Ocean acoustic inversion with estimation of *a posteriori* probability distributions

Peter Gerstoft[a)]
SACLANT Undersea Research Centre, I-19138 La Spezia, Italy

Christoph F. Mecklenbräuker[b)]
Ruhr-University Bochum, D-44780 Bochum, Germany

(Received 2 June 1997; accepted for publication 16 April 1998)

Inversion methods are applied in ocean acoustics to infer parameters which characterize the environment. The objective of this paper is to provide such estimates, and means of evaluating the inherent uncertainty of the parameter estimates. In a Bayesian approach, the result of inversion is the *a posteriori* probability density for the estimated parameters, from which all information such as mean, higher moments, and marginal distributions can be extracted. These are multidimensional integrals of the *a posteriori* probability density, which are complicated to evaluate for many parameters. Various sampling options are examined and it is suggested that ''importance sampling'' based on a directed Monte Carlo method, such as genetic algorithms, is the preferred method. The formulation of likelihood functions and maximum-likelihood objective functions for multifrequency data on a vertical array is discussed. *A priori* information about the parameters may be used in the formulation. Shallow-water acoustic data obtained at several frequencies using a vertical array is used to illustrate the applicability of the technique. [S0001-4966(98)05307-7]

PACS numbers: 43.30.Pc, 43.60.Pt [DLB]

INTRODUCTION

From a Bayesian point of view, the solution to an inverse problem is fully characterized by *a posteriori* probability distributions of the unknown parameters. Information about these parameters is assessed by moments of the *a posteriori* distributions, such as the mean, covariance, and marginal distributions. This improves on the usual practice of calculating only a single point-estimate of the parameters: Accuracy of the inversion can be estimated in this way.

In a real problem, the numerical evaluation of *a posteriori* distributions is limited by computational resources. Importance sampling, as performed by simulated annealing (SA) and genetic algorithms (GA), can considerably reduce the number of operations required. The main advantage of the above concept is that it not only provides the best possible parameter estimates, but also calculates moments of the *a posteriori* distributions associated with these parameters. The Monte Carlo method[1] and the simulated annealing method[2] were developed as methods to evaluate multidimensional integrals.

Part of the inverse problem is to find an environment in which a forward model can produce a replica with a good match to the observed data. A representative environment is chosen empirically and a set of parameters $\mathbf{m}$ is selected as unknown. It is assumed that the true model vector is contained in the parameter set $\mathcal{M}$. An objective function $\phi(\mathbf{m})$ that compares the data and the replica is selected. Then, optimization is carried out to find the optimum model parameter vector $\hat{\mathbf{m}}=(\hat{m}^1,\ldots,\hat{m}^M)$ that minimizes the selected objective function.

To understand the inverse problem and its solution, it is important to study uncertainties, ambiguities, and resolution aspects of the unknown parameters. For deterministic local methods, these are limited to a neighborhood surrounding the local estimate, e.g., Refs. 3 and 4. A good discussion of these aspects for underwater acoustics is found in Ref. 5. Estimation of uncertainties from global methods have been discussed in Refs. 6–11 for the related geophysical problem, and in Refs. 12 and 13 for ocean acoustic problems. Retrieving parameters using global optimization methods has frequently been discussed in ocean acoustics.[12,14–16]

The superiority of stochastic global methods for uncertainty estimation stems from their ability to sparsely sample the parameter space $\mathcal{M}$.[17] For present applications, fast estimation of the moments of the *a posteriori* distributions is more important than precision.

The motivation for the present study is that *a posteriori* distributions as described in Ref. 12 have been used with success in a series of previous papers. It has worked well for making comparisons between retrieved parameters and giving an indication of the convergence of the optimization method. However, it is based on an empirical weighting and fails to give a performance indication for *different* inversion approaches. Only a likelihood based *a posteriori* distribution can show the improvement in performance when using more frequencies, or the differing performance using a near or far array.

Before measurement, the information about the models is reflected in the *a priori* distribution $\rho(\mathbf{m})$ and after the experiment, the information about the models is reflected in

[a)]Now at Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093-0704; Electronic mail: gerstoft@mpl.ucsd.edu
[b)]Now at Siemens A.G., A-1100 Vienna, Austria; Electronic mail: christoph.mecklenbraeuker@siemens.at

the *a posteriori* distribution $\sigma(\mathbf{m})$. These distributions are related through the likelihood function $\mathscr{L}(\mathbf{m})$, which is a measure for the goodness of fit between the observed data and the data generated using a computational acoustic model and the environment $\mathbf{m}$ (Bayes Theorem)

$$\sigma(\mathbf{m}) = \mathscr{L}(\mathbf{m})\rho(\mathbf{m}). \tag{1}$$

When maximizing $\sigma(\mathbf{m})$ the maximum *a posteriori* (MAP) estimate of the parameters is obtained and when maximizing $\mathscr{L}(\mathbf{m})$ the maximum likelihood (ML) parameter estimate is obtained.

