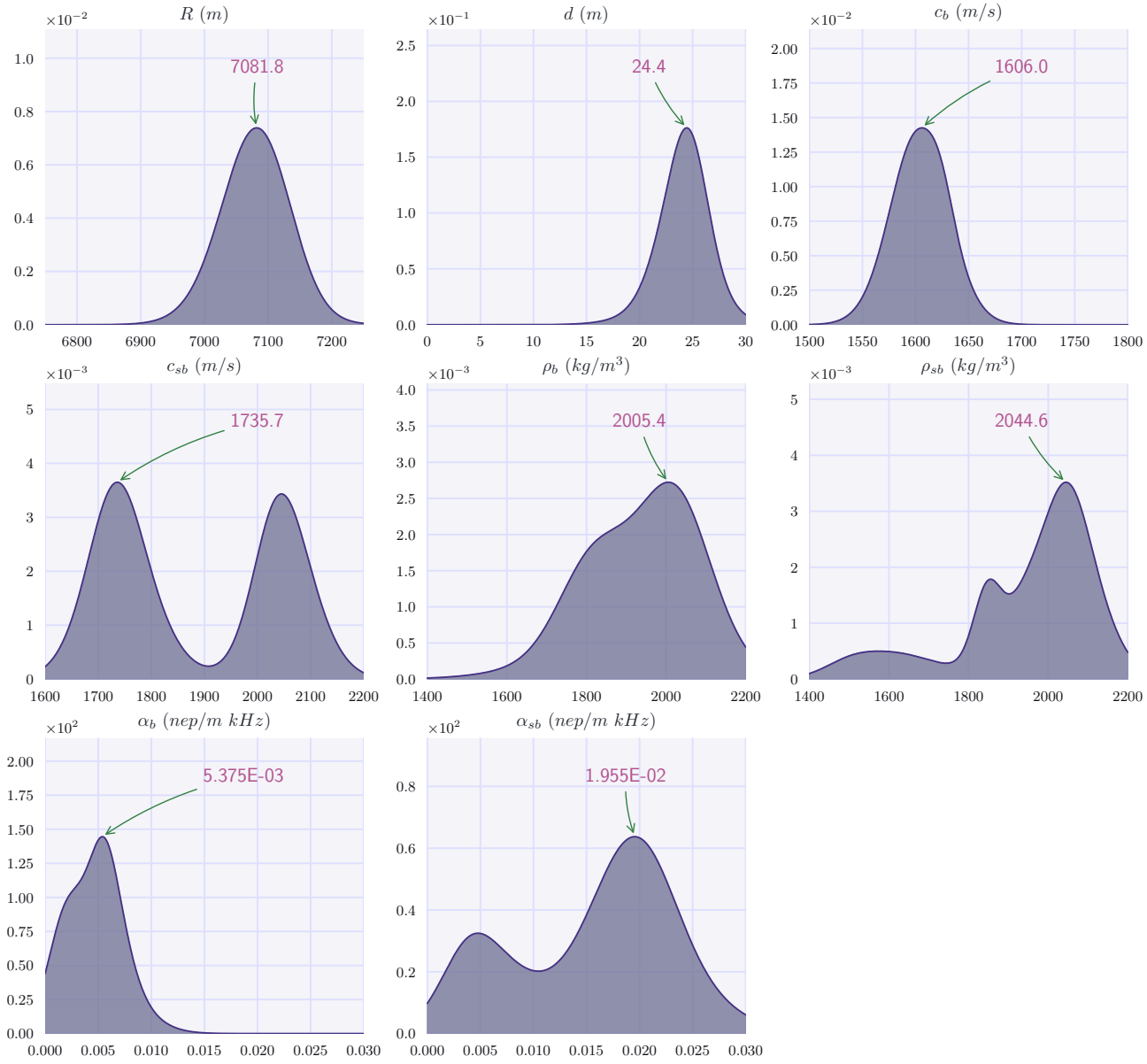


Figure 6.43: Marginal population distributions of the unknown parameters of the sea environment of the SW06 experiment using the **PSCS**. The distributions show statistics of the final population of GA.

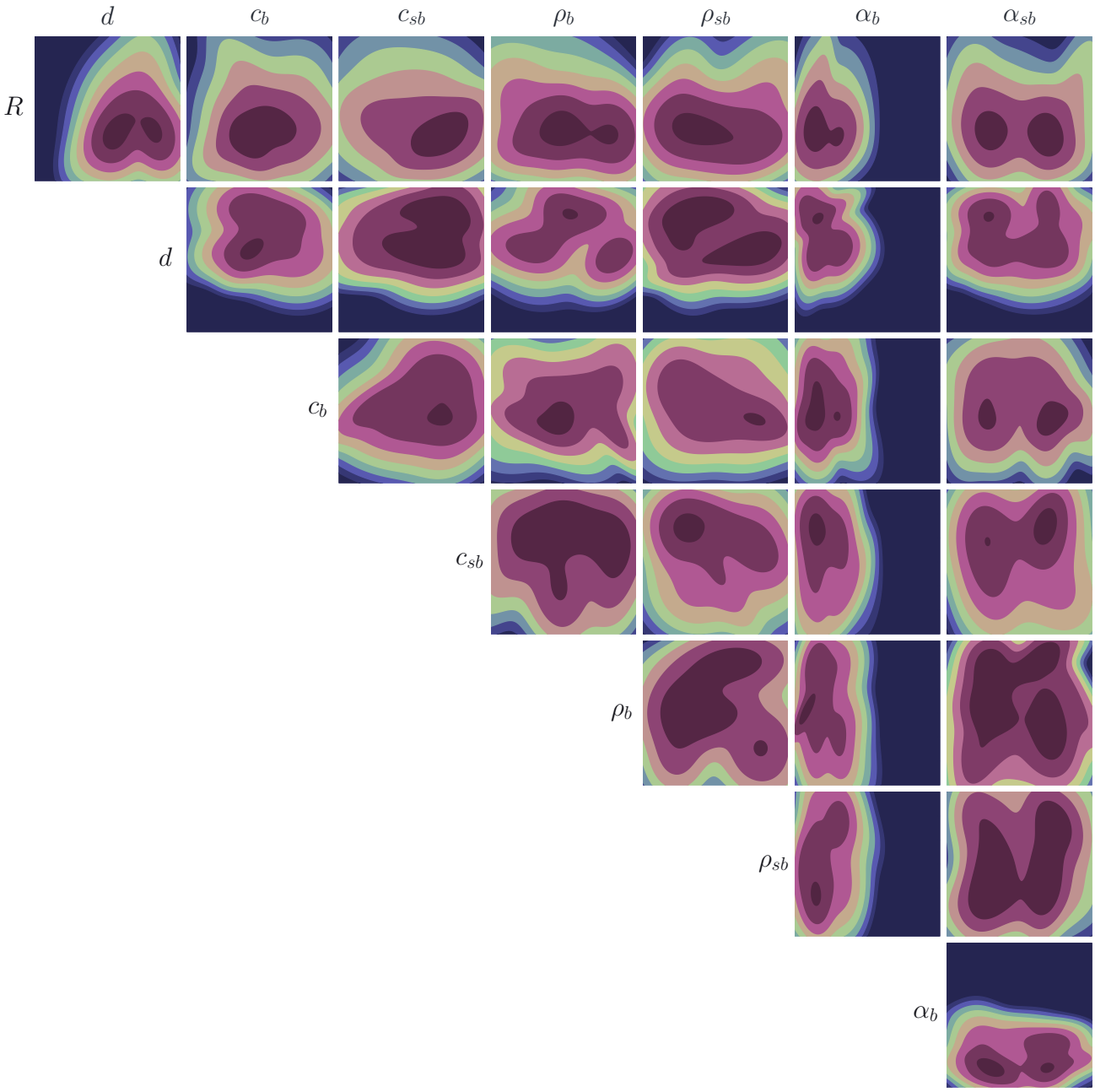


Figure 6.44: Joint population distributions corresponding to the inversion results using the SW06 data using the **SSCS**. The distributions show statistics of the final population of GA. Each distribution has been normalized independently.

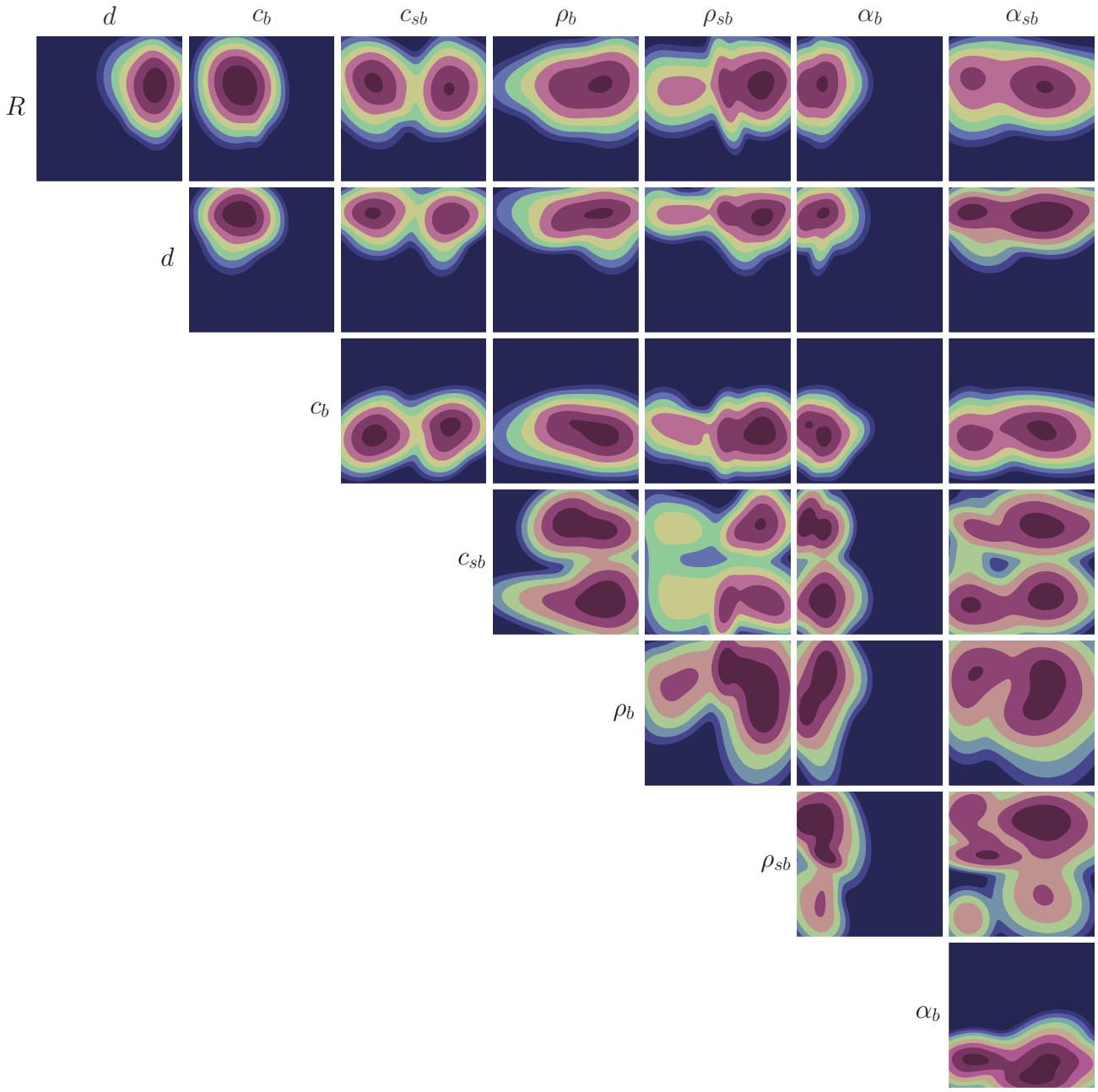


Figure 6.45: Joint population distributions corresponding to the inversion results using the SW06 data using the **PSCS**. The distributions show statistics of the final population of GA. Each distribution has been normalized independently.

6.3.2 Second formulation

The second formulation keeps the geometry of the previous one, but relaxes the assumption of the constant sound speed profile in the sediment layer by assuming that the sound speed in this medium is varying linearly with the depth.

The analytical form of the sediment sound speed profile is given by the formula:

$$c_b(z) = c_{b1} + s(z-h),\ z \in [h, h+d]. \tag{6.14}$$

where c_{b1} at the top of the sediment and s is the sound speed slope.

Therefore, the model is parameterized by the following vector $\mathbf{m}$ as:

$$\mathbf{m} = [R, d, c_{b1}, s, c_{sb}, \rho_b, \rho_{sb}, \alpha_b, \alpha_{sb}]^T. \tag{6.15}$$

The geoacoustic model is illustrated in Figure 6.46. Similarly to the previous example we derive the corresponding new marginal and joint distributions (Figures 6.48 and 6.47), whereas, Table 6.16 summarizes the optimal estimations using both inversion schemes.

