EuroRegio2016
June 13-15, 2016
Porto ▪ Portugal

Hidden Markov Models feature extraction for Inverting underwater acoustic signals using Wavelet Packet coefficients

Costas Smaragdakis, Michael Taroudakis

Department of Mathematics and Applied Mathematics, University of Crete,
Institute of Applied and Computational Mathematics, FORTH,
Voutes University Campus, 70013 Heraklion, Crete, Greece.
(kesmarag@tem.uoc.gr, taroud@uoc.gr)

Abstract
In current work we present a novel framework for characterizing and inverting underwater acoustic signals using pattern recognition clustering techniques which have been widely used in speech recognition, financial forecasting and generally in time series analysis. The new model is as a probabilistic model-based approach to characterize waveforms using time-frequency features using a discrete Wavelet Packet Transform (WPT). Unlike to our previous inversion works using statistical features of the sub-band coefficient of the wavelet transform of the signals, we take into account the sequential patterns of the signals in order to obtain a more precise characterization and hence to get more reliable inversion results. A set of Hidden Markov Models (HMMs) adapted to a group of training signals are applied to achieve the final feature extraction vector for each signal with posterior distributions. The training of the HMMs is performed by exploiting the k-mean like clustering approach with hmm posterior distribution as similarity measure. Each of the described feature vector has as its elements the posterior conditional probabilities given each one of the Hidden Markov Models. These posterior feature vectors constitute the inputs of a Mixture Density Network (MDN) whose outputs represent posterior probabilities densities of the inverted parameters. Experimental geo-acoustic inversion results based on a simple Pekeris environment are presented and are compared to the results obtained by a Statistical Characterization Scheme also developed by authors as a first measure of the robustness of the proposed new scheme.

Keywords: Inverse Problems, Wavelet Packet Transform, Hidden Markov Model Clustering, Geoacoustic inversions

PACS no. 43.60.Hj, 43.60.Np

1 Introduction

The goal of the work presented here is to study an approach for inverting acoustic signals recorded in the marine enviroment for the recovery of the environmental parameters, using sequential features of the signal for its characterization and a neural network which provides conditional posterior distribution functions of the recoverable parameters. Generally, inverse problems in underwater acoustics are associated with measurements of the acoustic signals in the time domain. A pair of observables $\mathbf{d}$ and the corresponding unknown environment parameters $\mathbf{m}$ forms the input and the output of a inverse problem respectively. The general form of a non-linear inverse problem is given by

$$T(\mathbf{m},\mathbf{d}) = 0, \quad (1)$$

where T is an apropriate non-linear function. In the current work, we propose a geoacoustic inversion scheme that combines several mathematical and machine learning ideas implemented in three successive steps: In the first step, a recorded discrete signal from a single receiver is associated with the effective part of the Wavelet Packet Transform spectogram after decomposing it in L levels keeping both the detail and the approximation coefficients in each level. In the second step, a clustering scheme of Hidden Markov Models is built using an approach described in Section 3. The extracted features from the first step are used in order to obtain probabilistic assigments to each one of the cluster nodes. Thus, a vector of the posterior probabilities is considered including information about the sequential patterns of the recording. In the last step of the proposed scheme, the second order feature vector of the signal consists the input of a probabilistic feed-forward neural network. For a given input vector this network provides a conditional posterior distribution function for the model parameters $p(\mathbf{m}\,|\,\mathbf{d})$. Finaly, a first validation test of the new scheme is presented in section 6 using a simple geoacoustic inversion case in which, the recoverable environmental parameters are those of the semi-infini sea-bed.