To find the best solution, global optimization is carried out by minimizing an objective function $\phi(\mathbf{m})$. In this paper, the objective function is chosen proportional to the log-likelihood function, but often it is selected on a more empirical basis. The likelihood function depends on the stochastic model for the data: differing probabilistic models for signal and noise result in different likelihood and objective functions.

I. EVALUATION OF *A POSTERIORI* DISTRIBUTIONS

Due to multidimensionality, often $M > 10$, the *a posteriori* distribution is not susceptible to graphic display, and mainly integral properties of the distribution are of interest. From the *a posteriori* probability distribution, information will be extracted to describe the solution. The following quantities are of interest: the MAP solution $\hat{\mathbf{m}}^{\mathrm{MAP}}$ where

$$\hat{\mathbf{m}}^{\mathrm{MAP}} \equiv \arg \max_{\mathbf{m} \in \mathscr{M}} \sigma(\mathbf{m}), \tag{2}$$

the expectation $\mathsf{E}_\sigma[\mathbf{m}]$ where

$$\mathsf{E}_\sigma[\mathbf{m}] \equiv \int_{\mathscr{M}} \mathbf{m}\sigma(\mathbf{m})d\mathbf{m}, \tag{3}$$

where $d\mathbf{m} = dm^1 \cdots dm^M$, the covariance matrix $\mathsf{Cov}_\sigma[\mathbf{m}]$ where

$$\mathsf{Cov}_\sigma[\mathbf{m}] \equiv \mathsf{E}_\sigma\{[\mathbf{m} - \mathsf{E}_\sigma(\mathbf{m})][\mathbf{m} - \mathsf{E}_\sigma(\mathbf{m})]^T\}, \tag{4}$$

the one-dimensional (1D) marginal *a posteriori* probability densities $\sigma^i(m^i)$ for parameter m^i

$$\sigma^i(m^i) \equiv \int \sigma(\mathbf{m})dm^1 \cdots dm^{i-1}dm^{i+1} \cdots dm^M, \tag{5}$$

and higher dimensional marginals are defined similarly to Eq. (5). The marginal distributions are the most important in interpreting the inverse result.

A. Integration of *a posteriori* distributions

For solving the inverse problem, Eq. (2), global optimization methods such as simulated annealing and genetic algorithms have been used. However, in order to characterize the solution by Eqs. (3)–(5) the multidimensional integration must be carried out. The evaluation of these integrals has been addressed in Refs. 10 and 11 for the similar geophysical inversion problem. They suggested and evaluated the following three methods for estimation of the integral.

1. Numerical integration (grid search)

Although the most precise and direct method, numerical integration is extremely computationally intensive for M parameters each discretized to k values. It requires k^M forward model evaluations. For $k = 100$ and $M = 10$, this is a prohibitively large number 10^{20}, as for each of these points, a forward model that takes about one CPU-second must be evaluated. This approach is practical only for a very small number of parameters, e.g., $M \leqslant 4$.

2. Monte Carlo integration

In Monte Carlo integration the integration points are selected at random from a uniform distribution. It is not necessary to evaluate the integral at all points as in a grid search. In contrast to classical computing methods, the efficiency of Monte Carlo integration depends only weakly on the dimension and geometric details of the problem. Thus even for high dimensions, M, and complicated boundaries of the parameter set $\mathscr{M}$, the numerical effort remains moderate. The integral is evaluated at randomly selected points from a uniform distribution. The disadvantage is that many of these points will be located in areas contributing little to the integral.

3. Importance sampling

In importance sampling, some knowledge about the integrand is exploited such that a nonuniform distribution is used for generating the integration points. Most of the function evaluations are concentrated in areas which contribute significantly to the integral instead of distributing the points evenly. The integrals can be evaluated using fewer forward models at a reduced variance of the estimated integral. The model space $\mathscr{M}$ is sampled nonuniformly according to a generating distribution g. The integrand is evaluated only at these sample points. Both GA and SA use a generating distribution to select the next point in the model space.

Consider the evaluation of the multidimensional integral,[18]

$$\theta = \int_{\mathscr{M}} f(\mathbf{m})\sigma(\mathbf{m})d\mathbf{m} \equiv \mathsf{E}_\sigma[f(\mathbf{m})], \tag{6}$$

where $f\sigma$ represents any of the integrands given in Eqs. (3)–(5). The integral is estimated by the weighted arithmetic mean using N_{obs} independent identical distributed (i.i.d.) samples $\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_{N_{\mathrm{obs}}}$ from the distribution $g(\mathbf{m})$,

$$\hat{\theta} = \frac{1}{N_{\mathrm{obs}}} \sum_{i=1}^{N_{\mathrm{obs}}} \frac{f(\mathbf{m}_i)\sigma(\mathbf{m}_i)}{g(\mathbf{m}_i)}. \tag{7}$$

It can be shown, Appendix A and Ref. 18, that the variance of $\hat{\theta}$ is minimized if the generating distribution is selected

$$g^{\mathrm{MV}}(\mathbf{m}) = \frac{|f(\mathbf{m})|\sigma(\mathbf{m})}{\int_{\mathscr{M}} |f(\mathbf{m})|\sigma(\mathbf{m})d\mathbf{m}}. \tag{8}$$

Thus in order to reduce the variance, the generating distribution should be selected proportional to $|f|\sigma$. In practical cases, this optimal solution cannot be select due to complicated behavior of $|f|\sigma$. Further, if the integrand is known *a*

priori then there is no reason for estimating it. However, it is expected that selections of g which are ''close'' to g^{MV} will give satisfactory performance.

