Probabilistic PCA and Ocean Acoustic Tomography Inversion with an Adjoint Method

Mohamed Berrada[a], Fouad Badran[b] and Sylvie Thiria[a]

[a]Laboratoire d'Océanographie et du Climat - Expérimentation et Approches Numériques, Université Pierre et Marie Curie, Tour 45-55 - 5ème étage - 4, place Jussieu, 75005 Paris, France

[b]Laboratoire CEDRIC, Conservatoire National des Arts et Métiers, 292, rue Saint Martin, 75003 Paris, France

mohamed.berrada@locean-ipsl.upmc.fr

We present an Ocean Acoustic Tomography (OAT) inversion in a shallow water environment. The idea is to determine the sound spreed profile $c(z)$, z is depth, knowing the acoustic pressures caused by a multiple frequencies source and collected by a sparse receiver array. The variational approach minimizes a cost function which measures the adequacy between the measurements and their forward model equivalent. This method introduces also a regularization term in the form $(c(z) - c_b(z))^T B^{-1}(c(z) - c_b(z))$, which supposes that $c(z)$ follows an *a priori* normal law. To circumvent the problem of estimating B^{-1}, we propose to model the celerity vectors by a probabilistic PCA. In contrast to the methods which use PCA as a regularization method and filter the useful information, we take a sufficient number of axes which allow the modelization of useful information and filter only the noise. The probabilistic PCA introduces a reduced number of non correlated latent variables η which act as new control parameters introduced in the cost function. This new regularization term, expressed as $\eta^T\eta$, reduces the optimization computation time. In the following we apply the probabilistic PCA to an OAT problem, and present the results obtained when performing twin experiments.

1 Introduction

The use of the variational method in both geoacoustic and OAT inversions is recent [1, 2]. In the following we use for OAT inversion a cost function which introduces a background term in the form $(c(z) - c_b(z))^T B^{-1}$ $(c(z) - c_b(z))$, where c(z) is the sound speed profile. This term corresponds to an *a priori* information on the physical parameters, which suppose that they are normally distributed. Usually the estimation of B^{-1} requires a data subset statistically representative of the problem. Due to the high dimensionality of the vectors c(z) and the strong correlation between their components it becomes difficult to estimate the matrix B^{-1}.
We propose to model $c(z)$ using the probabilistic Principal Component Analysis (PCA) [3]. This model assumes that $c(z)$ is made of two terms: the first is normally distributed in a linear subspace of reduced dimensionality, the remaining term being an isotropic normally distributed noise. The linear subspace is generated by the first principal components of the covariance empirical matrix of the data. The component of the data projections on the linear subspace are non correlated, zero mean and normally distributed. After normalization the components constitute the latent variables of the model and can be taken as the new control variables of an adjoint-based optimization method. Thus the variational minimization of the cost function acts on a reduced number of non correlated latent variables.
The paper is organized as follows. Section 2 reviews the forward model based on the width angle PE (WAPE) and the non local boundary conditions (NLBC), it presents the usual cost function with its background term. Section 3 introduces the probabilistic PCA approach and gives a complete methodology to minimizes the cost function with respect to the latent variables. Section 4 introduces an actual experiment in geoacoustic called the Yellow Shark experiment (YS). Section 5 presents the performances obtained when applying the variational PCA inversion to the Yellow Shark experiment.