2 Wavelet Packet Transform for Feature Extraction

We consider a set of model parameter vectors $\{\mathbf{m}^{(1)},\ldots,\mathbf{m}^{(\mathrm{I})}\}$ and a set of discrete acoustic signals $\{\mathbf{s}^{(1)},\ldots,\mathbf{s}^{(\mathrm{I})}\}$ corresponding to the model vectors of same index for a specific unerwater environment . The signals are typically synthetic obtained by some approprtaie propagation model. We denote by D the set of model parameter – signal pairs, so that

$$\mathrm{D} = \{(\mathbf{m}^{(1)},\mathbf{s}^{(i)}),\ldots,(\mathbf{m}^{(\mathrm{I})},\mathbf{s}^{(\mathrm{I})})\} \quad (2)$$

We apply a Wavelet Packet Transform (WPT) [1] which is an extension of the typical Wavelet Transform (WT) the foundamental difference from which is that in WPT both the detail and approximation coefficients in each level are decomposed in order to form a full binary tree of coefficients as shown in Fig 1. After the decomposition of the discrete signal with N samples by L levels the spectogram of the signal is defined. This spectogram is a matrix with 2^L rows (frequency scales) and $N \times 2^{-L}$ columns (time windows) as depicted in Fig 2. For feature extraction we choose the submatrix of the spectogram that includes values above a certain threshold. Thus, a given acoustic signal $\mathbf{s}^{(i)}$ decomposed in L levels using the WPT can be associated with a time sequence of vectors formed by the N' columns $\mathbf{x}_1^{(i)},\ldots,\mathbf{x}_{N'}^{(i)}$ of the submatrix of the signal's wavelet packet spectogram. Therefore each one of the signals in the dataset D is characterized by the set $\mathbf{X}^{(\mathrm{i})}$:

$$\mathbf{s}^{(i)} \to \mathbf{X}^{(i)} = \{\mathbf{x}_1^{(i)},\ldots,\mathbf{x}_{N'}^{(i)}\}, \quad i = 1,\ldots,\mathrm{I} \quad (3)$$

3 Hidden Markov Models Clustering

Our goal is to use a clustering scheme that can perform grouping of the signals based on the feature vectors obtained by the wavelet feature extraction procedure described above. In this scheme each signal can be associated with a set of posterior probability densities which are associated with the likelihood of these features to be described by each cluster. For this purpose we are using Hidden Markov Models as cluster nodes. With that choice, our clustering scheme is able to take into account the sequential characteristics of the feature vectors of the signal. Thus, the clustering scheme takes as inputs the wavelet packet features of a signal and provides a unique feature vector for it.

Following Elliot et al. [2] we can construct a Markov chain of latent variables so that each one of the observed feature vectors $\mathbf{x}_n$ depends on the value of a corresponding latent variable $\mathbf{z}_n$. A key property of this model is the conditional independency between $\mathbf{z}_{n-1}$ and $\mathbf{z}_{n+1}$ if the value of $\mathbf{z}_n$ is known, which was proven in [3].

Each latent variable is dictated by a discrete multinomial variable represented with a K-dimensional vector $\mathbf{z}_n = (z_{n1}, \ldots, z_{nK})^T$ in which, only one of the elements equals 1, and all other elements equal 0. Despite the fact that latent variables are unobserved, they govern the way that the observed data are generated. Note that K is the number of states as determined in Markov chain theory.