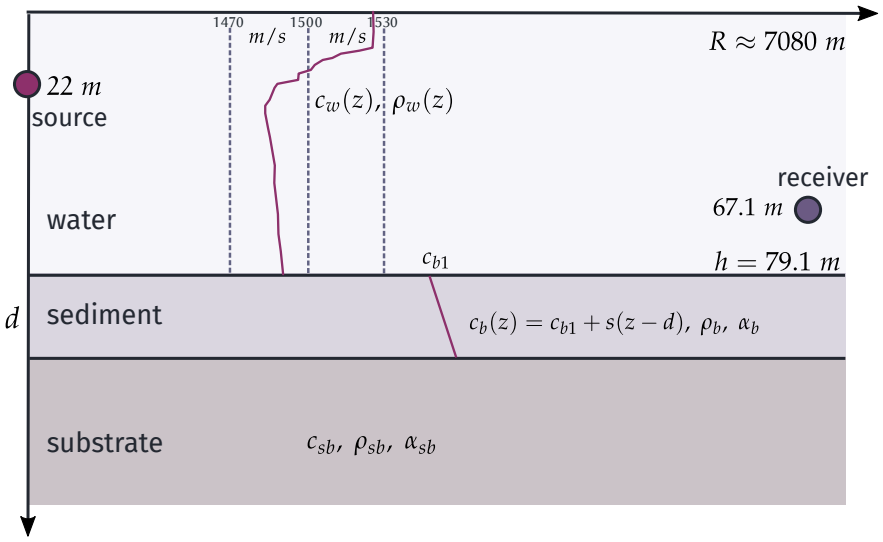


Figure 6.46: The range-independent environment that is considered quite close to the track of the Shallow Water 06 experiment corresponding to the second formulation.

Parameters	Units	Search Space	Estimated Value (PSCS)	Estimated Value (SSCS)
Range	$R\ (m)$	[6750, 7250]	7055.8	7011.8
Sediment thickness	$d\ (m)$	[0, 30]	24.9	18.6
Top sediment sound speed	$c_{b1}\ (m/s)$	[1500, 1800]	1633.9	1634.5
Sediment sound speed slope	$s\ (1/s)$	$[-5, 5]$	-2.53	0.33
Substrate sound speed	$c_{sb}\ (m/s)$	[1600, 2200]	1727.3	1887.1
Sediment density	$\rho_b\ (kg/m^3)$	[1400, 2200]	1771.6	1911.7
Substrate density	$\rho_{sb}\ (kg/m^3)$	[1400, 2200]	2094.3	1609.0
Sediment attenuation	$\alpha_b\ (nep/mkHz)$	[0, 0.03]	0.006637	0.005676
Substrate attenuation	$\alpha_{sb}\ (nep/mkHz)$	[0, 0.03]	0.019611	0.02076

Table 6.16: Estimated values of the recoverable parameters using the second formulation.

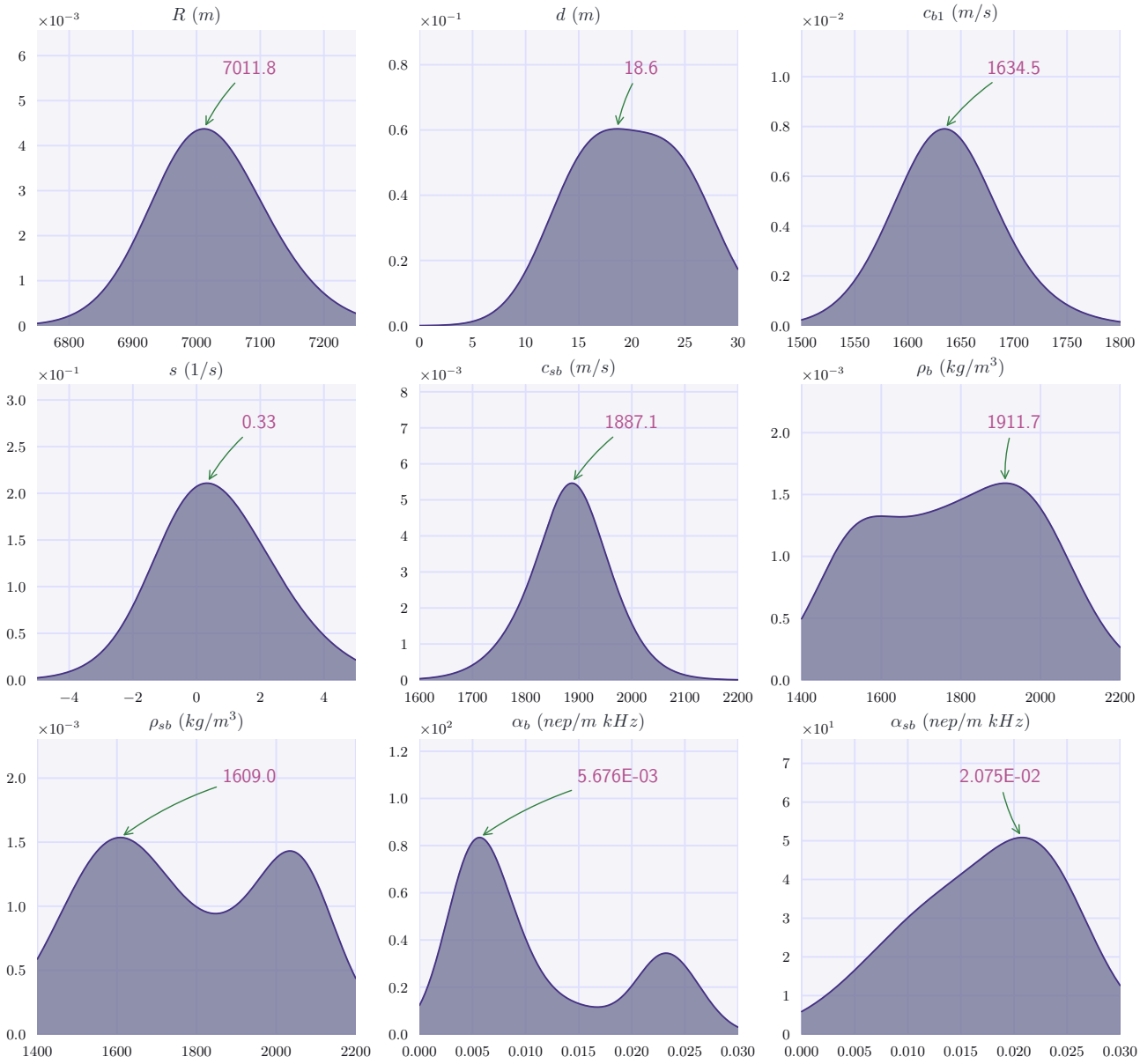


Figure 6.47: Marginal population distributions of the unknown parameters of the sea environment of the SW06 experiment using its range dependent formulation using the **SSCS**. The distributions show statistics of the final population of GA.

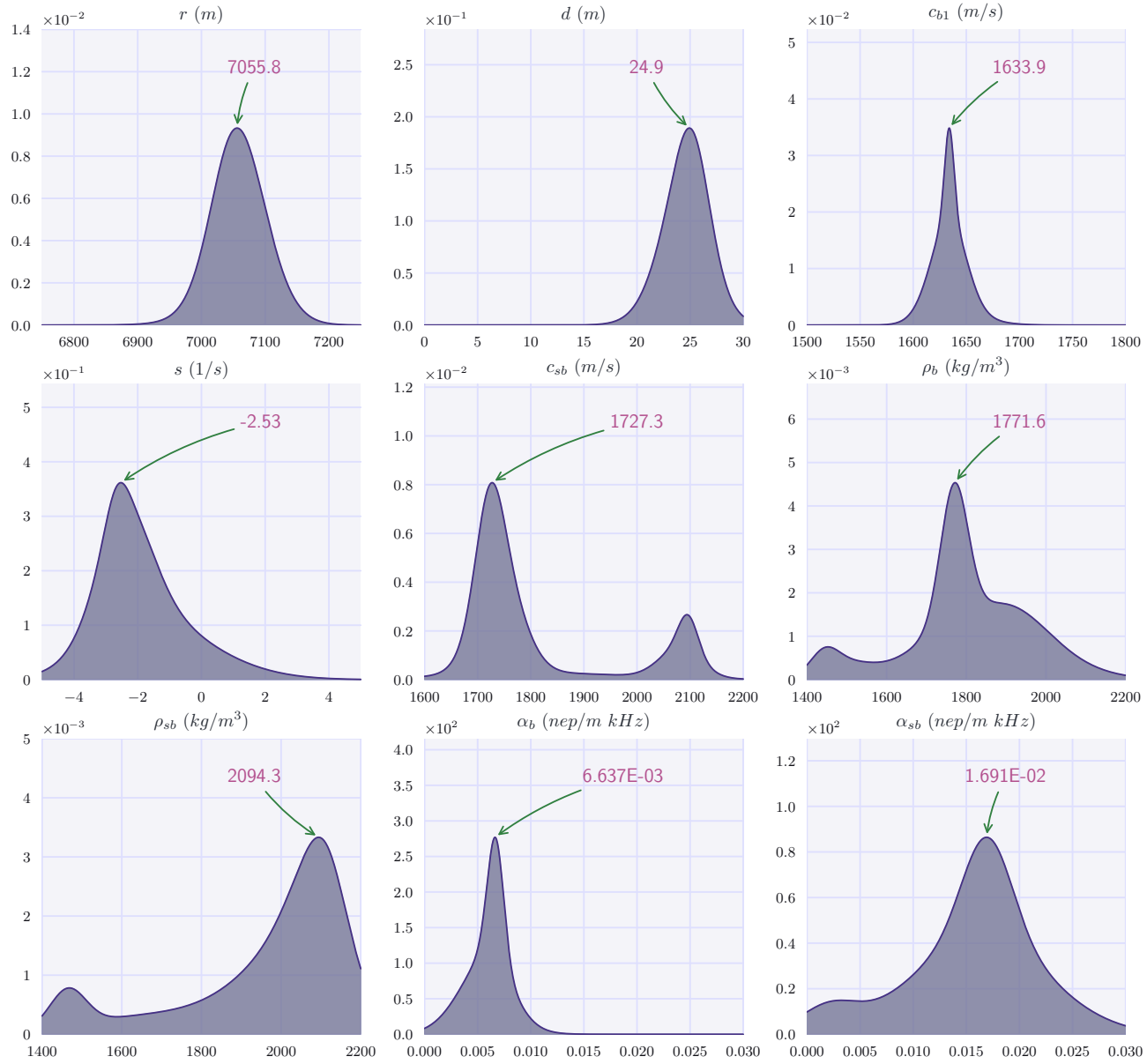


Figure 6.48: Marginal population distributions of the unknown parameters of the sea environment of the SW06 experiment using its range dependent formulation using the **PSCS**. The distributions show statistics of the final population of GA.