If the integral is evaluated using a generating distribution $g(\mathbf{m})$ without correcting for $g(\mathbf{m})$ in the denominator, it is seen from Eq. (A1) that the estimate will be biased. Since global estimation methods concentrate most of their numerical effort around the optimal values, they will tend to overestimate parts of the integral for these regions unless special bias corrections are introduced. This will result in an unknown error in the estimates.

Global optimization methods use a generating distribution for selecting the next model vector. This distribution will in general change as the optimization evolves. Thus they are carrying out importance sampling, but with an unknown generating distribution. For SA at a constant temperature T it can be shown that after a large number of iterations the sampling distribution is proportional to $\exp(-\phi(\mathbf{m})/T)$. For more details see Ref. 10 and Sec. I B.

When the MAP solution Eq. (2) is found using SA or GA, a large number N_{obs} of candidate solutions $\mathbf{m}_i$ ($i=1,...,N_{\mathrm{obs}}$) are drawn at random from the model set $\mathscr{M}$. By using the values of the objective function at these sample points $\mathbf{m}_i$, importance sampling can be used in evaluating the integrals Eqs. (3)–(5). When using GA or SA with fast cooling (as usual), the precise distribution of the samples is not known. It should also be made clear that the generating distribution is related in a nonlinear fashion (through GA or SA) to the objective function and not to the preferred distribution $|f|\sigma = |f|\mathscr{L}\rho$. This is not considered a problem as both $|f|$ and ρ are much smoother than $\mathscr{L}$ and the objective function is related to $\mathscr{L}$.

For the N_{obs} observations, the *a posteriori* probability for the kth model vector is estimated by

$$\hat{\sigma}(\mathbf{m}_k) = \frac{\mathscr{L}(\mathbf{m}_k)\rho(\mathbf{m}_k)}{\sum_{j=1}^{N_{\mathrm{obs}}} \mathscr{L}(\mathbf{m}_j)\rho(\mathbf{m}_j)}. \tag{9}$$

For the ith parameter m^i in the model vector the marginal probability distribution for obtaining the particular value κ can be found by summing Eq. (9):

$$\hat{\sigma}^i(\kappa) = \sum_{k=1}^{N_{\mathrm{obs}}} \hat{\sigma}(\mathbf{m}_k)\,\delta(m_k^i - \kappa), \tag{10}$$

where δ is the delta function. In the particular implementation, several independent GA searches are carried out in parallel. It is found for several parallel runs that saving the last obtained model vectors in a population suffices in each of the parallel runs.[12]

When displaying the marginal probability distributions, Eq. (5), they are all scaled so that the areas under each curve are one when the search interval is scaled from 0 to 1. This dictates the y axis and all plots from the same inversion have the same y axis, there is no scale on the y axis as it is mostly used for comparisons. With this choice of scaling the distributions depend on the search interval for each parameter.

B. Comparison to previous probability estimates

Previously,[12] the *a posteriori* distributions were estimated based on a semiempirical approach. Knowing that the likelihood function is usually related to the objective function $\phi(\mathbf{m})$ through an exponential $\mathscr{L} = \exp(-\phi(\mathbf{m})/\hat{\nu})$ (Ref. 6, cf. Appendix B), where $\hat{\nu}$ is the estimated noise power, the following scaling was used:

$$\mathscr{L}_{\mathrm{emp}}(\mathbf{m}) = \exp(-[\phi(\mathbf{m}) - \phi(\mathbf{m}_0)]/T), \tag{11}$$

where ϕ is any objective function and $\mathbf{m}_0$ is the estimated parameter vector, corresponding to the optimal value of the objective function. T is the ''temperature.'' Experimentally, it was found that a good value for T was the average of the 50 best objective functions obtained during the optimization, minus the best value of the objective function. It should be noted that this value of T is not intended to estimate the noise, but to produce a reasonable value with which to estimate the uncertainties of the parameters. The advantage of this scheme is that it works irrespective of the stochastic model for the data or likelihood function used. But for multifrequency inversions it is not suitable as different distributions cannot be directly compared. In this case, a likelihood-based *a posteriori* density, Eqs. (1) and (3)–(5) give better results.