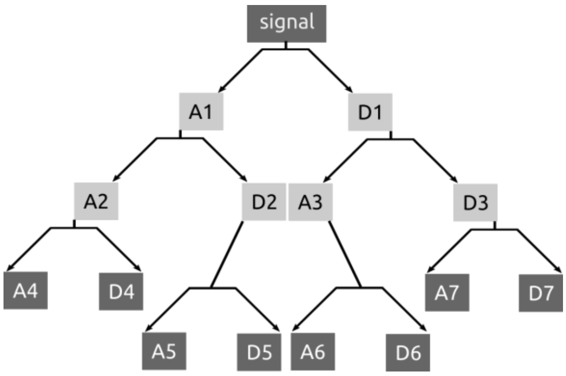

Figure 1: Discrete Wavelet Packet decomposition by L=3 levels

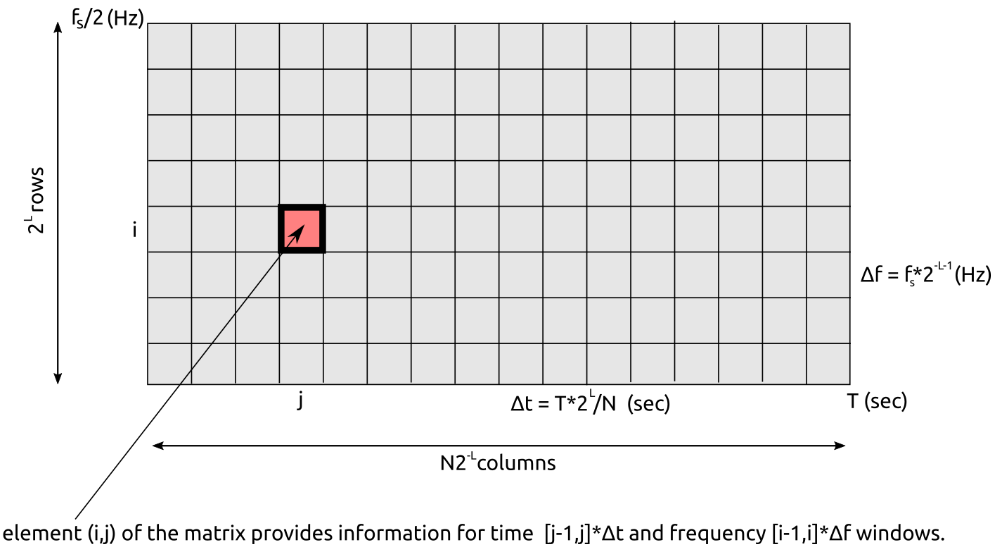

Figure 2: Time-Frequency resolution matrix of an acoustic signal.

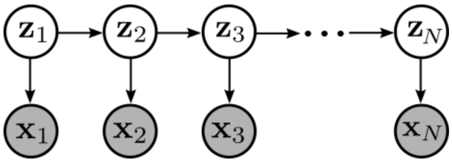

Figure 3: Markov chain of latent variables.

In order to determine the probabilistic model to be used to signal characterization, we have to define the marginal probabilities for the first latent variable $\mathbf{z}_1$ which is the only one totally independent. So, we choose a vector $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$ where

$$\pi_k = p(z_{1k} = 1), \quad k = 1, \ldots, K \tag{4}$$

This vector must obey the relation $\sum_{k=1}^{K} \pi_k = 1$.

Also, we denote by $\mathbf{A}$ the matrix with elements the transition probabilities from one state to another. For example, the element A_{ij} expresses the probability to move from the i-th to the j-th state. Therefore:

$$p(z_{nj} = 1 \mid z_{n-1,i} = 1) = A_{ij}, \quad i, j = 1, \ldots, K \tag{5}$$

Note that each row of the transition matrix has sum equals to 1. The last step is to define the conditional distributions of the feature vectors given the corresponding latent variables. In our case we have chosen to model these distributions via a Mixture of Gaussian Distributions with diagonal covariance matrices, governed by a set of parameters $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_k\}$, so that

$$p(\mathbf{x}_n \mid z_{nk} = 1, \theta_k) = \sum_{q=1}^{Q} w_q^k \mathrm{N}\,(\mathbf{x}_n; \boldsymbol{\mu}_q^k, \boldsymbol{\Sigma}_q^k), \quad k = 1, \ldots, K \tag{6}$$

where $w_q^k \in \{0,1\}$ with $\sum_{k=1}^{K} w_q^k = 1, \quad q = 1, \ldots, Q$.