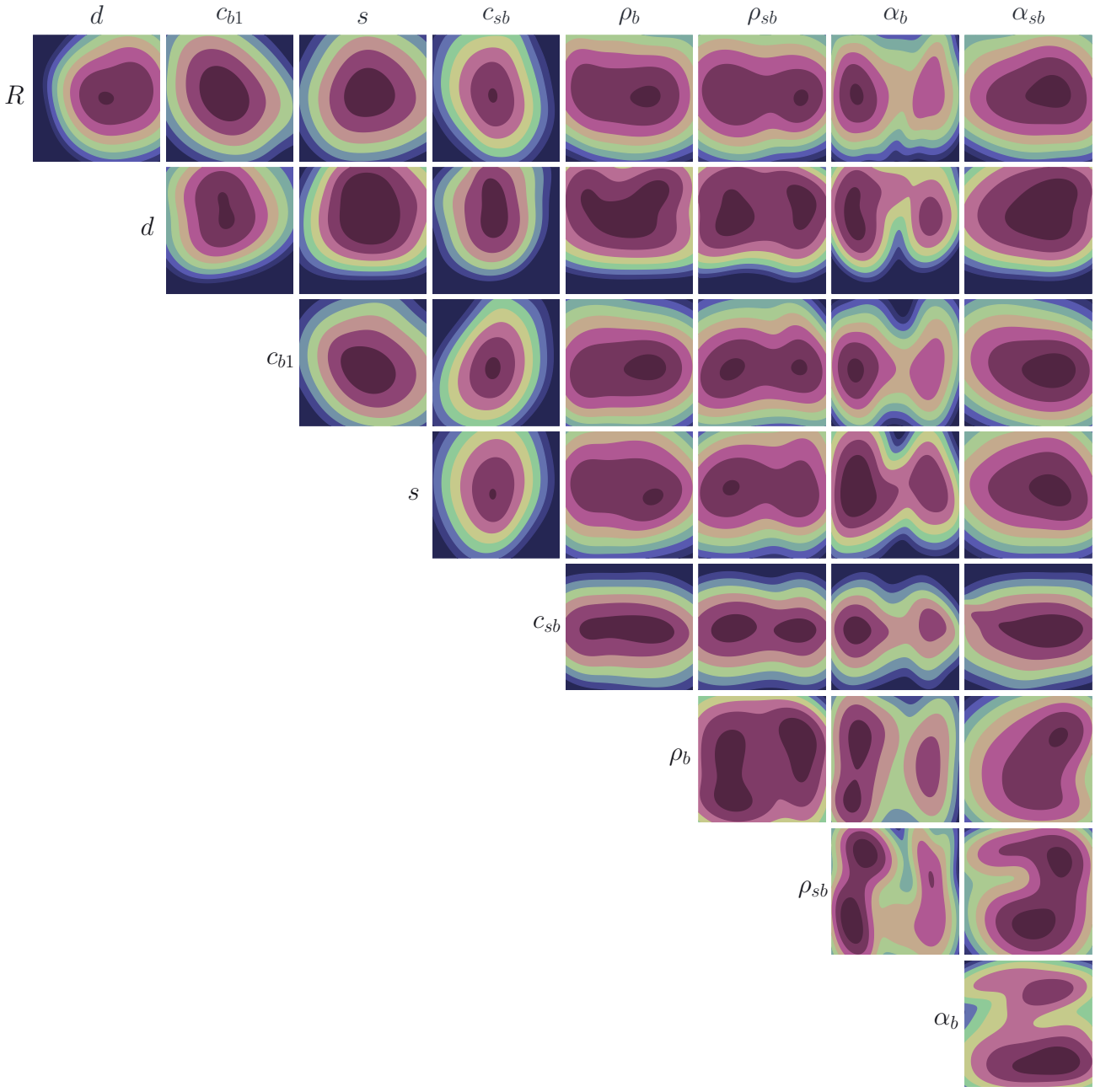


Figure 6.49: Joint population distributions corresponding to the inversion results using the SW06 data and the range-dependent formulation of the environment using the **SSCS**. The distributions show statistics of the final population of GA. Each distribution has been normalized independently.

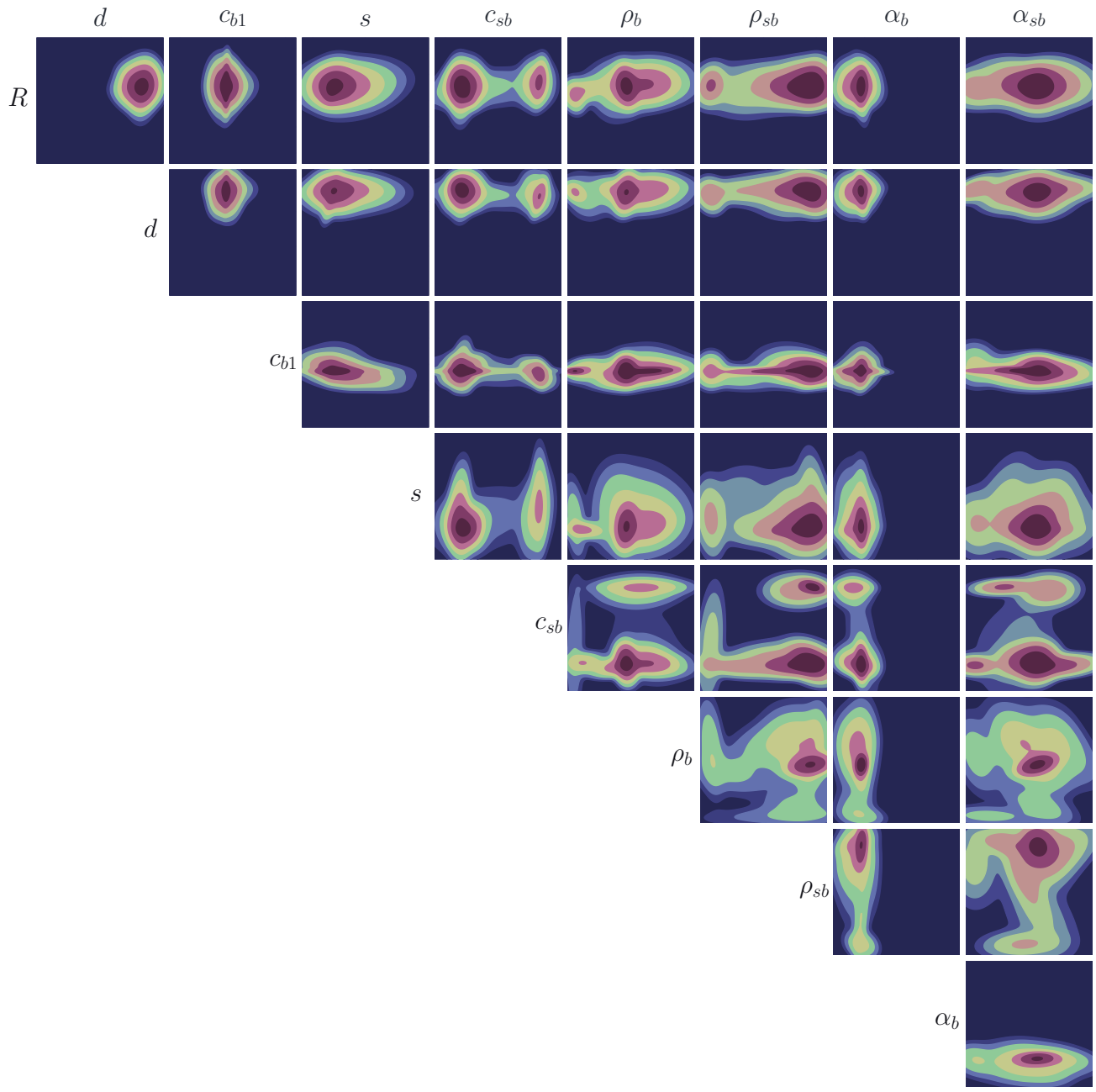


Figure 6.50: Joint population distributions corresponding to the inversion results using the SW06 data and the range-dependent formulation of the environment using the **PSCS**. The distributions show statistics of the final population of GA. Each distribution has been normalized independently.

6.3.3 Third formulation

The last formulation, to our knowledge at least, consists of the first modelling of the SW06 experiment using a range-depended environment or to be more precise a piecewise range-independent description of the actual source-receiver track. We follow the same parameterization as the second formulation but we further assume that the water depth is not constant but varies every 1 km.

The goal of this approach is to study if by using a modelling of the environment closer to the actual one, we would get better inversion results. To achieve a good representation of the seabed, we make use of the chirp sonar [58] data obtained during the experiment (Figure 6.51). To keep the number of the recoverable parameters unchanged, we additionally assume that the thickness of the sediment is the same at each region.The whole setup appears in Figure 6.52.

For this case the coupled normal mode program MODE4 developed at FORTH, capable of treating range-depended problems of are used to obtain the replica fields.

Table 6.17 contains the estimated values of the recoverable parameters using inversion scheme based on the two characterization methods. Furthermore, Figures 6.54 to 6.56 present the marginal and joint distributions obtained by utilized the statistical and probabilistic characterization schemes.

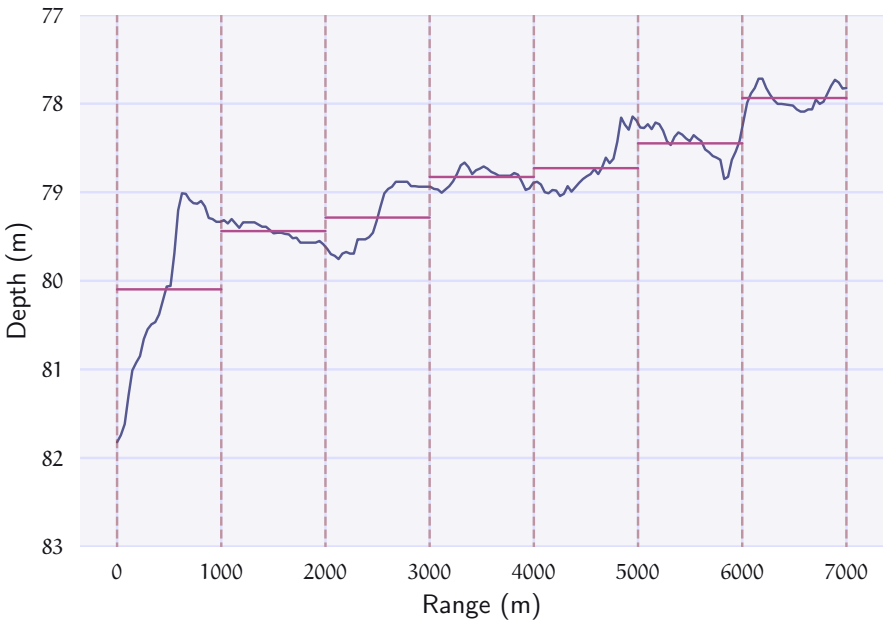


Figure 6.51: Water depth estimated by chirp sonar data along the source/receiver track obtained during the experiment. Here, we consider seven range independent segments with depths marked by the purple horizontal lines.

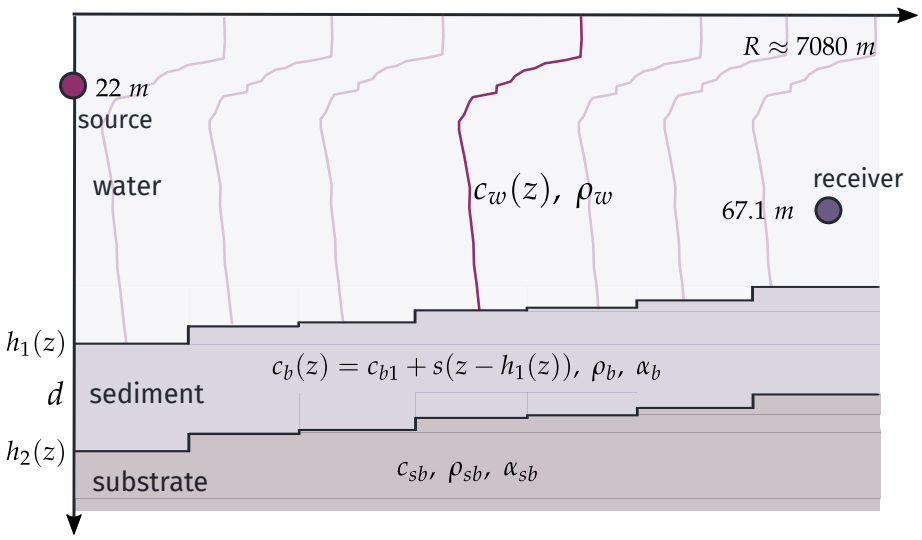


Figure 6.52: The considered range-dependent environment of the SW06 experiment corresponding to the third formulation.

Parameters	Units	Search Space	Estimated Value (PSCS)	Estimated Value (SSCS)
Range	$R\ (m)$	[6750, 7250]	7054.3	6960.7
Sediment thickness	$d\ (m)$	[0, 30]	25.0	24.7
Top sediment sound speed	$c_{b1}\ (m/s)$	[1500, 1800]	1631.8	1636.6
Sediment sound speed slope	$s\ (1/s)$	[−5, 5]	−2.58	−0.37
Substrate sound speed	$c_{sb}\ (m/s)$	[1600, 2200]	1724.3	1918.9
Sediment density	$\rho_b\ (kg/m^3)$	[1400, 2200]	1791.6	1645.8
Substrate density	$\rho_{sb}\ (kg/m^3)$	[1400, 2200]	2096.7	1926.1
Sediment attenuation	$\alpha_b\ (nep/mkHz)$	[0, 0.03]	0.006156	0.003574
Substrate attenuation	$\alpha_{sb}\ (nep/mkHz)$	[0, 0.03]	0.01721	0.008348

Table 6.17: Estimated values of the recoverable parameters using the third RD formulation.