C. Computational procedure

Depending on the model used for the error or noise distribution of the data, a specific likelihood function results. Prior knowledge about the error distribution is required. Usually this is not completely available and some simple and reasonable approximations must be used. Two special cases are considered for multifrequency vertical array data. The noise distribution on each hydrophone is assumed complex Gaussian and zero mean. First, in Sec. I C 1 it is assumed that the noise is independent on each *hydrophone* and second, in Sec. 1 C 2 it is assumed independent for each *significant mode*.

Often, a distinction is made between errors due to noise in the data and errors due to an incomplete forward model, because neither the theory nor the environmental model is adequate. If both error types belong to the same distribution, there is no reason to consider them separately.[19] Here only one error term is considered.

Recently, there has been progress in describing both errors using Kriging.[20] Both noise and modeling errors are assumed zero-mean Gaussian and are independent. The modeling errors are assumed to possess a given correlation structure depending on the ''distance'' between two environmental models. This correlation structure is chosen empirically. Clearly, the same values of the model parameters correspond to the same values of the model errors.

1. Multifrequency matched field processing

The relation between the observed complex-valued data vector $\mathbf{q}(\omega_l)$ on an N-element hydrophone antenna array and the predicted data $\mathbf{p}(\mathbf{m},\omega_l)$ at an angular frequency ω_l is described by the model

$$\mathbf{q}(\omega_l) = \mathbf{p}(\mathbf{m},\omega_l) + \mathbf{e}(\omega_l), \tag{12}$$

where $\mathbf{e}(\omega_l)$ is the error term. The predicted data is given by $\mathbf{p}(\mathbf{m},\omega_l)=\mathbf{w}(\mathbf{m},\omega_l)S(\omega_l)$, where the complex deterministic source term $S(\omega_l)$ is unknown. The transfer function $\mathbf{w}(\mathbf{m},\omega_l)$ is obtained using an acoustic propagation model and an environmental model $\mathbf{m}$.[21]

The errors are assumed to be additive, they stem from many sources: errors in describing the environment, errors in the forward model, instrument and measurements errors, and noise in the data. For the predicted acoustic field ''reasonably close'' to the true field, this error term is assumed complex Gaussian distributed, stationary with zero mean and diagonal covariance matrix $\nu(\omega_l)\mathbf{I}$, where the error power spectrum ν is unknown. Thus the data $\mathbf{q}(\omega_l)$ on the receiving array are also complex Gaussian distributed with mean $\mathbf{p}(\omega_l,\mathbf{m})$ and the covariance matrix $\nu(\omega_l)\mathbf{I}$. For the derivation of a maximum-likelihood estimate, it is further assumed that the data are uncorrelated across frequency and time. The source term $S(\omega_l)$ varies across time snapshots whereas the error power spectral density $\nu(\omega_l)$ is constant. In the following, $\mathbf{q}_l=\mathbf{q}(\omega_l)$, etc., is abbreviated, where $\{\omega_l|l=1,...,L\}$ is the processed frequencies. Under the above assumptions the covariance matrix $\mathbf{R}_l=\mathsf{E}[\mathbf{q}_l\mathbf{q}_l^\dagger]=\mathbf{p}_l(\mathbf{m})\mathbf{p}_l^\dagger(\mathbf{m})+\nu_l\mathbf{I}$. The likelihood function[22] becomes

$$\mathscr{L}(\mathbf{m})\propto\prod_{l=1}^{L}\ \frac{1}{\nu_l^N}\exp\left(-\frac{\phi_l(\mathbf{m})}{\nu_l}\right), \tag{13}$$

where (the dagger refers to the Hermitian transpose and ''tr'' is the trace operation)

$$\phi_l(\mathbf{m})=\mathrm{tr}\ \hat{\mathbf{R}}_l-\frac{\mathbf{w}_l^\dagger(\mathbf{m})\hat{\mathbf{R}}_l\mathbf{w}_l(\mathbf{m})}{\mathbf{w}_l^\dagger(\mathbf{m})\mathbf{w}_l(\mathbf{m})}. \tag{14}$$

Optimization for ν_l yields the closed form ML solution

$$\hat{\nu}_l=\frac{1}{N}\ \phi_l(\mathbf{m}). \tag{15}$$

The $N\times N$ Hermitian matrix $\hat{\mathbf{R}}_l$ denotes the estimated cross-spectral density matrix of the observed data in ''phone-space,'' see Sec. 1 C 3. With these definitions the log-likelihood function is

$$\log|\mathscr{L}(\mathbf{m})|\propto\log\left[\prod_{l=1}^{L}\ \phi_l^{-N}(\mathbf{m})N^N\exp(-N)\right]\propto-\log[\phi(\mathbf{m})], \tag{16}$$

where the ML-objective function $\phi(\mathbf{m})$ to be minimized is

$$\phi(\mathbf{m})=\prod_{l=1}^{L}\ \phi_l(\mathbf{m})=\prod_{l=1}^{L}\left(\mathrm{tr}\ \hat{\mathbf{R}}_l-\frac{\mathbf{w}_l^\dagger(\mathbf{m})\hat{\mathbf{R}}_l\mathbf{w}_l(\mathbf{m})}{\mathbf{w}_l^\dagger(\mathbf{m})\mathbf{w}_l(\mathbf{m})}\right). \tag{17}$$