As it is easily understood, the whole model is governed by the three families of distributions described above. So, we can denote as a Hidden Markov Model (HMM) the set of parameters $\boldsymbol{\lambda}$.

$$\boldsymbol{\lambda} = \{\boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\theta}\} \tag{7}$$

The joint distribution for the HMM is given by

$$p(\mathbf{X}, \mathbf{Z} \mid \boldsymbol{\lambda}) = p(\mathbf{z}_1 \mid \boldsymbol{\pi}) \Big[\prod_{n=2}^{N} p(\mathbf{z}_n \mid \mathbf{z}_{n-1}, \mathbf{A})\Big] \prod_{n=1}^{N} p(\mathbf{x}_n \mid \mathbf{z}_n, \boldsymbol{\theta}) \tag{8}$$

Let us denote by X, Z the supersets of the features and latent variables respectively, for each one of the signals in the data set $\{\mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(\mathrm{I})}\}$, thus:

$$\begin{aligned} \mathrm{X} &= \{\mathbf{X}^{(1)},\ldots,\mathbf{X}^{(\mathrm{I})}\} = \left\{\{\mathbf{x}_1^{(1)},\ldots,\mathbf{x}_N^{(1)}\},\ldots,\{\mathbf{x}_1^{(\mathrm{I})},\ldots,\mathbf{x}_N^{(\mathrm{I})}\}\right\} \\ \mathrm{Z} &= \{\mathbf{Z}^{(1)},\ldots,\mathbf{Z}^{(\mathrm{I})}\} = \left\{\{\mathbf{z}_1^{(1)},\ldots,\mathbf{z}_N^{(1)}\},\ldots,\{\mathbf{z}_1^{(\mathrm{I})},\ldots,\mathbf{z}_N^{(\mathrm{I})}\}\right\} \end{aligned} \tag{9}$$

The joint probability distribution of X,Z given a HMM descibed by $\boldsymbol{\lambda}$ is obtained as following

$$p(\mathrm{X},\mathrm{Z} \mid \boldsymbol{\lambda}) = \prod_{i=1}^{\mathrm{I}} p(\mathbf{X}^{(i)},\mathbf{Z}^{(i)} \mid \boldsymbol{\lambda})\,. \tag{10}$$

Marginizing over the latent superset we get the log-based likehood distribution for the features

$$\ln p(\mathrm{X} \mid \boldsymbol{\lambda}) = \ln \sum\nolimits_{\mathrm{Z}} p(\mathrm{X},\mathrm{Z} \mid \boldsymbol{\lambda})\,. \tag{11}$$

For minimizing the (1.10) we use the Expectation Maximization Algorithm (EM) [4] which is an iterative optimization procedure for statistical models including hidden variables. In each iteration EM starts with $\boldsymbol{\lambda}^{\mathrm{old}}$, performs two distinct steps and provides an updated $\boldsymbol{\lambda}^{\mathrm{new}}$ as described below:

1. **E step:** Evaluate the probabiliuty $p(\mathrm{Z} \mid \mathrm{X},\boldsymbol{\lambda}^{\mathrm{old}})$
2. **M step:** Evaluate the new estmation of the HMM parameters

$$\boldsymbol{\lambda}^{\mathrm{new}} = \arg\max\nolimits_{\lambda} \mathrm{Q}(\boldsymbol{\lambda},\boldsymbol{\lambda}^{\mathrm{old}}) \tag{12}$$

   where,

$$\mathrm{Q}(\boldsymbol{\lambda},\boldsymbol{\lambda}^{\mathrm{old}}) = \sum\nolimits_{\mathrm{Z}} p(\mathrm{Z} \mid \mathrm{X},\boldsymbol{\lambda}^{\mathrm{old}}) \ln p(\mathrm{X},\mathrm{Z} \mid \boldsymbol{\lambda}) \tag{13}$$

3. **Termination criterion:** Check for convergence of the log-likelihood function

$$|\ln p(\mathrm{X} \mid \boldsymbol{\lambda}^{\mathrm{new}}) - \ln p(\mathrm{X} \mid \boldsymbol{\lambda}^{\mathrm{old}})| < \mathrm{tol} \tag{14}$$

   If the criterion is not satisfied, set $\boldsymbol{\lambda}^{\mathrm{old}} \leftarrow \boldsymbol{\lambda}^{\mathrm{new}}$ and return to **step 1**.