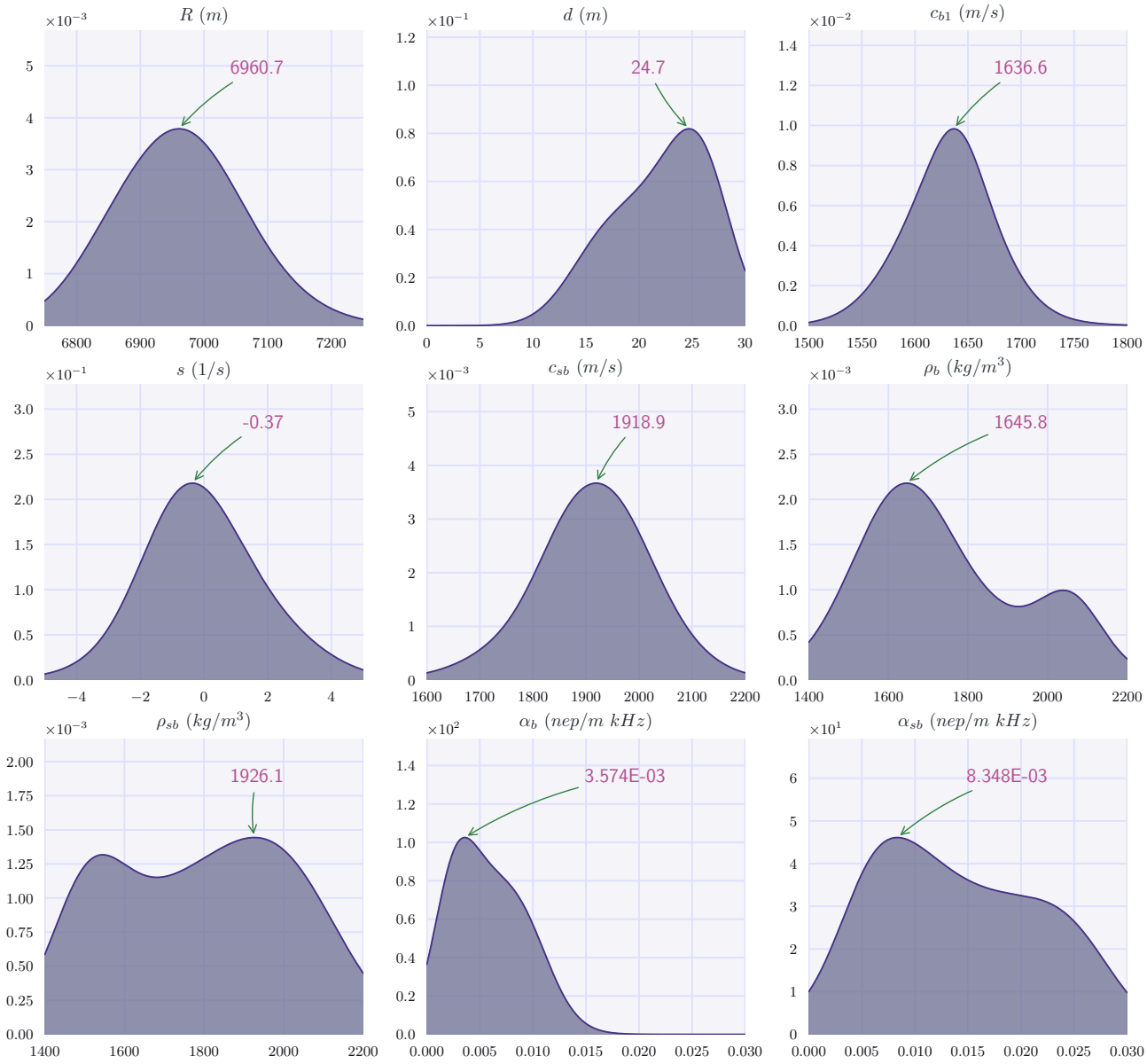


Figure 6.53: Marginal population distributions of the unknown parameters of the sea environment of the SW06 experiment using its range dependent formulation through the **SSCS**. The distributions show statistics of the final population of GA.

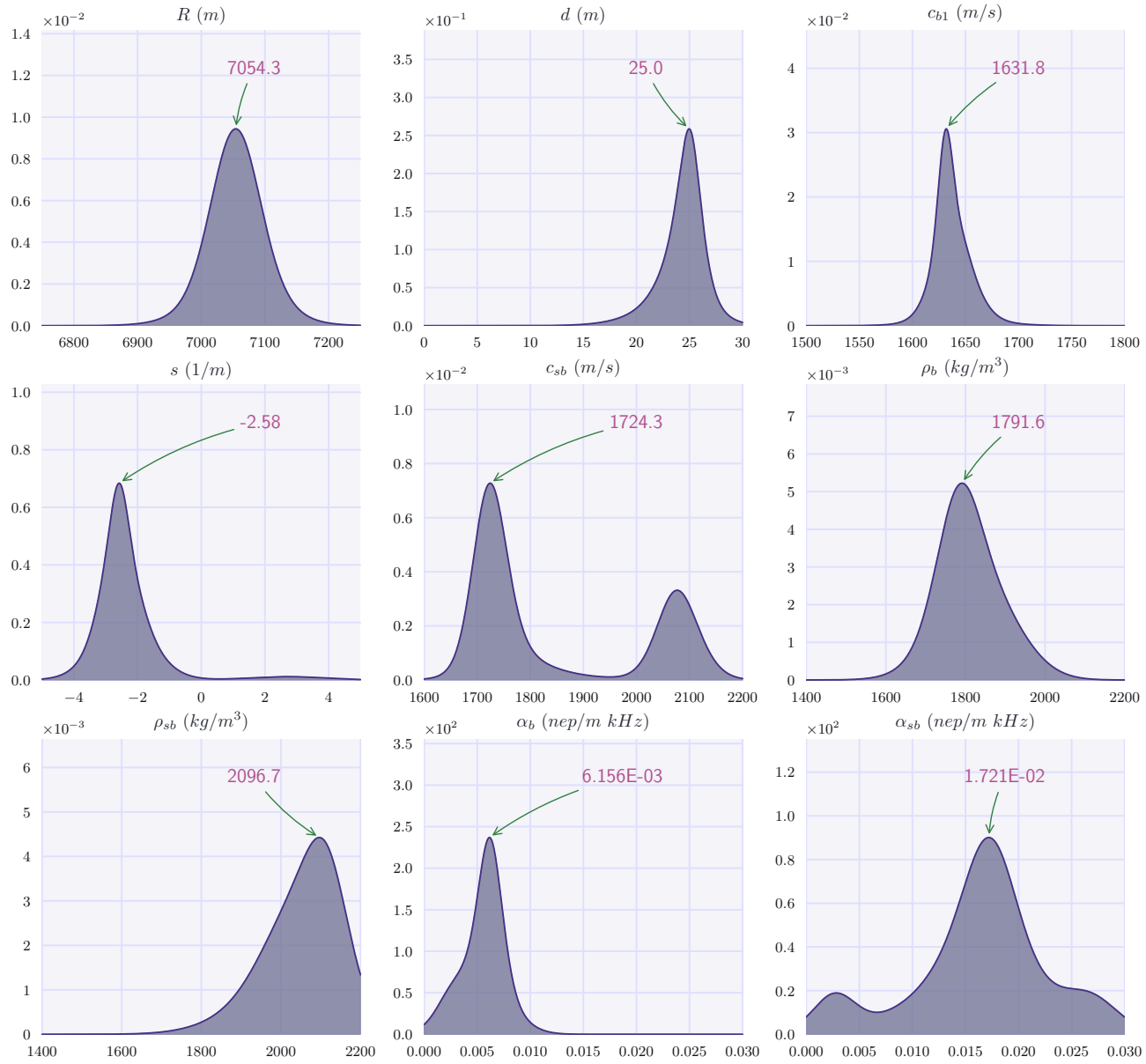


Figure 6.54: Marginal population distributions of the unknown parameters of the sea environment of the SW06 experiment using its range dependent formulation through the **PSCS**. The distributions show statistics of the final population of GA.

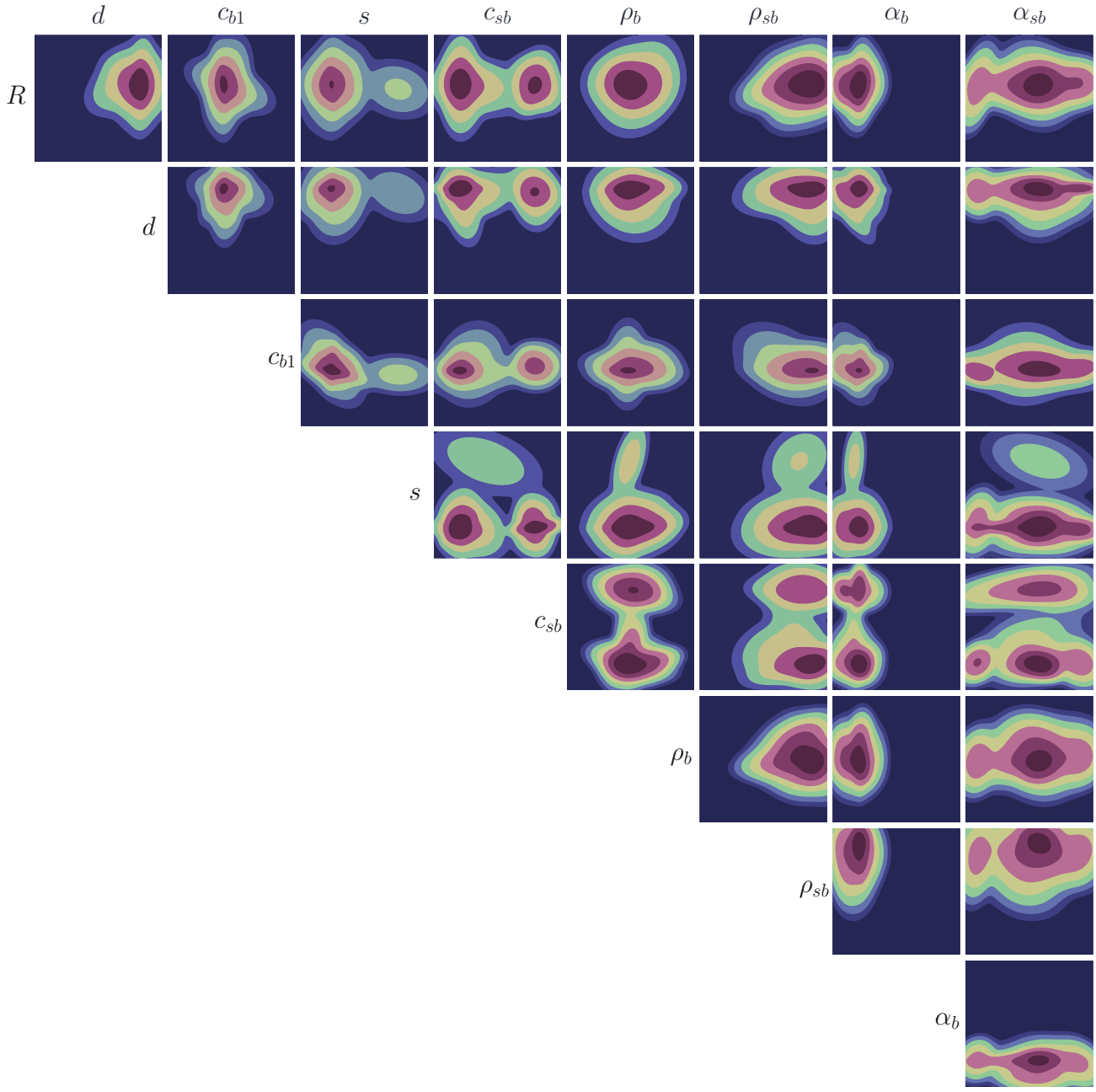


Figure 6.55: Joint population distributions corresponding to the inversion results using the SW06 data and the range-dependent formulation of the environment through the **SSCS**. The distributions show statistics of the final population of GA. Each distribution has been normalized independently.