For details, see Appendix B. Using a global optimization procedure, the minimum $\hat{\phi}^{\mathrm{ML}}$ for the ML solution $\hat{\mathbf{m}}^{\mathrm{ML}}$ is estimated. The estimate, Eq. (15), is biased. The bias stems from the degrees of freedom in the estimated parameters: source signal S and nonlinear parameters $\mathbf{m}$.[23] For simplicity this bias is neglected here. The noise power spectral density is estimated, Eq. (15), $\hat{\nu}_l^{\mathrm{ML}}=1/N\,\phi_l(\hat{\mathbf{m}}^{\mathrm{ML}})$, and the likelihood function is given by

$$\mathscr{L}(\mathbf{m})=p(\mathbf{m}|\mathbf{q})\propto\prod_{l=1}^{L}\ (\hat{\nu}_l^{\mathrm{ML}})^{-N}\exp\left(-\frac{\phi_l(\mathbf{m})}{\hat{\nu}_l^{\mathrm{ML}}}\right)$$

$$\propto\prod_{l=1}^{L}\ \exp\left(-N\,\frac{\phi_l(\mathbf{m})-\hat{\phi}_l^{\mathrm{ML}}}{\hat{\phi}_l^{\mathrm{ML}}}\right). \tag{18}$$

The problem, as addressed above, is then to integrate this multidimensional probability distribution. Often this integral can be evaluated with sufficient accuracy using the information from the global search. In some cases it might be necessary to increase the sampling of the model space in order to obtain convergence.

The likelihood function, Eq. (18), has a stronger maximum when more hydrophones are used. When inverting observed data, there is a limit to how much useful information can be obtained by adding additional hydrophones, as they then become strongly correlated. At high signal-to-noise ratio (SNR) it is expected that the main error contribution is due to inadequate forward modeling. Further, the number of uncorrelated hydrophones is approximately the same as the number of propagating modes, because this limits the degrees of freedom in the random part of the acoustic wave field. The number of uncorrelated hydrophones is estimated as the rank of the covariance matrix.

2. Multifrequency matched mode processing

Normal modes provide a complete description of the field at long ranges, and thus one can equivalently process the data in the phone-space or in the modal-space. The matched mode approach is described by Tolstoy,[24] Hinich,[25] and Shang.[26] Modal processing is discussed here as an alternative noise estimate when there are more hydrophones than propagating modes.

The observed field of N sensors is assumed approximately expressed via a set of J *significant* normal modes, expressed in a $N\times J$ matrix $\mathbf{V}(\omega_l,\mathbf{m})=[\mathbf{v}_1(\omega_l,\mathbf{m}),\ldots,\mathbf{v}_J(\omega_l,\mathbf{m})]$. The typical vector $\mathbf{v}_j(\mathbf{m})$ contains spatial samples of the jth normal mode at the receiver array locations. The set of normal modes will be determined based on the environment. The corresponding complex valued modal amplitudes (the breve refers to the mode-space) $\breve{\mathbf{q}}(\omega)=(\breve{q}_1,\ldots,\breve{q}_J)'$.

$$\mathbf{q}(\omega_l)\approx\sum_{j=1}^{J}\ \mathbf{v}_j(\omega_l,\mathbf{m})\breve{q}_j=\mathbf{V}(\omega_l,\mathbf{m})\breve{\mathbf{q}}(\omega). \tag{19}$$

There should be more hydrophones than modes, $N>K$. This relationship can be inverted in a least-squares sense and estimates the vector of modal amplitudes $\breve{\mathbf{q}}_l=\breve{\mathbf{q}}(\omega_l)$ in the mode-space from the observation $\mathbf{q}_l=\mathbf{q}(\omega_l)$ in phone-space.

$$\breve{\mathbf{q}}_l=[\mathbf{V}_l^\dagger(\mathbf{m})\mathbf{V}_l(\mathbf{m})]^{-1}\mathbf{V}_l^\dagger(\mathbf{m})\mathbf{q}_l. \tag{20}$$

Note that the modes $\mathbf{V}_l(\mathbf{m})$ and thus modal amplitudes depend on the environment. When optimizing the environment $\mathbf{m}$ the modal amplitudes will change with the environment.