Having the tools for maximizing the likelihood function ready, we shall pass to our main goal of clustering using a set of HMMs.

We radomly construct a set of J HMMs as $\mathrm{L} = \{\boldsymbol{\lambda}^{1},\ldots,\boldsymbol{\lambda}^{\mathrm{J}}\}$. We use a simple k-means [4] like approach to achive the clustering of the acoustic signals. The difference is that the k-means algorithm uses the Euclidean norm for doing the cluster assignations, while our algorithm uses the posterior probabilities of the set of HMMs instead. The procedure is described by the following steps:

1. Draw J signals from the data set at random and assign each with a member of L.
2. Evaluate the log-probabilities for all pairs signal-HMM $d_j^{(i)} = \ln p(\mathbf{X}^{(i)},\boldsymbol{\lambda}^{j})$.
3. Assign each $\mathbf{X}^{(i)}$ to the j^*-th HMM, where $j^* = \arg\min_j (d_j^{(i)})$.
4. Adapt all the HMMs in respect to all their assigned feature vectors.
5. If the assignations still changing return to step 2.

Having created and adapted the set L of HMMs we are able to use the vector of log-scale probabilities as signal features

$$\mathbf{m}^{(i)} \to \mathbf{s}^{(i)} \to \mathbf{X}^{(i)} \to \mathbf{d}^{(i)} = \{d_1^{(i)},\dots,d_J^{(i)}\}, \quad i=1,\dots,I \tag{15}$$

These features have taken into account the sequential structure of the acoustic signals. After performing a proper standarding (zero means and unit variances) these features are using for visualizing the estimate values as the final step of our inversion scheme.

4 Posterior Distributions via a Mixture Density Network

Using the characterizations (15), we are transforming the data set D to define the training set G

$$G = \{(\mathbf{m}^{(1)},\mathbf{d}^{(1)}),\dots,(\mathbf{m}^{(I)},\mathbf{d}^{(I)})\} \tag{16}$$

We denote by M the dimension of the model parameters $\mathbf{m}^{(i)}$ and J the dimension of the feature vectors $\mathbf{d}^{(i)}$. It should be clear that both vectors in each pair are normalized in order to have zero mean and one standard deviation.

Extending the approach proposed by Williams [5] to use mixture multivariate Gaussian densities with full covariance matrices, we build a Mixture Density Network (MDN) that takes the elements of a feature vector $\mathbf{d}$ as input and yields the marginal posterior distribution of the unknown model parameter vector $\mathbf{m}$.

$$p(\mathbf{m}\,|\,\mathbf{d}) = \sum_{k=1}^{K} \pi_k(\mathbf{d}) \mathrm{N}\,(\mathbf{m}\,|\,\boldsymbol{\mu}_k(\mathbf{d}),\boldsymbol{\Sigma}_k(\mathbf{d})) \tag{17}$$

where N denotes the normal distribution function , $\boldsymbol{\mu}_k(\mathbf{d})$ is the mean value vector and $\boldsymbol{\Sigma}_k(\mathbf{d})$ is the covariance matrix of the k^{th} component of the mixture. Williams has used the Cholesky factorization to the inverse of the covariance matrices which is nessesery in the calculation of the normal distribution

$$\boldsymbol{\Sigma}_k^{-1}(\mathbf{d}) = \mathbf{U}_k^T(\mathbf{d})\mathbf{U}_k(\mathbf{d}) \tag{18}$$

where $\mathbf{U}_k$ is an upper triangular matrix with strictly positive diagonal elements.