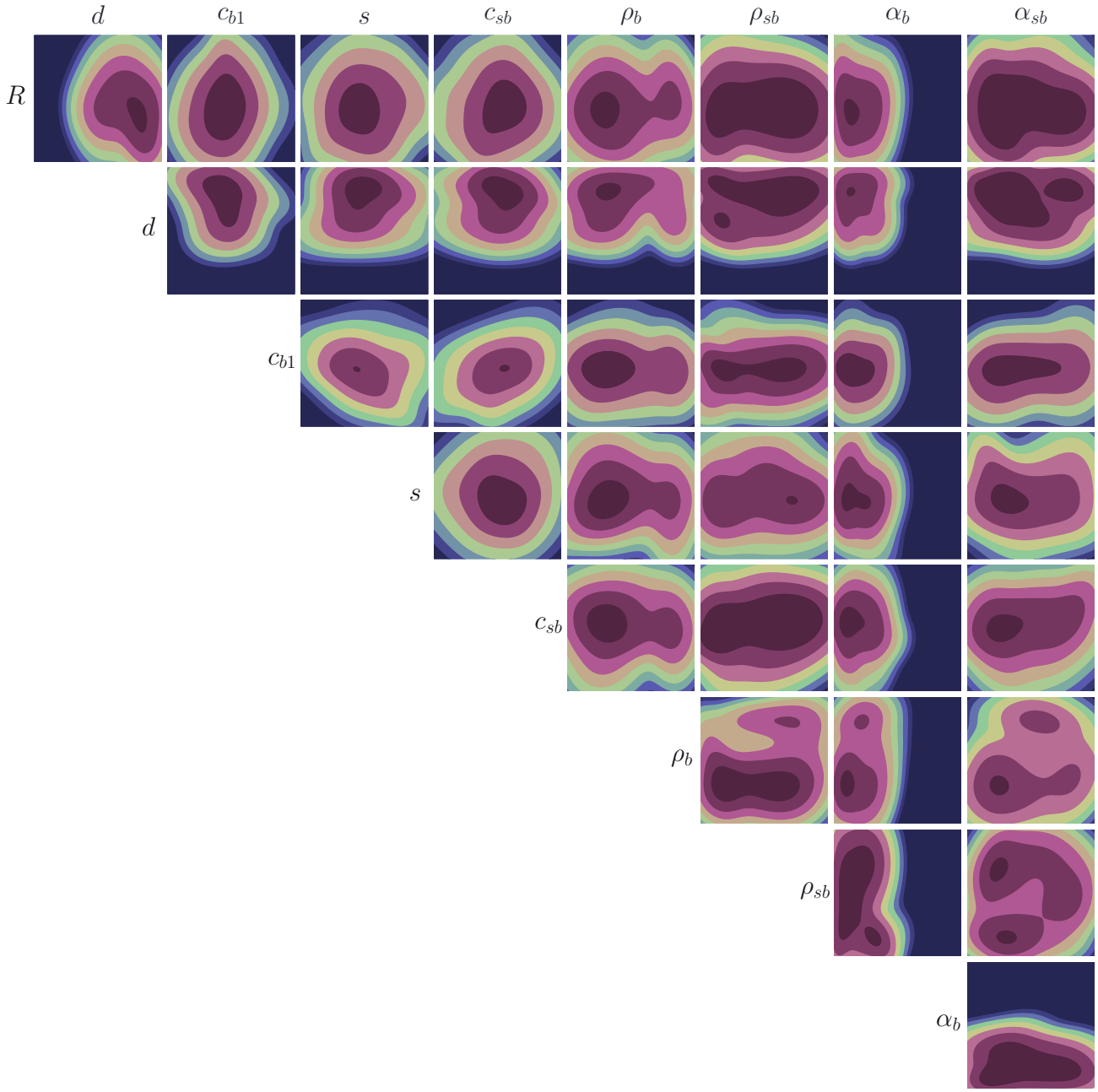


Figure 6.56: Joint population distributions corresponding to the inversion results using the SW06 data and the range-dependent formulation of the environment through the **PSCS**. The distributions show statistics of the final population of GA. Each distribution has been normalized independently.

6.3.4 Evaluation of the inversion results

The evaluation of the quality of the inversion results, using the experimental data, will be based on three different criteria:

1. Considering the comparisons between actual and recovered (simulated) signals.
2. Comparisons of the results with those reported by other researchers using alternative techniques.
3. Taking into account geological observations of the environmental properties produced by measurements using mechanical means (ground true).

By comparing the simulated waveforms using the inversion results of the two methods and the MODE1 normal mode program for the calculation of the synthetic acoustic fields (Figure 6.57), it can be noticed that the differences are not significantly large, especially for the earlier times that are characteristic of the propagation of the lower order modes. This is an interesting result, suggesting that the inversion schemes based on the statistical and probabilistic characterization of the recorded signal can give estimations of the model parameters that reproduce the actual measured signal in a very good quality. Moreover, an additional positive observation is the better fitting among the actual (denoised) and reproduced signals, when the more general formulation of the sea environment is considered.

Table 6.18 presents a comparison of the model parameters estimated in this work (six first rows) with estimations obtained by other studies (seventh to eleventh row) as well older published results using the statistical signal characterization scheme which however was based on slightly different modelling of the sound speed profile. In addition we present reference values from Bonnel and Chapman [57] which are referred to typical values for the type of the sediment layers existing in the area of experiment. When two values are separated by a slash, the modeling of the forward propagation problem assumes a linear sound speed profile in the sediment, where the value before the slash denotes the estimation of the sound speed at the top of the sediment layer, whereas the value after the slash the sentiment sound speed at the sediment-substrate interface. Note that it is very difficult to assess the reliability of our proposed scheme by means of the results obtained by other researchers as there is no definite knowledge of the parameters of the sea-bed. Overall, the value of the sound speed in the sediment layer estimated by our proposed scheme was very similar to that estimated by other

researchers. In particular, the PSCS method returns sound speed profile in the sediment which is very close to the ones given by Jiang (IJC-2008) et al. [132] and Tan et al. (TGYH-2014) [133].

Also, a geological survey in the area of the experiment by Goff et al. [134] reported an estimation of the sound speed in the substrate of 1728 m/s which is very close to the estimated value by the PSCS scheme. Large variations of the values for the sound speed in the substrate estimated by the various methods are reported. This can be explained by the fact that the substrate does not affect too much the sound field measured in the water column.

Finally, geological studies presented in the literature generally agree with the description of the sound speed profile in the sediment layer with a negative slope. Following these studies the negative slope can be explained by the possible existence of a different sediment material in the interior of the layer in which the sound is propagated slower. Moreover, the negative recovered sound speed profile is consistent with chirp sonar data [135], sample data and in-situ probes [136, 137] and shallow core measurements [134]. The agreement of the values obtained by the PSCS method with reference values is noticeably.

Study	$d\ (m)$	$c_b\ (m/s)$	$\rho_b\ (kg/m^3)$	$c_{sb}\ (m/s)$	$\rho_{sb}\ (kg/m^3)$
PSCS (RI)	24.4	1606.0	2005.4	1735.7	2044.6
PSCS (RI-slope)	24.9	1633.9/1570.9	1771.6	1727.3	2094.3
PSCS (RD-slope)	25.0	1631.8/1567.3	1791.6	1724.3	2096.7
SSCS (RI)	17.2	1636.3	1783.6	2019.8	1663.5
SSCS (RI-slope)	18.6	1634.5/1640.6	1911.7	1887.1	1609.0
SSCS (RD-slope)	24.7	1636.6/1627.5	1645.8	1918.9	1926.1
JC-2008 (RI-slope)	21.1	1636.6/1579.8	1680.0	1740.5	-
BC-2011 (RI)	26.9	1603.0	1890.0	2199.0	2280.0
BNGM-2012 (RI)	26.5	1621.0	1660.0	1858.0	2470.0
BDC-2013 (RI)	25.0	1604.0	1800.0	2132.0	1480.0
TGYH-2014 (RI-slope)	21.5	1650.0/1562.6	2100.0	1993.0	-
SSCS-2014 (RI)	22.3	1590.0	1916.4	1746.8	2054.0
SSCS-2017 (RI-slope)	25.3	1638.4	1575.3	1938.4	2214.9
Reference values	20.0	1600.0	1800.0	1700.0	2100.0

Table 6.18: Geoacoustic inversion results that refer to the SW06 experiment. JC-2008 [132], BC-2011 [57] ,BNGM-2012 [138], BDC-2013 [58], TGYH-2014 [133], SSCS-2014 [106] and SSCS-2017 [129].

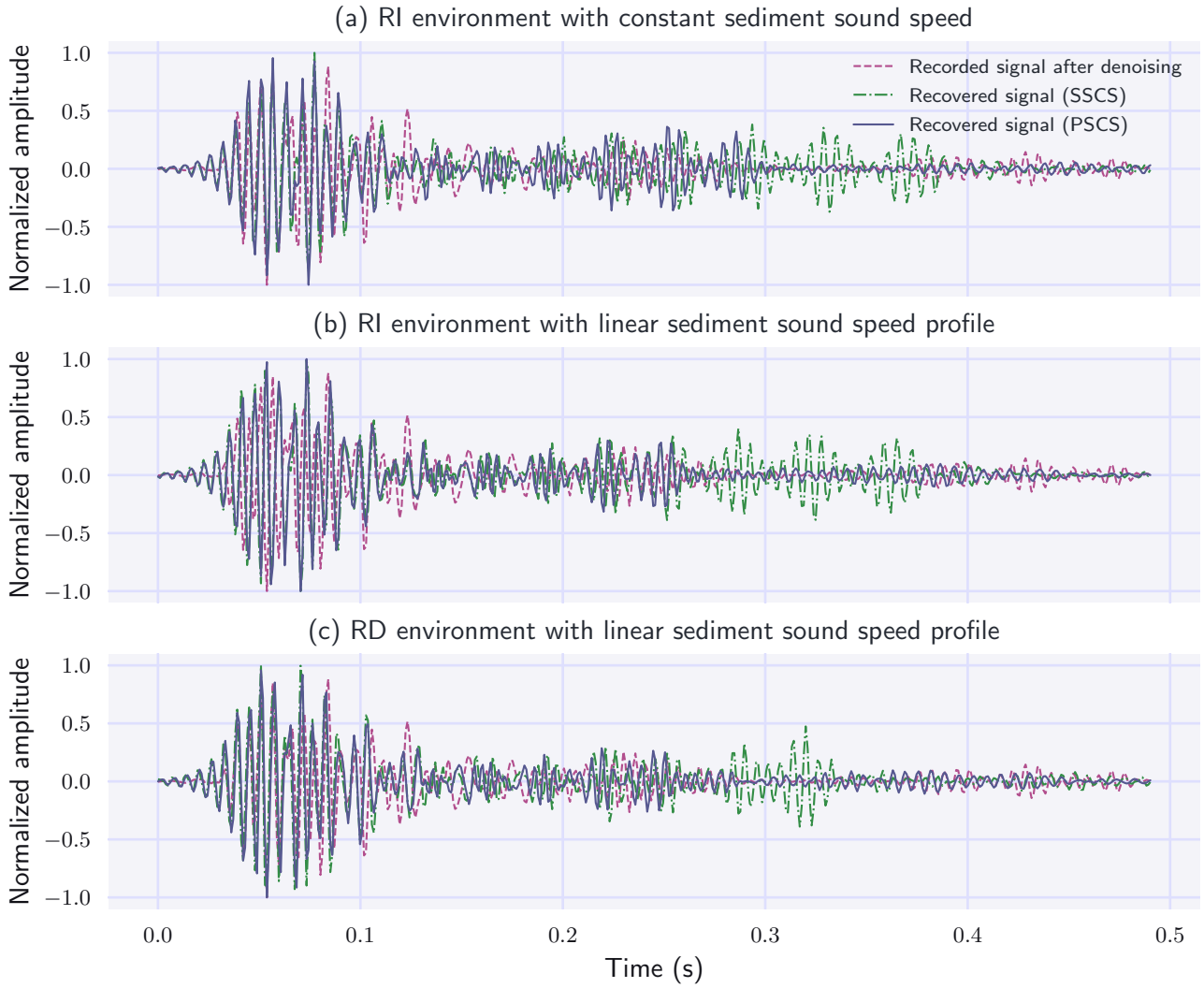


Figure 6.57: Recorded and recovered signals in the time domain for the three inversion test case of the sw06 experiment.

chapter 7

CONCLUSIONS

7.1 Conclusions

In this thesis a new signal characterization method based on Hidden Markov Models (HMMs) was presented. The proposed method can in principle be applied to any form of signal. However, only characterizations of the scheme in acoustical signals associated with application of acoustical oceanography were presented here. The proposed signal characterization method follows previous attempts to characterize acoustic signals by statistical means.

Thus, the main contribution of the work presented here is the introduction of a novel time-invariant probabilistic characterization scheme for acoustic signals. The proposed approach assigns the stationary wavelet packet coefficients of a signal to a Hidden Markov Model (HMM) Providing a novel approach for signal feature extraction.

In addition, in order that the characterization of the signal is reliable, a new denoising scheme using dictionary learning and sparse decomposition of the raw signals was introduced. This denoising scheme also improves the quality of the signal characterizations obtained by the Statistical Signal Characterization Scheme (SSCS) also considered in this thesis.

The new method was Applied with data taken from simulated or real experiments related to problems of acoustical tomography of geacoustic inversions. A GA has been employed to drive the optimization procedure, using the Kullback-Leibler divergence between the synthetic or measured signal and replica signals as the objective (fitness) function.

It was shown that good signal characterization and subsequent inversion results are obtained when the signal denoising scheme is applied. Following the analysis of the results presented

here the PSCS scheme shows better behaviour over the SSCS in particular all applications studied. In particular, the values corresponding to the maxima of the marginal probability densities obtained through the GA in all the cases studied which were considered as the estimated values of the recoverable parameters were closer to the reference ones of the actual marine environment in relation to those obtained by SSCS method which has been applied with the same optimization tool (GA). Another advantage of the new technique compared to the SSCS is that the probability densities gave narrower confidence intervals.

Although the new PSCS gave better results than the SSCS, this method is considerably more computationally demanding. To ensure the efficient application of the PSCS characterization scheme at a reasonable time, we developed a fully-parallelized implementation of HMM. Also, the optimization procedure using GA was carried out in parallel for both inversion schemes.