2 Adjoint-Based Optimization

Let G_f be a forward model which represents the prediction of acoustic propagation in an oceanic environment:

$$G_f(c(z)) = \psi(r, z), \tag{1}$$

where f indicate frequency of acoustic signal source, $c(z)$ is the sound speed profile in the water column (control parameters). The field $\psi(r, z)$ therein is related to the complex pressure $p(r, z)$ according to

$$p(r, z) = \frac{\psi(r, z)\exp(ik_0 r)}{\sqrt{k_0 r}}, \tag{2}$$

where $k_0 = 2\pi f / c_0$ is a reference wave number.
Our forward model G_f is based on the Wide-Angle PE (WAPE) due to Claerbout [4], an analytical Thomson's source term [5], a Direchlet boundary condition at the surface and a boundary condition at the water-bottom interface ($z = z_b$) called Non Local Boundary Condition (NLBC) due to Yevick and Thomson [6]. For a stratified medium with varying density $\rho(z)$ and absorption loss $\alpha(z)$ the system can be described as follows:

$$\begin{cases} 2ik_0\left(1 + \frac{1}{4}(\mathcal{N}^2 - 1)\right)\frac{\partial\psi}{\partial r} + \frac{\partial^2\psi}{\partial z^2} + k_0^2(\mathcal{N}^2 - 1)\psi \\ \qquad + \frac{i}{2k_0}\rho\frac{\partial}{\partial z}\rho^{-1}\frac{\partial^2\psi}{\partial z\partial r} = 0, \\ \psi(0, z) = S_f(z), \\ \psi(r, 0) = 0, \\ \text{NLBC} \quad \left[\frac{\partial}{\partial z} - i\beta\right]\psi(r + \Delta r, z_b) = \\ \qquad i\beta\sum_{j=1}^{K+1}\mathbf{g}_{1,j}\psi\left[(K + 1 - j)\Delta r, z_b\right], \end{cases} \tag{3}$$

where $\mathcal{N}(z) = n(z)[1 + i\alpha(z)]$ and $n(z) = c_0/c(z)$ the refraction index. The interval $0 \to r + \Delta r$ is divided into $K+1$ intervals of width Δr ($r + \Delta r = (K + 1)\Delta r$). The NLBC transforms the PE problem having a transverse radiation condition at infinity, into an equivalent one in a bounded domain, with the convolution coefficients $\mathbf{g}_{1,j}$ and

$$\beta = \frac{\rho_w}{\rho_b}k_0\sqrt{\frac{(\mathcal{N}_b^2 - 1)(1 + \frac{1}{4}\nu^2) + \nu^2}{1 + \frac{1}{4}\nu^2}}, \tag{4}$$

where $\nu^2 = 4i/k_0\Delta r$, and the subscripts w and b indicate the water column and bottom, respectively. For further details of the NLBC derivation including algebraic expressions for the coefficients $\mathbf{g}_{1,j}$ see [6] and [2].
In OAT inversion problem, we seek to determine the sound speed profile $c(z)$ knowing the acoustic pressures

caused by a multiple frequencies source ($f_l,\ l = 1, ..., L$) and collected by a sparse receiver array. The variational formulation of this problem is to introduce a cost function that we write:

$$J(c(z)) = \frac{1}{T}\sum_{l=1}^{L} J_{o,l}(c(z)) + J_b(c(z)), \qquad (5)$$

where $J_{o,l}(c(z))$ is a likelihood term and $J_b(c(z))$ is the background term. The parameter T is a coefficient. It must be chosen so as to achieve a better balance between the likelihood term $\sum_{l=1}^{L} J_{o,l}$ and the background term J_b. Thereafter, we will talk about the two components of the cost function (5).

Likelihood: The quantity $J_{o,l}(c(z))$ is a likelihood term witch quantify the mismatch between the measurements (acoustic signals $s_j(t),\ j = 1, ..., N$) across an N-element vertical array at the frequencies $f_l,\ l = 1, ..., L$ and the predicted (replica) field vector $\psi_l = G_{f_l}(c(z))$. In this paper we have chosen a likelihood term, used in a meta-heuristic inversion method by Hermand and Gerstoft [7], and that we write

$$J_{o,l}(c(z)) = tr\hat{\mathbf{R}}_l - \frac{\psi_l^{\dagger}\hat{\mathbf{R}}_l\psi_l}{\psi_l^{\dagger}\psi_l}, \qquad (6)$$