So, each normal distribution in the mixture depends on the $\boldsymbol{\mu}_k$ vector and $\mathbf{U}_k$ matrix and therefore after performing simple calculations it can be expressed as following :

$$\mathrm{N}\,(\mathbf{m}\,|\,\boldsymbol{\mu}_k(\mathbf{d}),\mathbf{U}_k(\mathbf{d})) = (2\pi)^{-M/2}\det(\mathbf{U}_k(\mathbf{d}))\exp\left(-0.5\,\|\,\mathbf{U}_k(\mathbf{d})(\mathbf{m}-\boldsymbol{\mu}_k(\mathbf{d}))\,\|_2^2\right) \tag{19}$$

In our implementation, we use a feed-forward neural network of a single hidden layer with H nodes, as illustrated in Figure 4, and ℓ outputs which generates, with proper transformation functions, a mixture Gaussian distribution with K components. The outputs of the network express the elements of $\pi_k, \boldsymbol{\mu}_k, \mathbf{U}_k$ using appropriate activation functions.

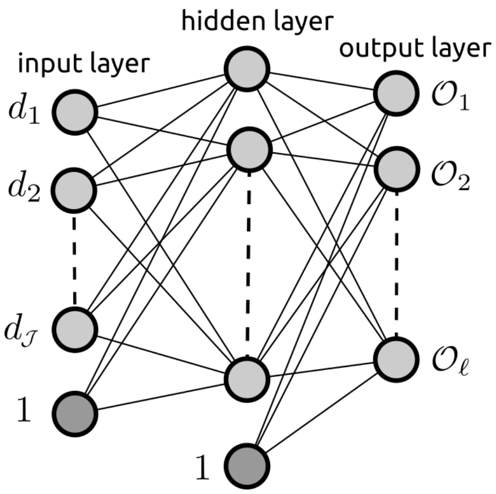

Figure 4: Mixture Density Network with a single hidden layer.

The network have K outputs $\{\mathrm{O}_{\pi}^{k}\}_{k=1,\ldots,K}$ that determine the mixing weights $\pi_k(\mathbf{x})$, $K \times M$ outputs $\{\mathrm{O}_{\boldsymbol{\mu}}^{k,j}\}_{k=1,\ldots,K}^{j=1,\ldots,M}$ for the mean values $\boldsymbol{\mu}_k(\mathbf{d})$, $K \times M$ outputs $\{\mathrm{O}_{\mathbf{U}}^{k,j}\}_{k=1,\ldots,K}^{j=1,\ldots,M}$ corresponding to diagonal elements of the matrices $\mathbf{U}_k$ and $K \times (M^2 - M)/2$ outputs $\{\mathrm{O}_{\mathbf{U}}^{k,i,j}\}_{k=1,\ldots,K}^{i=1,\ldots,M,\, j=i+1,\ldots,M}$ that determine the upper-diagonal elements of the matrices $\mathbf{U}_k$. For mixing weights we use the *softmax* function over the O_{π}^{k}, thus:

$$\pi_k(\mathbf{d}) = \frac{\exp(\mathrm{O}_{\pi}^{k})}{\sum_{j=1}^{K} \exp(\mathrm{O}_{\pi}^{j})} \tag{20}$$

Since the means $\boldsymbol{\mu}_k(\mathbf{d})$ get real values, the network outputs consist of:

$$\{\boldsymbol{\mu}_k\}_j = \mathrm{O}_{\boldsymbol{\mu}}^{k,j} \tag{21}$$

As diagonal elements of the matrices $\mathbf{U}_k$ we take the exponentials of the corresponding outputs while for upper diagonal elements we take values directly from the network outputs. Therefore each $\mathbf{U}_k$ upper diagonal matrix is obtained by the index-form formula:

$$\{\mathbf{U}_k\}_{i,j} = \begin{cases} exp(\mathrm{O}_{\mathbf{U}}^{k,i}), & i = j \\ \mathrm{O}_{\mathbf{U}}^{k,i,j}, & i < j \\ 0, & i > j \end{cases} \tag{22}$$