Overall, we consider that the new scheme developed for the acoustic signal characterization and its use for tomographic and geoacoustic inversions can be considered as a promising new tool for the applications of the acoustical oceanography related to the monitoring of the marine environment with potentially more applications in other fields requiring signal characterizations.

7.2 Future Directions

In this section we discuss some future research plans relative to the work presented here. We will split this section in two parts:

7.2.1 Possible characterization improvements

— Best wavelet packet decomposition —

So far, we have characterized a given acoustic signal in terms of a time-frequency decomposition using the stationary wavelet packets and keeping features from the last level of the signal's decomposition. Recall that in all characterization cases introduced in this work the maximum level of decomposition was four.

An optimum wavelet packet divides the time-frequency space in such a way that the resulted sub-tree are best adapted in order to approximate a given signal.

In view of this we plan to study an optimum decomposition employing a criterion that introduces the decomposition entropy. The best basis would be occurred by the fast dynamic programming algorithm. Motivated by the work presented by Coifman and Wickerhauser [139].

— Alternative emission distributions —

In this thesis the emission probabilities consist of Gaussian distributions. We believe that by using a more flexible form of emission distributions will achieve a more accurate characterization of acoustic signals.

7.2.2 Characterization of seismic signals

Following the successful characterization of underwater acoustic signals, we would like to apply the proposed probabilistic feature extraction procedure to seismic waves recorded by typical seismic observatories.

In order to characterize seismic recordings we will follow the steps mentioned below:

- We will use a suitable time-frequency feature extraction mechanism, using historical data to find the optimal wavelet decomposition tree, which is the tree that minimizes the decomposition entropy as presented above.
- Assuming that these coefficients illustrate some specific structures varying with time, the extracted wavelet packet coefficients of a seismic signal will be modelled by a proper left-to-right hidden Markov model (HMM).
- The association of the extracted time-frequency features with the HMM will be achieved, after performing a training process.

Therefore, the final features of the seismic signal will consist of the parameters describing the HMM. Following the work proposed in this thesis, the Kullback-Leibler Divergence (KLD) will be used, in order to quantify the relative entropy between two HMMs, calculated numerically with the Monte-Carlo (MC) technique. Therefore, considering the associated HMMs of two seismic recordings, KLD can be considered was a similarity measure of them. Our goal will be to study seismic sequences in specific locations as a novel contributions of seismic monitoring.

A

Derivation of the Light-Bulb Implosion Source Excitation Function

Raw data of the light bulb explosion implies that the effective bandwidth of the acoustic signals was from 30 to 200 Hz. In the SW06 experiment, a light-bulb implosion was set for a depth of 22 m using a messenger weight to break the bulb.

The implosion of the light bulb generates a signal ion the water which consists of two parts. The first part is characterized by a negative pressure phase because of a decrease in pressure during the collapse of the light bulb surface. This effect is a consequence of the instantaneous water flow into the light bulb volume. The second and the most significant part is a damped sine wave as a result of the oscillation of the light bulb gas volume in the water. After the implosion, the gas forms a spherical void in the water with an initial radius. The void collapses due to the hydrostatic pressure to a minimum value and when the internal pressure overtakes the external one, the void starts to increase its volume. This process continues and a damped sine waveform is generated and transmitted in the water.

A damped sine waveform has the analytic form

$$S(t) = AH(t)\exp(-\alpha t)\sin(\omega_c t). \tag{A.1}$$

where A is a normalization constant, H is the Heaviside function, ω_c is an angular frequency of the pulse, and α is the decay parameter.

All these parameters should be estimated in order to model the actual excitation function of the experimental light bulb source.

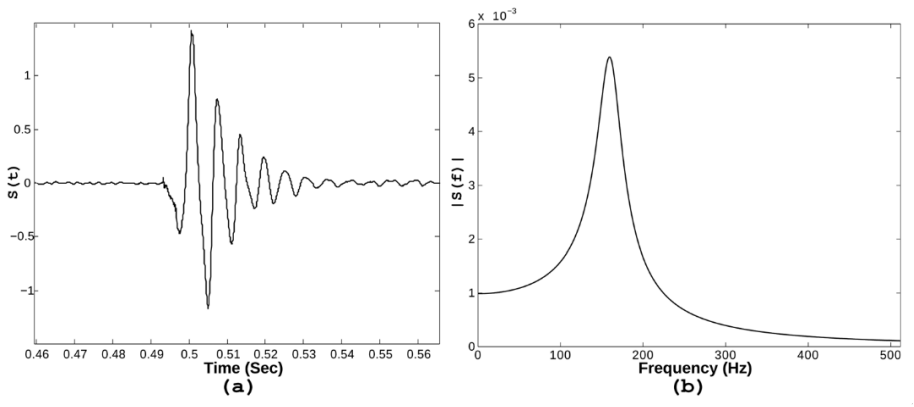


Figure A.1: (a) The actual recorded signal close to the light-bulb source. (b) The simulated source excitation function in the frequency domain.

The ω_c can be estimated by measuring the time difference between two peaks of consecutive positive pulses and then by taking its inverse. For the signal illustrated in Figure A.1 (a), this value is 320π rad/s.

The decay parameter is estimated by measuring the amplitude of two consequent peaks $S(t_1)$ and $S(t_2)$ of the signal given the times corresponding to these peaks. In example, if we know the pair of coordinates $[t_1, S(t_1)]$ and $[t_2, S(t_2)]$ then the α parameter takes the value

$$\alpha = \frac{1}{t_2 - t_1} \ln \frac{S(t_1)}{S(t_2)}. \tag{A.2}$$

For our data, the proper value of α is equal to 92.8 rad/s. Note that the estimation of the decay parameter can be found for any of consecutive peaks without any significant difference. The modeled source is written in the frequency domain as follows:

$$S(\omega) = \frac{\omega_c}{(\alpha + i\omega)^2 + \omega_c^2}, \tag{A.3}$$

where ω denotes the angular frequency. Figure A.1 (b) illustrates the marginal spectrum of the source excitation function for frequency in the range $[0, 512]$ Hz.

B

ALGORITHMS

B.1 ISTA Algorithm

Given a vector $\mathbf{s} \in \mathbb{R}^L$ and a matrix (dictionary) $\mathbf{D} \in \mathbb{R}^{L,M}$ we introduce the following optimization problem:

$$\arg\min_{\mathbf{a}\in\mathbb{R}^M} \left\{ \frac{1}{2}\|\mathbf{s} - \mathbf{D}\mathbf{a}\|_2^2 + \lambda\|\mathbf{a}\|_1 \right\}, \tag{B.1}$$

where λ a positive real number.

The scope of this problem is to represent the vector $\mathbf{s}$ as good as possible using only few columns of the given dictionary matrix $\mathbf{D}$. The number of the used columns of $\mathbf{D}$ is controlled by the so called sparsity term λ.

In the applications of this thesis we deal with such problems by employing the Iterative Shrinkage and Thresholding Algorithm (ISTA).

Each iteration of the ISTA generalized gradient descent algorithm includes two phases. First, it performs a single step gradient descent procedure to the reconstruction error term $\frac{1}{2}\|\mathbf{s} - \mathbf{D}\mathbf{a}\|_2^2$ and then performs an update based on the sparsity term $\lambda\|\mathbf{a}\|_1$ using the proximal mapping theory. The i-th element of the proximal mapping of the sparsity term calculated at the coefficient vector $\mathbf{a}$ has been proven to be [140]

$$\text{prox}_h(\mathbf{a})_i = \begin{cases} (\mathbf{a})_i - \lambda \,\text{sign}((\mathbf{a})_i - \lambda) & \text{if } |(\mathbf{a})_i| > \lambda \\ 0 & \text{otherwise.} \end{cases} \tag{B.2}$$

The whole procedure is described in the following algorithm:

Algorithm 1

procedure ISTA($\mathbf{D}$,$\mathbf{s}$) ▷ Required the dictionary.
  Initialize $\mathbf{a}^0$ ▷ Randomly chosen
  $k \leftarrow 0$
  repeat
    $\tilde{\mathbf{a}}^{k+1} \leftarrow \mathbf{a}^k - t\,\mathbf{D}^T(\mathbf{D}\mathbf{a}^k - \mathbf{s})$ ▷ update from reconstruction term
    $\mathbf{a}^{k+1} \leftarrow \text{prox}_h(\tilde{\mathbf{a}}^{k+1})$ ▷ update from the sparsity term
    $k \leftarrow k + 1$
  until $\|\mathbf{a}^{k+1} - \mathbf{a}^k\|_2 < \text{tol}$
  $\hat{\mathbf{a}} \leftarrow \mathbf{a}^{k+1}$
  return $\hat{\mathbf{a}}$

B.2 Dictionary Learning Algorithm

The following algorithm is used in connection with the ISTA algorithm (B.1) for deriving a proper dictionary matrix $\mathbf{D}$ to be capable of representing a set of M vectors $\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_M\}$.

Algorithm 2

procedure DICTLEARNING
  $\mathbf{A} \leftarrow 0^{M,M}$, $\mathbf{B} \leftarrow 0^{L,M}$ ▷ No prior information
  $k \leftarrow 0$
  for $t = 1,\ T$ **do**
    Draw $\{\mathbf{s}_{t_i}\}_{i \in [1,\ \eta] \cap \mathbb{Z}}$ from the set $\{\mathbf{s}_j\}_{j \in [1,\ M] \cap \mathbb{Z}}$
    for $i = 1,\ \eta$ **do**
      $\hat{\mathbf{a}_{t_i}} = \mathrm{ISTA}(\mathbf{D}, \mathbf{s}_{t_i})$, ▷ using **Algorithm 1**
    if $t < \eta$ **then**
      $\theta \leftarrow t\eta$
    else
      $\theta \leftarrow \eta^2 - \eta + t$
    $\beta \leftarrow (\theta - \eta + 1)/(\theta + 1)$
    $\mathbf{A} \leftarrow \beta\mathbf{A} + \sum_{i=1}^{\eta} \hat{\mathbf{a}}_{t_i} \hat{\mathbf{a}}_{t_i}^T$,
    $\mathbf{B} \leftarrow \beta\mathbf{B} + \sum_{i=1}^{\eta} \mathbf{s}_{t_i} \hat{\mathbf{a}}_{t_i}^T$,
    for $j = 1,\ N$ **do**
      $\mathbf{u}_j \leftarrow \frac{1}{\mathbf{A}_{jj}}(\mathbf{b}_j - \mathbf{D}\mathbf{a}_j) + \mathbf{d}_j$
      $\mathbf{d}_j \leftarrow \frac{1}{\|\mathbf{u}_j\|_2}\mathbf{u}_j$ ▷ Update the j column of dictionary
  return D ▷ Return the trained dictionary

The corresponding optimization problem is expressed as:

$$\min_{\mathbf{D}} \frac{1}{M} \sum_{j=1}^{M} \min_{\mathbf{a}_j} \left\{ \frac{1}{2} \|\mathbf{s}_j - \mathbf{D}\mathbf{a}_j\|_2^2 + \lambda \|\mathbf{a}_j\|_1 \right\}. \tag{B.3}$$

This algorithm is performed for a predefined number of steps T. At each step the dictionary is adapted based on η ($\eta \leq M$) randomly chosen vectors $\{\mathbf{s}_{i_1}, \mathbf{s}_{i_2}, \ldots, \mathbf{s}_{i_\eta}\}$. The adaptation rate of the dictionary descreses at each training step. The theoretical background of the algorithm is discused in Section 4.6.2.

C

IMPLEMENTATIONS

During my doctoral research, some of the Python code that I developed are available as free software.

Stationary Wavelet Transform (https://github.com/kesmarag/sp)

Hidden Markov Model (https://github.com/kesmarag/ml-hmm)

Gaussian Mixture Model (https://github.com/kesmarag/ml-gmm)

BIBLIOGRAPHY

[1] Michael Taroudakis, George Tzagkarakis, and Panagiotis Tsakalides. Characterization of an underwater acoustic signal using the statistics of the wavelet subband coefficients. *Theoretical and Computational Acoustics 2005*, 8:42–130, October 2006.

[2] Finn B. Jensen, William A. Kuperman, Michael B. Porter, and Henrik Schmidt. *Computational Ocean Acoustics*. Modern Acoustics and Signal Processing. Springer-Verlag, New York, 2 edition, 2011.

[3] Joseph B. Keller. Rays, waves and asymptotics. *Bulletin of the American Mathematical Society*, 84(5):727–750, September 1978.

[4] Terry Foreman. Ray Modeling Methods for Range Dependent Ocean Environments. Technical, Naval Ocean Research and Development Activity, December 1983.

[5] Michael B. Porter and Homer P. Bucker. Gaussian beam tracing for computing ocean acoustic fields. *The Journal of the Acoustical Society of America*, 82(4):1349–1359, October 1987.