A simple relationship between the observed modal amplitudes $\breve{\mathbf{q}}_l$ and synthetic generated modal amplitudes is assumed

$$\breve{\mathbf{q}}_l = \breve{\mathbf{p}}_l(\mathbf{m}) + \breve{\mathbf{e}}_l(\mathbf{m}), \tag{21}$$

where $\breve{\mathbf{p}}_l(\mathbf{m}) = S_l \breve{\mathbf{w}}_l(\mathbf{m})$ is the complex-valued modal amplitudes of the synthetic data and $\breve{\mathbf{e}}_l$ represents the error term for each mode. It is assumed that the noise covariance matrix is diagonal for the J significant modes and the noise power $\breve{\nu}$ is identical for all J modes. Using a similar approach to that in Sec. 1 C 1, the objective function is

$$\breve{\phi}_l(\mathbf{m}) = \mathrm{tr}\, \hat{\hat{\mathbf{R}}}_l - \frac{\breve{\mathbf{w}}_l^\dagger(\mathbf{m}) \hat{\hat{\mathbf{R}}}_l \breve{\mathbf{w}}_l(\mathbf{m})}{\breve{\mathbf{w}}_l^\dagger(\mathbf{m}) \breve{\mathbf{w}}_l(\mathbf{m})}, \tag{22}$$

where $\hat{\hat{\mathbf{R}}}_l$ is the estimated covariance matrix of the modes. But using the expression for the modes, Eq. (19), the objective function in mode-space, Eq. (22), is seen to be equivalent to the objective function in phone-space, i.e., Eq. (14) expressed in the phone-space

$$\breve{\phi}_l(\mathbf{m}) \approx \phi_l(\mathbf{m}). \tag{23}$$

The noise estimate is obtained using the approximation in Eq. (19),

$$\hat{\hat{\nu}}_l = \frac{1}{J}\, \hat{\hat{\phi}}_l^{\mathrm{ML}} \approx \frac{1}{J}\, \hat{\phi}_l^{\mathrm{ML}}. \tag{24}$$

The likelihood function becomes

$$\mathscr{L}(\mathbf{m}) = p(\mathbf{m}|\mathbf{q}) \propto \prod_{l=1}^{L} (\hat{\hat{\nu}}_l^{\mathrm{ML}})^{-J} \exp\left(- \frac{\breve{\phi}_l(\mathbf{m})}{\hat{\hat{\nu}}_l^{\mathrm{MD}}} \right)$$
$$\propto \prod_{l=1}^{L} \exp\left(-J\, \frac{\phi_l(\mathbf{m}) - \hat{\phi}_l^{\mathrm{ML}}}{\hat{\phi}_l^{\mathrm{ML}}} \right). \tag{25}$$

The advantage of this formulation is that it does not depend directly on the number of hydrophones, but only on the number of propagating modes. For many hydrophones ($N \gg J$) this likelihood function seems more realistic. A formal definition for J is not yet clear. In either case, J is less or equal to the number of propagating modes. For simplicity, J is assumed to be independent of frequency. Only the objective function is affected by the choice of J. All the propagating modes are incorporated in the forward model.

3. Estimation of the covariance matrix

In order to estimate the covariance matrix $\mathbf{R}_l$, the received time signal is divided into K time frames. Each frame was short-time Fourier transformed using the multiple-windows technique described in Refs. 27 and 28,

$$\mathbf{q}_{k,p}(\omega) = \sum_{t=0}^{T-1} \nu_t^p \mathbf{q}(t+kT) e^{-j\omega t}, \quad \text{for } \begin{cases} k = 0,\ldots,K-1 \\ p = 0,\ldots,P-1 \end{cases}, \tag{26}$$

where ν^p is a special set of P orthonormal data tapers.[27,28] The correlation matrix $\mathbf{R}$ was estimated at each selected frequency ω_l as the ensemble average

$$\hat{\mathbf{R}}(\omega_l) = \frac{1}{KP} \sum_{k=0}^{K-1} \sum_{p=0}^{P-1} \mathbf{q}_{k,p}(\omega_l) \mathbf{q}_{k,p}^\dagger(\omega_l). \tag{27}$$

In order to obtain a good estimation of the noise, it is required that $KP \gg N$, where N is the number of hydrophones. In order to ''just'' estimate the signal and the unknown parameters $\mathbf{m}$, the number of averages KP can be much smaller for a received signal with sufficient SNR. In this paper, the evaluation of the inversion accuracy is addressed. Therefore modeling errors in $\mathbf{p}(\mathbf{m},\omega)$ and additive noise must be distinguished. This implies the necessity of a stable noise estimate (27), and thus a larger number of averages than if only parameter estimates $\hat{\mathbf{m}}$ are needed.

D. Including the *a priori* probability distribution

Usually, when solving inverse problems, the question is, ''What is the environmental model for this given data set?'' This is normally an ill-posed question. A better question is, ''What can be inferred from the data about the environmental model given some environmental information?'' Thus some *a priori* information should be included in the inverse problem. *A priori* information is always used in global inversion schemes. The model structure is selected based on *a priori* knowledge and uniform *a priori* distributions are used between the minimum and maximum bounds for the parameters.

One possibility is to include the *a priori* model in the objective function, e.g., Refs. 6 and 29. This has the disadvantage that the distribution must be known explicitly. For Gaussian *a priori* distribution the objective function consists of two terms, one measuring the match between observed and synthetic data and the second penalizing the deviation from the *a priori* model. This approach is used in linearized inversions in order to regularize the solution.