For simplicity we shall refer to the neural network weight parameters as $\mathbf{W}$ and to biases as $\mathbf{b}$. The error function for the mixture density network defined by the sum of the negative logarithm of the likelihood functions for each member of the data set [4]:

$$E(\mathbf{W},\mathbf{b}) = \sum_{i=1}^{I} E_i(\mathbf{W},\mathbf{b}), \quad (23)$$

where,

$$E_i(\mathbf{W},\mathbf{b}) = -\ln \sum_{k=1}^{K} \pi_k(\mathbf{d}_i,\mathbf{W},\mathbf{b}) \mathrm{N}\,(\mathbf{m}_i \mid \boldsymbol{\mu}_k(\mathbf{d}_i,\mathbf{W},\mathbf{b}), \mathbf{U}_k(\mathbf{d}_i,\mathbf{W},\mathbf{b})) \quad (24)$$

is the error function of the i-th pair of the training data set G defined above. The goal of the training procedure of the network is to find $\mathbf{W},\mathbf{b}$ so that $E(\mathbf{W},\mathbf{b})$ takes its minimum value. In order to minimize the above error function, we use the stochastic gradient descent (SGD) optimization scheme. The derivatives of the error function can be evaluated by employing the typical backpropagation procedure[6].

Recall that by the procedure described above, we have trained the neural network to provide as output the marginal distribution of the recovable parameters of an environment insonified by the acoustic waves of a known source given mesurements of the acoustic field at a specific location. Using this output we are able to estimate the probable values for these parameters and their confidence intervals.

5 Test Case: Pekeris Environment

As a first inversion example to validate the efficiency and the robustness of the new inversion scheme, we consider a typical shallow water Pekeris environment. The inverse problem corresponds to the recovery of the velocity c_b and of the density ρ_b of the sea-bed which is modeled as half space when the measurements of the acoustic field due to a tomographic source (modeled with a Gaussian spectrum) are available at a single hydrophone. We assume that the measurements are made in the presence of additive white Gaussian noise, resulting to low Signal-to-Noise Ratio (SNR) 10dB. The constant environmental parameters as well as the actual values of the recoverable parameters appear in Table 1, along with the operational parameters of the sound source.

We have considered all the parameters that control the inversion procedure via a trial and error technique. In this example, we have performed WPT using 4 level Daubechies 4 (db4) wavelet filters. For the second phase of our scheme, we have a cluster consisted by 5 HMMs, each of them having latent variables with 7 possible states. The emission distributions have been considered to be mixture of Gaussians with 2 components. Finally, in the last component of the model that illustrates the posterior information of the parameters, we have used a single layer MDN with 100 nodes and $\tanh$ based activation function.

The training set G is defined by generating a set of synthetic signals according to the unknown parameters c_b, ρ_b. The substrate sound speed takes values in $[1600 m/s, 1800 m/s]$ using a step of $5\, m/s$ and the density of the substrate takes values in $[1200 kg/m^3, 1400 kg/m^3]$ using a step of $5\, kg/m^3$, thus creating a set of a total 1681 parameter-feature pairs $(\mathbf{m},\mathbf{d})$.

Figure 5 presents the conditional posterior joint probability distribution for the two unknown parameters in logarithmic scale. Table 2 illustrates the inversion results of the proposed procedure. As inversion result with the proposed procedure we have considered to be the parameters for which the posterior joint probability distribution gets its maximum value.

Table 1: Environmental, source and receiver parameters for the Pekeris environment.

Parameters	**Actual Values**
Water Depth $h(m)$	200
Source Depth $z_s(m)$	100
Receiver Depth $z_R(m)$	100
Range $r(m)$	8000
Central Frequency $f_o(Hz)$	130
Bandwidth $\Delta f(Hz)$	90
Water Density $\rho_w(kg/m^3)$	1000
Sound Speed in Water $c_w(m/s)$	1500