[6] Evan Kruse Westwood. Acoustic propagation modeling in shallow water using ray theory. *The Journal of the Acoustical Society of America*, 85(6):2682–2682, June 1989.

[7] K. Fuchs and G. Müller. Computation of Synthetic Seismograms with the Reflectivity Method and Comparison with Observations. *Geophysical Journal of the Royal Astronomical Society*, 23(4):417–433, September 1971.

[8] B. L. N. Kennett. Seismic waves in a stratified half space. *Geophysical Journal of the Royal Astronomical Society*, 61(1):1–10, April 1980.

[9] C. L. Pekeris. THEORY OF PROPAGATION OF EXPLOSIVE SOUND IN SHALLOW WATER. In *Geological Society of America Memoirs*, volume 27, pages 1–116. Geological Society of America, 1948.

[10] Fred D. Tappert. The parabolic approximation method. In Joseph B. Keller and John S. Papadakis, editors, *Wave Propagation and Underwater Acoustics*, Lecture Notes in Physics, pages 224–287. Springer Berlin Heidelberg, Berlin, Heidelberg, 1977.

[11] James A. Davis, DeWayne White, and Raymond C. Cavanagh. NORDA Parabolic Equation Workshop, 31 March - 3 April 1981. Technical Report NORDA-TN-143, NORDA-TN-143, Naval Ocean Research and Development Activity Stennis Space Center MS, September 1982.

[12] Ding Lee, William L. Siegmann, and Oregory A. Kriegsmann. A wide-angle three-dimensional parabolic wave equation. *The Journal of the Acoustical Society of America*, 74(S1):S96–S96, November 1983.

[13] Ralph A. Stephen. A review of finite difference methods for seismo-acoustics problems at the seafloor. *Reviews of Geophysics*, 26(3):445–458, August 1988.

[14] J. Robert Fricke. Acoustic scattering from elemental Arctic ice features: Numerical modeling results. *The Journal of the Acoustical Society of America*, 93(4):1784–1796, April 1993.

[15] Michael J. Buckingham. Theory of acoustic propagation around a conical seamount. *The Journal of the Acoustical Society of America*, 80(1):265–277, July 1986.

[16] Michael I. Taroudakis. A coupled–mode formulation for the solution of the helmholtz equation in water in the presence of a conical sea-mount. *Journal of Computational Acoustics*, 04(01):101–121, March 1996.

[17] Wenyu Luo and Henrik Schmidt. Three-dimensional propagation and scattering around a conical seamount. *The Journal of the Acoustical Society of America*, 125(1):52, January 2009.

[18] John S. Perkins and Ralph N. Baer. An approximation to the three-dimensional parabolic-equation method for acoustic propagation. *The Journal of the Acoustical Society of America*, 72(2):515–522, August 1982.

[19] R. B. Evans. A coupled mode solution for acoustic propagation in a waveguide with stepwise depth variations of a penetrable bottom. *The Journal of the Acoustical Society of America*, 74(1):188–195, July 1983.

[20] M. I. Taroudakis, G. A. Athanassoulis, and J. P. Ioannidis. A variational principle for underwater acoustic propagation in a three-dimensional ocean environment. *The Journal of the Acoustical Society of America*, 88(3):1515–1522, September 1990.

[21] Erich Zauderer. *Partial Differential Equations of Applied Mathematics*. John Wiley & Sons, October 2011. Google-Books-ID: pO9cwd5UQP4C.

[22] Charles Allan Boyles. *Acoustic waveguides: applications to oceanic science*. John Wiley & Sons, Incorporated, February 1984. Google-Books-ID: wxRRAAAAMAAJ.

[23] Harry Nyquist. Certain topics in telegraph transmission theory. *Trans. AIEE*, 47:617–644, April 1928.

[24] D. M. F. Chapman and Dale D. Ellis. The group velocity of normal modes. *The Journal of the Acoustical Society of America*, 74(3):973–979, September 1983.

[25] Walter Munk and Carl Wunsch. Ocean acoustic tomography: Rays and modes. *Reviews of Geophysics*, 21(4):777–793, May 1983.

[26] E. C. Shang. Ocean acoustic tomography based on adiabatic mode theory. *The Journal of the Acoustical Society of America*, 85(4):1531–1537, April 1989.

[27] Munk. *Ocean acoustic tomography*. Cambridge University Press, Cambridge, England, 1995.

[28] William Menke. *Geophysical Data Analysis: Discrete Inverse Theory*. Elsevier, October 1989.

[29] Michael D. Collins, W. A. Kuperman, and Henrik Schmidt. Nonlinear inversion for ocean-bottom properties. *The Journal of the Acoustical Society of America*, 92(5):2770–2783, November 1992.

[30] C. E. Lindsay and N. R. Chapman. Matched field inversion for geoacoustic model parameters using adaptive simulated annealing. *IEEE Journal of Oceanic Engineering*, 18(3):224–231, July 1993.

[31] Peter Gerstoft. Inversion of acoustic data using a combination of genetic algorithms and the Gauss–Newton approach. *The Journal of the Acoustical Society of America*, 97(4):2181–2190, April 1995.

[32] Mark R. Fallat and Stan E. Dosso. Geoacoustic Inversion for the Workshop '97 Benchmark Test Cases Using Simulated Annealing. *Journal of Computational Acoustics*, 06(01n02):29–43, March 1998.

[33] Mark R. Fallat, Peter L. Nielsen, and Stan E. Dosso. Hybrid geoacoustic inversion of broadband Mediterranean Sea data. *The Journal of the Acoustical Society of America*, 107(4):1967–1977, March 2000.

[34] S.m. Jesus and A. Caiti. Estimating geoacoustic bottom properties from towed array data. *Journal of Computational Acoustics*, 04(03):273–290, September 1996.

[35] A. Caiti, S. M. Jesus, and A. Kristensen. Geoacoustic seafloor exploration with a towed array in a shallow water area of the Strait of Sicily. In *OCEANS 96 MTS/IEEE Conference Proceedings. The Coastal Ocean - Prospects for the 21st Century*, volume 1, pages 156–161 vol.1, September 1996.

[36] A. Tolstoy. *Matched field processing for underwater acoustics*. Singapore : World Scientific, 1993.

[37] T. Bohlen, S. Kugler, G. Klein, and F. Theilen. 1.5d inversion of lateral variation of Scholte-wave dispersion. *GEOPHYSICS*, 69(2):330–344, March 2004.

[38] Simone Kugler, Thomas Bohlen, Thomas Forbriger, Sascha Bussat, and Gerald Klein. Scholte-wave tomography for shallow-water marine sediments. *Geophysical Journal International*, 168(2):551–570, February 2007.

[39] A. Tolstoy, O. Diachok, and L. N. Frazer. Acoustic tomography via matched field processing. *The Journal of the Acoustical Society of America*, 89(3):1119–1127, March 1991.

[40] A. Tolstoy. Linearization of the matched field processing approach to acoustic tomography. *The Journal of the Acoustical Society of America*, 91(2):781–787, February 1992.

[41] A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky. An overview of matched field methods in ocean acoustics. *IEEE Journal of Oceanic Engineering*, 18(4):401–424, October 1993.

[42] S. E. Dosso, M. L. Yeremy, J. M. Ozard, and N. R. Chapman. Estimation of ocean-bottom properties by matched-field inversion of acoustic field data. *IEEE Journal of Oceanic Engineering*, 18(3):232–239, July 1993.

[43] Michael B. Porter and A. Tolstoy. The matched field processing benchmark problems. *Journal of Computational Acoustics*, 02(03):161–185, September 1994.

[44] A. Tolstoy. Simulated performance of acoustic tomography via matched field processing. *Journal of Computational Acoustics*, 02(01):1–10, March 1994.

[45] Stan E. Dosso and Peter L. Nielsen. Quantifying uncertainty in geoacoustic inversion. II. Application to broadband, shallow-water data. *The Journal of the Acoustical Society of America*, 111(1):143–159, January 2002.

[46] Stan E. Dosso and Michael J. Wilmut. Bayesian focalization: Quantifying source localization with environmental uncertainty. *The Journal of the Acoustical Society of America*, 121(5):2567–2574, May 2007.

[47] Stan E. Dosso and Michael J. Wilmut. Uncertainty estimation in simultaneous Bayesian tracking and environmental inversion. *The Journal of the Acoustical Society of America*, 124(1):82–97, July 2008.

[48] Stan E. Dosso and Michael J. Wilmut. Bayesian multiple-source localization in an uncertain ocean environment. *The Journal of the Acoustical Society of America*, 129(6):3577–3589, June 2011.

[49] Stan E. Dosso, Charles W. Holland, and Malcolm Sambridge. Parallel tempering for strongly nonlinear geoacoustic inversion. *The Journal of the Acoustical Society of America*, 132(5):3030–3040, November 2012.

[50] En-Cen Lo, Ji-Xun Zhou, and Er-Chang Shang. Normal mode filtering in shallow water. *The Journal of the Acoustical Society of America*, 74(6):1833–1836, December 1983.

[51] Michael I. Taroudakis and John S. Papadakis. A modal inversion scheme for ocean acoustic tomography. *Journal of Computational Acoustics*, 01(04):395–421, December 1993.

[52] E. C. Shang, A. G. Voronovich, Y. Y. Wang, K. Naugolnykh, and L. Ostrovsky. New schemes of ocean acoustic tomography. *Journal of Computational Acoustics*, 08(03):459–471, September 2000.

[53] Valerii V. Goncharov and Alexander G. Voronovich. An experiment on matched-field acoustic tomography with continuous wave signals in the Norway Sea. *The Journal of the Acoustical Society of America*, 93(4):1873–1881, April 1993.

[54] Gopu R. Potty, James H. Miller, James F. Lynch, and Kevin B. Smith. Tomographic inversion for sediment parameters in shallow water. *The Journal of the Acoustical Society of America*, 108(3):973–986, August 2000.

[55] Gopu R. Potty, James H. Miller, and James F. Lynch. Inversion for sediment geoacoustic properties at the New England Bight. *The Journal of the Acoustical Society of America*, 114(4):1874–1887, October 2003.

[56] Michael I. Taroudakis and George Tzagkarakis. On the use of the reassigned wavelet transform for mode identification. *Journal of Computational Acoustics*, 12(02):175–196, June 2004.

[57] Julien Bonnel and N. Ross Chapman. Geoacoustic inversion in a dispersive waveguide using warping operators. *The Journal of the Acoustical Society of America*, 130(2):EL101–107, August 2011.

[58] Julien Bonnel, Stan E. Dosso, and N. Ross Chapman. Bayesian geoacoustic inversion of single hydrophone light bulb data using warping dispersion analysis. *The Journal of the Acoustical Society of America*, 134(1):120–130, July 2013.

[59] Walter Munk and Carl Wunsch. Ocean acoustic tomography: a scheme for large scale monitoring. *Deep Sea Research Part A. Oceanographic Research Papers*, 26(2):123–161, February 1979.

[60] E. K. Skarsoulis, G. A. Athanassoulis, and Uwe Send. Ocean acoustic tomography based on peak arrivals. *The Journal of the Acoustical Society of America*, 100(2):797–813, 1996.

[61] THETIS Group. Open-ocean deep convection explored in the Mediterranean. *EOS Transactions*, 75:217–221, 1994.

[62] E. K. Skarsoulis and U. Send. One-Step Analysis of Nonlinear Traveltime Data in Ocean Acoustic Tomography. *Journal of Atmospheric and Oceanic Technology*, 17(2):240–254, February 2000.

[63] Michael Taroudakis. Identifying Modal Arrivals in Shallow Water for Bottom Geoacoustic Inversions. *Journal of Computational Acoustics - J COMPUT ACOUST*, 8:307–324, 2000.

[64] Michael Taroudakis, Maria Markaki, and Eirini Mavritsaki. Matching modal arrivals in shallow water for tomographic inversions. In *5th European Conference on Underwater Acoustics*, pages 1317–1322, 2000.

[65] Michael Taroudakis and Maria Markaki. Tomographic inversions in shallow water using modal travel time measurements. *Acta Acustica united with ACUSTICA*, 87:647–658, 2001.

[66] Solomon Kullback. *Information Theory and Statistics*. Dover Publications, Mineola, N.Y, new edition edition edition, July 1997.

[67] George Tzagkarakis, Michael Taroudakis, and Panagiotis Tsakalides. A statistical geoacoustic inversion scheme based on a modified radial basis functions neural network. *The Journal of the Acoustical Society of America*, 122:1959–68, November 2007.

[68] Ourania Siskoglou. Bayesian inversion technique based on the statistical characterization of acoustic signals with applications in Ocean Acoustic Tomography. Master's thesis, University of Crete, Heraklion, 2017.

[69] P. Papadakis, C. Smaragdakis, M. Taroudakis, and A. Tolstoy. Hybrid inversion techniques for geoacoustic inversion. pages 199–206, Istanbul, 2010.