where † is the Hermitian transpose operator, tr is the trace operator, $\hat{\mathbf{R}}_l$ are the estimated spatial correlation matrices at the frequencies $f_l,\ l = 1, ..., L$ and $\psi_l^{\dagger}\hat{\mathbf{R}}_l\psi_l$ is the linear Bartlett processor. Matrices $\hat{\mathbf{R}}_l$ are computed using the acoustic signals $s_j(t)$ [7].
The gradient of $J_{o,l}(c(z))$ with respect of the parameters $c(z)$ is computed by the adjoint method. We write $J_{o,l}(c(z)) = J_{o,l}(G_{f_l}(c(z)))$ and the gradient $\nabla_c J_{o,l} = \mathbf{G}^*_{f_l}\nabla_\psi J_{o,l}$ where the linear operator $\mathbf{G}_{f_l} = \frac{\partial G_{f_l}}{\partial c}$ is the so-called tangent linear model and $\mathbf{G}^*_{f_l}$ represents the adjoint model. In the experiment presented in section 5 we implemented the adjoint-model by using the semi-automatic adjoint code generator YAO [8]. For further details see [1] and [2].

Background: The background term in (7) is the usual expression used for variational data assimilation in meteorology and oceanography [9, 10, 11]:

$$J_b(c(z)) = (c(z) - c_b(z))^T B^{-1}(c(z) - c_b(z)) \qquad (7)$$

which is based on probabilistic formalism. This term corresponds to an *a priori* probability on the physical parameters that we assume to follow a normal distribution $\mathcal{N}(c_b, B)$, c_b is the background and B is the variance-covariance matrix. Due to the high dimensionality of the vectors c(z) and the strong correlation between their components it becomes difficult to estimate the matrix B^{-1}. A more adequate approach for the minimization process consists in performing a transformation of the parameters into non correlated ones (or almost non correlated). Such transformations are used in meteorology and oceanography: in [9, 10] the $B^{1/2}$ transformation is computed by using a recursive filter, or in [11] an empirical decomposition introducing a physic knowledge is performed. The computation (or the approximation) of the non correlated parameter vector $u = B^{-1/2}(c(z) - c_b(z))$ allows to rewrite the cost function (5) which becomes $J(u) = \frac{1}{T}\sum_{l=1}^{L} J_{o,l}(u) + \frac{1}{2}u^T u$. This expression avoids calculating the inverse of matrix B, and provides a better preconditioning for the minimization process. Another transformation consists in using the probabilistic PCA model.

3 Probabilistic PCA

Consider a data set $\mathcal{A}$ made of sound speed profiles $\mathbf{c}$ evaluated at M points of the space discretization according to depth z. We denote by $\mathbf{c}_b$ the mean vector of $\mathcal{A}$. The Probabilistic PCA model [3, 12] allows a probabilistic interpretation of $\mathcal{A}$. It introduces an explicit latent variable $\eta \in \mathbb{R}^q$ ($q \ll M$) with a normal prior distribution $\mathcal{N}(0, \mathbf{I}_q)$ and assumes that the conditional vector parameter $\mathbf{c}/\eta$ is normally distributed with mean $\mathbf{W}\eta + \mu$ and isotropic covariance matrix $\kappa^2\mathbf{I}_M(M \times M)$, where $\mathbf{W}$ is a $(M \times q)$ matrix of range q and μ is a vector in $\mathbb{R}^M$.
The columns of $\mathbf{W}$ span a linear subspace $\mathbf{E}_q$ in $\mathbb{R}^M$ of dimension q and $\mathbf{W}\eta + \mu$ represents the associated affine linear variety, which contains the vector μ. Under these conditions, the profile $\mathbf{c}$ is normally distributed with mean vector μ and variance-covariance matrix:

$$B = \mathbf{W}\mathbf{W}^T + \kappa^2\mathbf{I}_M. \qquad (8)$$

Thus, η appears as a latent variable and $\mathbf{c}$ is a linear transformation of η plus an additive normal noise ε,

$$\mathbf{c} = \mathbf{W}\eta + \mu + \varepsilon, \qquad (9)$$

where $\varepsilon \hookrightarrow \mathcal{N}(0, \kappa^2\mathbf{I}_M)$.
For a given value of q, the parameters of probabilistic PCA are the matrix $\mathbf{W}$, κ^2 and μ. These parameters will be estimated by maximizing the likelihood of $\mathcal{A}$. It can be shown [3, 12] that the optimal solution μ is the mean $\mathbf{c}_b$ of the data set $\mathcal{A}$ and $\mathbf{W} = \mathbf{U}(\mathbf{L} - \kappa^2\mathbf{I}_q)^{1/2}\mathbf{R}$ where $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_q)$ is made of the first q eigenvectors of the empirical variance-covariance matrix of $\mathcal{A}$. $\mathbf{L}$ is a diagonal matrix $(q \times q)$, whose elements are the corresponding eigenvalues λ_i and $\mathbf{R}$ a rotation matrix of $\mathbf{E}_q$. This expression can be simplified by choosing $\mathbf{R} = \mathbf{I}_q$. Finally the determination of κ^2 gives:

$$\kappa^2 = \frac{1}{M - q}\sum_{i=q+1}^{M} \lambda_i. \qquad (10)$$

The sum $\sum_{i=q+1}^{M} \lambda_i$ represents the residual variance of the data not taken into account by the q first principal axes. Therefore, the κ^2 is the average of the residual variance of the $(M - q)$ remaining principal axes.
Thus, $\mathbf{W}\eta$ represents the element of $\mathbf{E}_q$ whose coordinates using $(\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_q)$ are $(\sqrt{\lambda_1 - \kappa^2}\eta_1, \sqrt{\lambda_2 - \kappa^2}\eta_2, ..., \sqrt{\lambda_q - \kappa^2}\eta_q)$, the variance of the i-th component ($1 \le i \le q$) equals $\lambda_i - \kappa^2$. This model assumes that the variance for the remaining principal axes ($i > q$) is constant and equal to κ^2.
Modeling $\mathcal{A}$ by a probabilistic PCA needs to take care when dealing with the choice of the value of q: the residual variance must be "low" enough in order to ensure that the model captures all the physical properties, and

the variances on the other principal axes ($i > q$) being "isotropic" enough. Formula (10) shows that κ^2 can be negligible in the case of OAT inversion where $M = 113$ and $q \leq 10$, κ^2 will be omitted in the following.
Taking into account this approximation, the probabilistic PCA assumes that a profile $\mathbf{c}$ is generated by:

$$\mathbf{c} = \mathbf{W}\eta + \mathbf{c}_b = \mathbf{U}\mathbf{L}^{1/2}\eta + \mathbf{c}_b, \qquad (11)$$

where η is the latent variable associated to $\mathbf{c}$, and which is normally distributed ($\eta \hookrightarrow \mathcal{N}(0, \mathbf{I}_q)$).
The cost function (5) becomes

$$\begin{aligned} J(\mathbf{c}) &= J(\mathbf{W}\eta + \mathbf{c}_b) \\ &= \frac{1}{T}\sum_{l=1}^{L} J_{o,l}(\mathbf{W}\eta + \mathbf{c}_b) + \eta^T\eta, \end{aligned} \qquad (12)$$

According to the latent variable η the cost function becomes

$$J(\eta) = \frac{1}{T}\sum_{l=1}^{L} J_{o,l}(\eta) + \eta^T\eta, \qquad (13)$$

which has to be minimized with respect to the latent variables η_i, $i = 1, ..., q$.