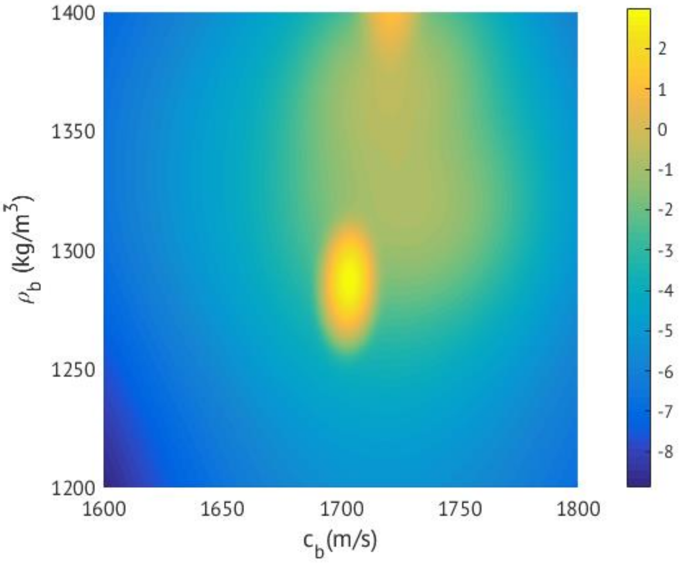

Figure 1: Conditional posterior joint probability distribution for the unknown parameters in logarithmic scale.

For comparison reasons, we have also provided the corresponding results obtained applying the Statistical Characterization Scheme [7] after the hybrid denoising strategy that has been used by the authors in [8].

Table 2: Inversion results using the proposed procedure as well as the Statistical Characterization Scheme using the hybrid denoising strategy.

Parameters	**Search space**	**Actual values**	**SCS(SD+CD)**	**Proposed procedure**
$\rho_b(kg/m^3)$	[1200, 1400]	1300.0	1342.1	1287.1
$c_b(m/s)$	[1600, 1800]	1700.0	1727.3	1703.5

6 Conclusions

A new procedure for the estimation of the unknown parameters of a shallow water environments using measurements of the acoustic field is presented, based on a Hidden Markov Clustering scheme employed for the characterization of the acoustic signal in association with a neural network. Using a simple test case corresponding a Pekeris environment and synthetic data for the recovery of sea-bed parameters we have shown that the proposed procedure gives very good results even when the acoustic data are considered with high level Gaussian noise. Future work includes the study of a proper regularization techniques to avoid under-fitting and over-fitting problems occurring while training each probabilistic model. Finally, the validation of the proposed scheme requires the applicability study in multidimensional inversion cases and data from real experiments.

Acknowledgements

Work supported partially by PEFYKA project within the KRIPIS action of the GSRT, funded by Greece and the European Regional Development Fund of the European Union under the NSRF and the O.P. Competitiveness and Entrepreneurship. The work is also carried out in the framework of the Hellenic Republic-Siemens Agreement partially funded by the Programmatic Agreement Between Research Centers- GSRT 2015-2017

References

[1] Mallat, S. *A Wavelet Tour of Signal Processing*, Academic Press, San Diego (USA), 2nd edition, 1999.

[2] Elliot, R; Aggoun, L; Moore J. *Hidden Markov Models: Estimation and Control*, Springer, New York (USA), 1995.

[3] Pearl, *J. Probabilistic Reasoning in Intelligent Systems,* Morgan Kaufmann, Los Angeles (USA), 2nd edition, 2014.

[4] Bishop, C. Pattern recognition and machine learning. Springer International Edition, New Delhi (India), 2006.

[5] Williams, P. Using neural networks to model conditional multivariate densities. *Neural Computations*, Vol 8 (4), 1996, pp 843-854.

[6] Rumelhart, D.E.; McClelland J.L; PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Volume 1: Foundations, MIT Press, 1986.

[7] Taroudakis, M.; Tzagkarakis, G.; Tsakalides, P. Classification of shallow water acoustic signals via alpha stable modeling of the one dimensional wavelet coefficients, J. Acoust. Soc. Am. Vol 119, 2006, pp 1396-1405.

[8] Taroudakis, M.; Smaragdakis, C. Denoising procedures for inverting underwater acoustics signals in applications of acoustical oceanography, *EuroNoise 2015, 31 May – 3 June, Maastricht*, pp 1393-1398