Here a simple approach is used: the obtained likelihood function is multiplied with the *a priori* distribution, according to Eq. (1). This distribution can be arbitrary, for example, a smoothed distribution obtained from inversion of other data. For illustration in Sec. II A 3 a simple triangular distribution is used:

$$\rho(m^i) \propto \begin{cases} (m^i - m_u^i)/(m_m^i - m_u^i), & \text{for } m_m^i < m^i < m_u^i \\ (m^i - m_l^i)/(m_m^i - m_l^i), & \text{for } m_l^i < m^i < m_m^i \\ 0, \quad \text{otherwise}, \end{cases} \tag{28}$$

where $m_l^i < m_m^i < m_u^i$ are the abscissa of lower bound, maximum, and upper bound of the *a priori* distribution, respectively.

A priori information is also used in the parametrization of the forward model. The choices made when doing this have a significant influence on the inverse solution, probably more than including *a priori* information for each parameter. In discretizing the environment, the physics should be carefully considered and described efficiently. Shape functions[30] are a useful method to obtain an efficient description which provides a mapping between the environmental model and the numerical forward model. This could, for example, be used to limit the search to only positive gradients in the sediment, or to obtain a more efficient description of the environment (see the example in Sec. II A 3).

It is assumed for simplicity that there is a vanishing correlation between the *a priori* distributions of the indi-

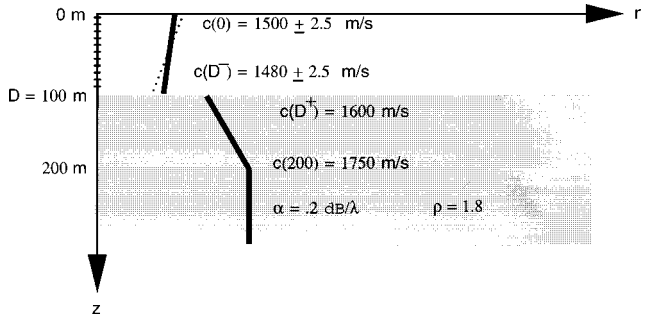

FIG. 1. The environment for the sspmis case. The source coordinates are (9.3 km, 78 m) for SNR=40 dB and (9.6 km, 82 m) for SNR=10 dB.

vidual parameters. Thus $\rho(\mathbf{m})=\rho^1(m^1)\rho^2(m^2)\cdots$. Although for correlated parameters, the model search can be limited using a correlated *a priori* distribution, this approach appears unpractical and, instead, shape functions are used for mapping correlated model vectors to a new representation with lower correlation.

II. EXAMPLES

The examples illustrate the integration of the *a posteriori* distributions and how phone-space and mode-space based likelihood functions affect the estimates. Only in Sec. II A 3, a nonuniform *a priori* distribution is used.

The objective function Eq. (17) was optimized and the SNAP normal mode code[21] was used as a forward model. For likelihood functions, either the empirical, the multifrequency matched field, or the multifrequency matched mode models are used. The GA parameters were as in Refs. 12 and 31: the reproduction rate was 0.5, the permutation probability was 0.05, and the crossover rate was 0.8.

A. SSP-mismatch case

This case corresponds to the sound-speed mismatch from the 1993 Matched field workshop,[32] Fig. 1. It is based on a synthetic data set from a normal-mode code using a 250-Hz source in shallow water. The data are received on the 20 hydrophones spanning the entire water column. White Gaussian noise was added to the data vectors to obtain a SNR of either 40 or 10 dB.[32] This corresponds precisely to the likelihood function developed in Sec. I C 1.

Only four parameters are unknown in this case; the source range and depth and the ocean sound speed at the top and the bottom. Each of the parameters can assume 51 discrete values. For an exhaustive search this requires evaluation of $51^4=7\times10^6$ forward models. It took 5 days of CPU time on a DEC-Alpha 500/266 to evaluate all models. For the genetic algorithm 4×10^4 forward models were evaluated in half an hour of CPU time.

1. Without a priori information

A uniform *a priori* distribution is assumed. To display the marginal distribution, the integral, Eq. (5), is evaluated. When using an exhaustive search, i.e., evaluating each point in the integration corresponding to Sec. I A 1, the result