4 The Yellow Shark (YS) experiment

This section presents a twin experiment based on real data to retrieve the control variable profiles $\mathbf{c}$ using the probabilistic PCA approach.
The data were collected during the Yellow Shark (YS) experiment [7, 13]. It was carried out in the south of Elba island in the Mediterranean sea during the summer of 1994 and collected a series of 181 sound speed profiles $\mathbf{c}$ over 9 km. The distance between two profiles was 50 m. For each profile, measurements are made every meter, giving rise to a vector of varying dimension (between 113 and 116) depending on the depth of the water column. Because of the assumption of a stratified medium and that the probabilistic PCA need to project vectors of the same size, only the 113 first measurements have been considered here (M=113). For more details about the Yellow Shark data one can refer to [7, 13].
As it is shown in Fig. 1, the profiles are very similar (less than 5 m/s variability), except for a small interval around 20 m depth, corresponding to the thermocline area.
The profiles $\mathbf{c}$ obtained during the YS experiment represent the behavior of the sound speed profile during the period of the experiment and constitute the data set $\mathcal{A}$. The probabilistic PCA gives a model of this behavior, Fig. 2 represents the Cumulative Percentage of Total Variation (CPTV) relative with total energy, for each of first 15 PCA axes. With $q = 2$ axes the CPTV is approximately 76%, whereas with $q = 4$ axes it is more than 90%, it reaches 96% (resp. 98%) for $q = 7$ (resp. $q = 10$) axes. Taking into account the remark about the choice of q given in the preceding section and Fig. 2 it is clear that $q = 7$ and $q = 10$ could realize a good compromise. Twin experiments, using the probabilistic PCA and based on variational inversion, will be presented in the following section.

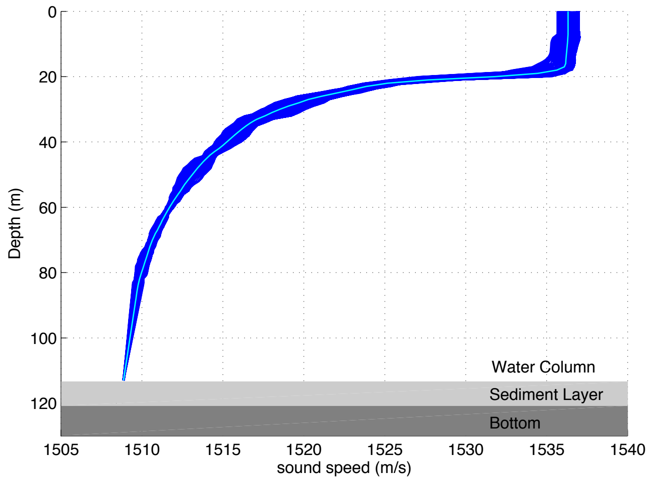

Figure 1: All YS sound speed profiles (blue) and the ensemble average (cyan). The sediment layer and bottom geoacoustic properties are supposed known.

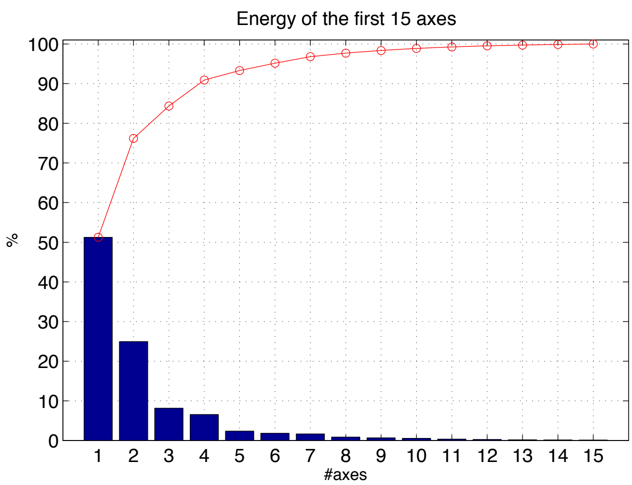

Figure 2: The energy of first 15 axes. The first ten axes concentrate almost 98% of total energy.