[70] Michael Taroudakis and Costas Smaragdakis. A hybrid approach for estimating range-dependent properties of shallow water environments. *The Journal of the Acoustical Society of America*, 136(4):2120–2120, October 2014.

[71] Yves Meyer. Ondelettes et fonctions splines. *Séminaire Équations aux dérivées partielles (Polytechnique)*, pages 1–18, 1986.

[72] S. G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, July 1989.

[73] Stephane Mallat. *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*. Academic Press, Amsterdam ; Boston, 3 edition edition, December 2008.

[74] I. Daubechies. *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, January 1992.

[75] Ronald R. Coifman, Yves Meyer, and Victor Wickerhauser. Wavelet Analysis and Signal Processing. In *In Wavelets and their Applications*, pages 153–178, 1992.

[76] Mladen Victor Wickerhauser. *Adapted Wavelet Analysis from Theory to Software*. A. K. Peters, Ltd., Natick, MA, USA, 1994.

[77] John A. Gubner and Wei-Bin Chang. Wavelet transforms for discrete-time periodic signals. *Signal Processing*, 42(2):167–180, March 1995.

[78] Pavel Rajmic and Zdenek Prusa. Discrete wavelet transform of finite signals: detailed study of the algorithm. *International Journal of Wavelets, Multiresolution and Information Processing*, 12(01):1450001, October 2013.

[79] Patrick Billingsley. *Probability and Measure*. Wiley, Hoboken, N.J, anniversary edition edition, February 2012.

[80] Joseph K. Blitzstein and Jessica Hwang. *Introduction to Probability*. Chapman and Hall/CRC, Boca Raton, 1 edition edition, July 2014.

[81] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, New York, April 2011.

[82] Charles Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. The Regents of the University of California, 1972.

[83] John P. Nolan. Parameterizations and modes of stable distributions. *Statistics & Probability Letters*, 38(2):187–195, June 1998.

[84] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, Cambridge, MA, 1 edition edition, August 2012.

[85] A P Dempster, N M Laird, and D B Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. (1):39.

[86] Wilson A. Sutherland. *Introduction to Metric and Topological Spaces*. Oxford University Press, Oxford, 2 edition edition, October 2009.

[87] G Mclachlan and K Basford. *Mixture Models: Inference and Applications to Clustering*, volume 38. 1988.

[88] A. P. Dawid. Conditional Independence in Statistical Theory. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(1):1–31, 1979.

[89] Leonard E. Baum and Ted Petrie. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *The Annals of Mathematical Statistics*, 37(6):1554–1563, December 1966.

[90] Leonard E. Baum and J. A. Eagon. An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. *Bulletin of the American Mathematical Society*, 73(3):360–363, 1967.

[91] Leonard E. Baum and George Sell. Growth transformations for functions on manifolds. *Pacific Journal of Mathematics*, 27(2):211–227, November 1968.

[92] Leonard E. Baum, Ted Petrie, George Soules, and Norman Weiss. A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *The Annals of Mathematical Statistics*, 41(1):164–171, February 1970.

[93] J. MacQueen. Some methods for classification and analysis of multivariate observations. The Regents of the University of California, 1967.

[94] Tom S. Verma and Judea Pearl. Causal Networks: Semantics and Expressiveness. *arXiv:1304.2379 [cs]*, March 2013. arXiv: 1304.2379.

[95] Dan Geiger and Judea Pearl. On the Logic of Causal Models. *arXiv:1304.2355 [cs]*, March 2013. arXiv: 1304.2355.

[96] A. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13(2):260–269, April 1967.

[97] J. Omura. On the Viterbi decoding algorithm. *IEEE Transactions on Information Theory*, 15(1):177–179, January 1969.

[98] F. R. Kschischang, B. J. Frey, and H. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47(2):498–519, February 2001.

[99] W. Reiher. Hammersley, J. M., D. C. Handscomb: Monte Carlo Methods. Methuen & Co., London, and John Wiley & Sons, New York, 1964. VII + 178 S., Preis: 25 s. *Biometrische Zeitschrift*, 8(3):209–209, January 1966.

[100] Reuven Y. Rubinstein. *Simulation and the Monte Carlo Method*. John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1981.

[101] Brian D. Ripley. *Stochastic Simulation*. Wiley-Interscience, New York, 1 edition edition, March 2006.

[102] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online Dictionary Learning for Sparse Coding. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 689–696, New York, NY, USA, 2009. ACM.

[103] Carlos Ramirez, Vladik Kreinovich, and Miguel Argaez. Why l1 is a good approximation to l0: A geometric explanation. *Journal of Uncertain Systems*, 7(3):203–207, July 2013.

[104] David E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Professional, Reading, Mass, 1 edition edition, January 1989.

[105] Félix-Antoine Fortin, François-Michel De Rainville, Marc-André Gardner, Marc Parizeau, and Christian Gagné. DEAP: Evolutionary Algorithms Made Easy. *Journal of Machine Learning Research*, 13:2171–2175, July 2012.

[106] Michael Taroudakis, Costas Smaragdakis, and Ross Chapman. Inversion of acoustical data from the "Shallow Water 06" experiment by statistical signal characterization. *The Journal of the Acoustical Society of America*, 136:EL336, October 2014.

[107] Viktoria Taroudaki and Dianne P. O'Leary. Near-Optimal Spectral Filtering and Error Estimation for Solving Ill-Posed Problems. *SIAM Journal on Scientific Computing*, 37(6):A2947–A2968, January 2015.

[108] Viktoria Taroudaki, Costas Smaragdakis, and Michael Taroudakis. *Deblurring acoustic signals for statistical characterization in applications of ocean acoustic tomography*, volume 140. October 2016.

[109] Viktoria Taroudaki, Michael Taroudakis, and Costas Smaragdakis. Statistical optimal filtering method for acoustical signal deblurring. *The Journal of the Acoustical Society of America*, 144(3):1689–1689, September 2018.

[110] M. G. Jafari and M. D. Plumbley. Fast Dictionary Learning for Sparse Representations of Speech Signals. *IEEE Journal of Selected Topics in Signal Processing*, 5(5):1025–1031, September 2011.

[111] George Tzagkarakis and Panagiotis Tsakalides. A statistical approach to texture image retrieval via alpha-stable modeling of wavelet decompositions. In *5th International Workshop on Image Analysis for Multimedia Interactive Services*, pages 21–23, 1997.

[112] George Tzagkarakis, Baltasar Beferull-Lozano, and Panagiotis Tsakalides. Rotation-Invariant Texture Retrieval via Signature Alignment Based on Steerable Sub-Gaussian Modeling. *IEEE Transactions on Image Processing*, 17(7):1212–1225, July 2008.

[113] John P. Nolan. Modeling Financial Data with Stable Distributions. In Svetlozar T. Rachev, editor, *Handbook of Heavy Tailed Distributions in Finance*, volume 1 of *Handbooks in Finance*, pages 105–130. North-Holland, Amsterdam, January 2003.

[114] G. E. P. Box and D. R. Cox. An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2):211–252, 1964.

[115] J. M. Bland and D. G. Altman. Transformations, means, and confidence intervals. *BMJ (Clinical research ed.)*, 312(7038):1079, April 1996.

[116] D. McGuinness, S. Bennett, and E. Riley. Statistical analysis of highly skewed immune response data. *Journal of Immunological Methods*, 201(1):99–114, February 1997.

[117] Henry C. Thode. *Testing For Normality*. CRC Press, January 2002.

[118] R. Bakis. Continuous speech recognition via centisecond acoustic states. *Acoustical Society of America Journal*, 59:S97, January 1976.

[119] C. C. Tappert. A Markov model acoustic phonetic component for automatic speech recognition †. *International Journal of Man-Machine Studies*, 9(3):363–373, May 1977.

[120] Lawrence R Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *PROCEEDINGS OF THE IEEE*, 77(2):30, 1989.

[121] Joseph Picone. Continuous Speech Recognition Using Hidden Markov Models. *Ieee Assp Magazine*, 1990.

[122] X. Menendez-Pidal, J. B. Polikoff, S. M. Peters, J. E. Leonzio, and H. T. Bunnell. The Nemours database of dysarthric speech. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*, volume 3, pages 1962–1965 vol.3, October 1996.

[123] Y. Cem Subakan, Johannes Traa, Paris Smaragdis, and Daniel Hsu. Method of moments learning for left-to-right Hidden Markov models. In *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 1–5, New Paltz, NY, USA, October 2015. IEEE.

[124] Andrew R. Liddle. Information criteria for astrophysical model selection. *Monthly Notices of the Royal Astronomical Society: Letters*, 377(1):L74–L78, May 2007. arXiv: astro-ph/0701113.

[125] Gilles Celeux and Jean-Baptiste Durand. Selecting hidden Markov model state number with cross-validated likelihood. *Computational Statistics*, 23(4):541–564, October 2008.

[126] Peter Gerstoft. Inversion of seismoacoustic data using genetic algorithms and a posteriori probability distributions. *The Journal of the Acoustical Society of America*, 95(2):770–782, February 1994.

[127] Michael I. Taroudakis and Maria G. Markaki. On the use of matched-field processing and hybrid algorithms for vertical slice tomography. *The Journal of the Acoustical Society of America*, 102(2):885–895, August 1997.

[128] Michael I. Taroudakis and Costas Smaragdakis. Inversions of statistical parameters of an acoustic signal in range-dependent environments with applications in ocean acoustic tomography. *The Journal of the Acoustical Society of America*, 134(4):2814–2822, October 2013.

[129] Michael Taroudakis, Costas Smaragdakis, and Ross Chapman. Denoising underwater acoustic signals for applications in acoustical oceanography. *Journal of Computational Acoustics*, January 2017.

[130] Y. Lin, A. E. Newhall, T. F. Duda, P. F. J. Lermusiaux, and P. J. Haley. Merging Multiple-Partial-Depth Data Time Series Using Objective Empirical Orthogonal Function Fitting. *IEEE Journal of Oceanic Engineering*, 35(4):710–721, October 2010.

[131] John A. Goff and James A. Austin. Seismic and bathymetric evidence for four different episodes of iceberg scouring on the New Jersey outer shelf: Possible correlation to Heinrich events. *Marine Geology*, 266(1):244–254, October 2009.

[132] Yong-Min Jiang and N. Ross Chapman. Bayesian geoacoustic inversion in a dynamic shallow water environment. *The Journal of the Acoustical Society of America*, 123(6):EL155–EL161, May 2008.

[133] Bien Aik Tan, Peter Gerstoft, Caglar Yardim, and William S. Hodgkiss. Recursive Bayesian synthetic aperture geoacoustic inversion in the presence of motion dynamics. *The Journal of the Acoustical Society of America*, 136(3):1187–1198, September 2014.

[134] John A. Goff, Barbara J. Kraft, Larry A. Mayer, Steven G. Schock, Christopher K. Sommerfield, Hilary C. Olson, Sean P.S. Gulick, and Sylvia Nordfjord. Seabed characterization on the New Jersey middle and outer shelf: correlatability and spatial variability of seafloor sediment properties. *Marine Geology*, 209(1-4):147–172, August 2004.

[135] A. Turgut, D. Lavoie, D. J. Walter, and W. B. Sawyer. Measurements of Bottom Variability During Swat New Jersey Shelf Experiments. In Nicholas G. Pace and Finn B. Jensen, editors, *Impact of Littoral Environmental Variability of Acoustic Predictions and Sonar Performance*, pages 91–98. Springer Netherlands, Dordrecht, 2002.

[136] Larry A. Mayer, Barbara J. Kraft, Peter Simpkin, Paul Lavoie, Eric Jabs, and Eric Lynskey. In-Situ Determination of the Variability of Seafloor Acoustic Properties: An Example from the Onr Geoclutter Area. In Nicholas G. Pace and Finn B. Jensen, editors, *Impact of Littoral Environmental Variability of Acoustic Predictions and Sonar Performance*, pages 115–122. Springer Netherlands, Dordrecht, 2002.

[137] B. J. Kraft, L. A. Mayer, P. Simpkin, P. Lavoie, E. Jabs, E. Lynskey, and J. A. Goff. Calculation of In Situ Acoustic Wave Properties in Marine Sediments. In Nicholas G. Pace and Finn B. Jensen, editors, *Impact of Littoral Environmental Variability of Acoustic Predictions and Sonar Performance*, pages 123–130. Springer Netherlands, Dordrecht, 2002.

[138] J. Bonnel, C. Gervaise, B. Nicolas, and J. I. Mars. Single-receiver geoacoustic inversion using modal reversal. *The Journal of the Acoustical Society of America*, 131(1):119–128, January 2012.

[139] R. R. Coifman and M. V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38(2):713–718, March 1992.

[140] Neal Parikh and Stephen Boyd. Proximal Algorithms. *Found. Trends Optim.*, 1(3):127–239, January 2014.

Acoustic Signal Characterization using Hidden Markov Models with applications in Acoustical Oceanography