5 Inversion results

First the mean profile vector of the YS data ($\mathbf{c}_b$) was computed and $\mathbf{c}^*$ the farthest profile from $\mathbf{c}_b$, using the Euclidian distance, was selected. In the twin experiments $\mathbf{c}_b$ will be the initial background term used during the minimization and $\mathbf{c}^*$ the profile to be retrieved.
The YS experiments are used here as a realistic test case. A multiple frequencies source was placed at $z_s = 69.2$ m depth; the water depth was $z_l = 113.1$ m. The transmitted signal was received on a vertical array (VRA) of 32 hydrophones, 2 m spaced, from 37.2 to 99.2 m depth. The acoustic signals $s_j(t),\ j = 1, ..., 32$ are generated accordingly with the WAPE acoustic model $G_{f_l}(\mathbf{c}^*)$, using seven different source frequencies {200, 250, 315, 400, 500, 630, 800} Hz and a VRA at a range of 1 km. The sediment layer and the bottom are represented by the geoacoustic properties proposed in [7, 13].
Before computing the matrices $\hat{\mathbf{R}}_l,\ l = 1, ..., 7$, that we used to determine the likelihood term (6), we added a normally distributed noise of amplitude $\nu = 0.05$ to the acoustic signals

$$s_j(t) = s_j(t) + e(t),\ j = 1, ..., 32 \qquad (14)$$

where $e(t) \hookrightarrow \mathcal{N}(0, \nu^2)$.
We took the background $\mathbf{c}_b$ as initial profile for the minimization. Note that it corresponds to the latent variable $\eta = 0$. The choice of the parameter T of the cost function (13) has been done by the "L-Curve" method [14]. The twin experiments are performed using both 7 and 10 PCA axes. In both experiments, the probabilistic model is in the best conditions, indeed, $\kappa^2 \approx 0$ and the variances are virtually constant in all directions truncated by the PCA, especially for 10 axes.
The Fig. 3 illustrates the true centred profile (solid line) and both estimated centred profiles that have been found: the estimated profile using 7 axes (dotted) and the estimated one using 10 axes (dashed).

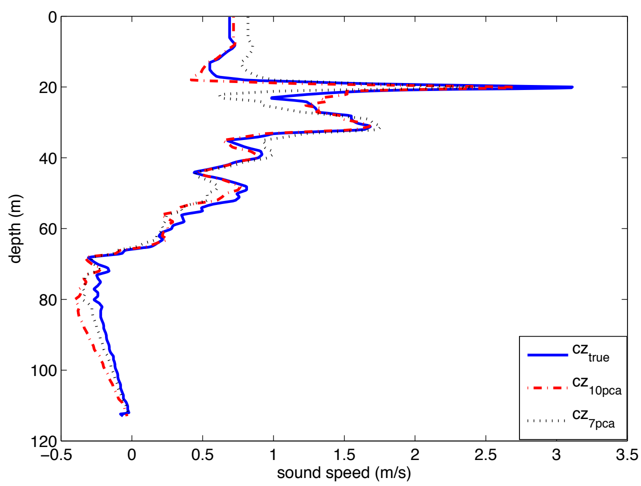

Figure 3: OAT results using probabilistic PCA approach. Starting with the background centred profile (which corresponds to zero) the true and both estimated centred profiles are shown. The true centred profile is indicated in solid line blue (cz_{true}), the estimated using 7 PCA axes is indicated in dotted black (cz_{7pca}) and the estimated using 10 PCA axes is indicated in dashed red (cz_{10pca}).

We note that in the area before and around the thermocline (approximately 30 m depth) the inversion result is much better with 10 axes that with 7. Between 30 and 80 m depth results are somewhat similar with a small advantage for 10 axes. After 80 m advantage east side 7 axes. One can say that overall the result is better when we take 10 axes compared to 7.
This can be explained by the fact that profiles reconstitution becomes better when increasing the PCA axes number. But from another point of view, the choice of the number q of the PCA axes generates a family of sound speed profile, which are increasingly rich with the value of q. We can consider the number q as a measurement of the degree of freedom of the corresponding generated family of sound speed. Thus, the risk that the minimizing process retrieves the noise, that we added to the acoustic signals, is more important for q = 10 than for q = 7. This risk was avoided here by the introduction of the background term (7) and the right choice of the parameter T.

6 Conclusion

In this paper we presented a variational approach to deal with an Ocean Acoustic Tomography (OAT) in shallow water inversion. This approach minimize a cost function, which measure the adequacy between the measurements and their forward model equivalence. We propose to add a background regularization term to this cost function, this suppose that the sound speed profile has an *a priori* normal distribution which depends on a variance-covariance matrix B. To deal with this regularization term in the cost function, we propose to model the sound speed profile by the probabilistic PCA model. We showed that this model could represent easily this regularization term and reduce significantly the number of the control parameters. We showed, in the context of a real data and a twin experiment with an added noise, that the variational inversion applied to a cost function with a background regularization term gives good results.

References

[1] F. Badran, M. Berrada, J. Brajard, M. Crépon, C. Sorror, S. Thiria, J.-P. Hermand, M. Meyer, L. Perichon, and M. Asch. Inversion of satellite ocean colour imagery and geoacoustic characterization of seabed properties: Variational data inversion using a semi-automatic adjoint approach. *Journal of Marine Systems*, 69:126–136, 2008.

[2] J.-P. Hermand, M. Meyer, M. Asch, and M. Berrada. Adjoint-based acoustic inversion for the physical characterization of a shallow water environment. *J. Acoust. Soc. Am.*, 119(6):3860–3871, 2006.

[3] M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statisitical Society*, 61:611–622, 1999.

[4] Claerbout. *Fundamentals of Geophysical Data Processing with applications to petroleum prospecting.* Blackwell Scientific Publications, 1976.

[5] D. J. Thomson. Wide-angle parabolic equation solutions to two range-dependent bench mark problems. *J. Acoust. Soc. Am.*, 87(4):1514–1520, 1990.

[6] David Yevick and David J. Thomson. Nonlocal boudary conditions for finite-difference parabolic equation solvers. *J. Acoust. Soc. Am.*, 106(1), July 1999.

[7] Jean-Pierre Hermand and Peter Gerstoft. Inversion of broad-band multitone acoustic data from the yellow shark summer experiments. *IEEE*, 21(4):324–346, October 1996.

[8] Sylvie Thiria, Fouad Badran, and Charles Sorror. Yao: Un logiciel pour les modèles numériques et l'assimilation de données. Technical report, LOCEAN-IPSL, Juin 2006.

[9] C. M. Hayden and R. J. Purser. Recursive filter objective analysis of meteorological fields: Applications to nesdis operational processing. *J. Appl. Meteor.*, 34:3–15, 1995.

[10] A. C. Lorenc. Interactive analysis using covariance functions and filters. *Quart. J. Roy. Meteor. Soc.*, 118:569–591, 1992.

[11] A.T. Weaver, E. Machu C. Deltel, S. Ricci, and N.Daget. A multivariate balance operator for variational ocean data assimilation. *Q. J. R. Meteorol. Soc.*, 131(613):3605–3625, 2005.

[12] Christopher M. Bishop. *Pattern Recognition and Machine Learning.* Springer, New York, 2006.

[13] Jean-Pierre Hermand. Broad-band geoacoustic inversion in shallow water from waveguide impulse response measurements on a single hydrophone: Theory and experimental results. *IEEE*, 24(1), 1999.

[14] Per Christian Hansen. Analysis of discrete ill-posed problems by means of the l-curve. *SIAM Rev.*, 34(4):561–580, 1992.