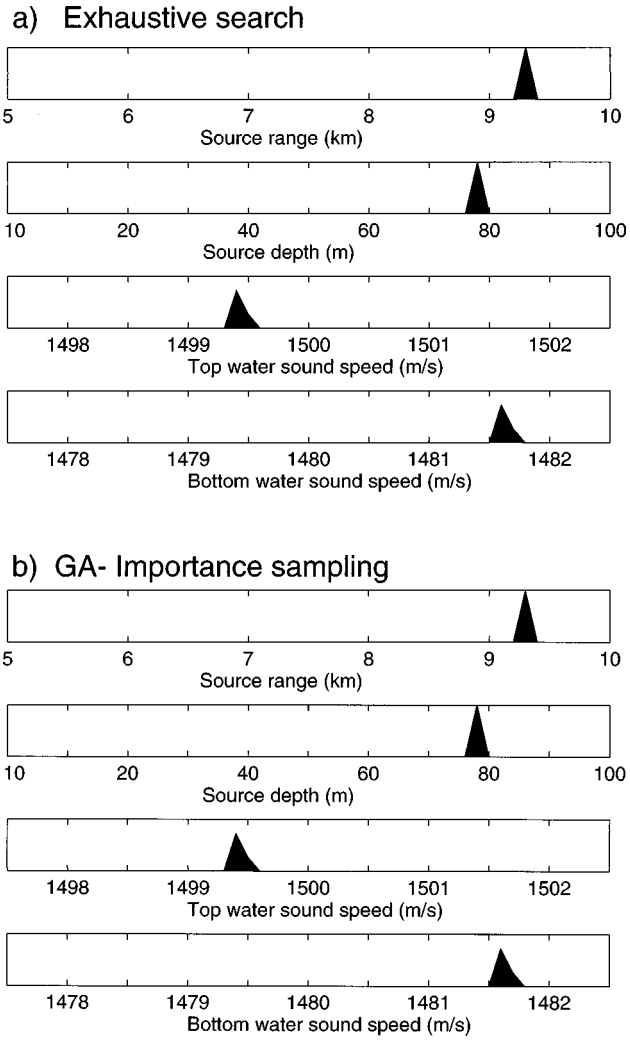

FIG. 2. The estimated *a posteriori* distribution for a SNR of 40 dB for the sspmis case. (a) using numerical integration, (b) using importance sampling.

of this integration is shown in Figs. 2(a) and 3(a). In generating these plots, the noise was assumed to be unknown.

When carrying out the optimization by SAGA,[31] 20 parallel runs, each sampling 2000 models with a population size of 64 were used. One result is the ML estimate of parameters which corresponds to the best obtained fit.

During this optimization the last population (64 individuals) in each of the 20 runs is saved in order to estimate the integrals. Thus the estimation of the integral is based on $20\times64=1280$ model vectors. It is seen in Figs. 2(b) and 3(b) that the evaluation of the integral resembles the distributions obtained when evaluating the integrals based on an exhaustive search. The use of all the GA-evaluated forward models ($20\times2000=4\times10^4$) in the evaluation of the integral did not have any effect. For practical reasons, it is preferred to base the evaluated integral on the last population in each run.

Note that for both SNR=40 and 10 dB the source range and depth are estimated quite accurately, whereas for more noise, the ability to resolve the sound-speed parameters is lost when the SNR is relaxed to 10 dB. The reason for this poor resolution of the parameters is due to a strong correlation between the ocean sound speed at the bottom with that at the top; see Fig. 4, where the two-dimensional marginal

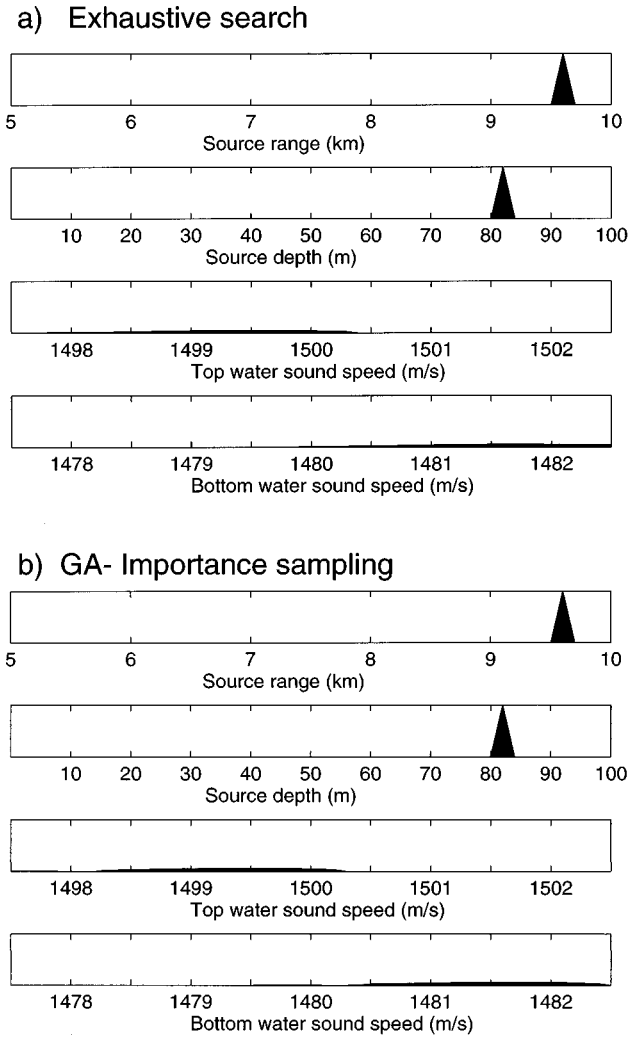

FIG. 3. The estimated *a posteriori* distribution for a SNR=10 dB for the sspmis case. (a) Using numerical integration, (b) using importance sampling.

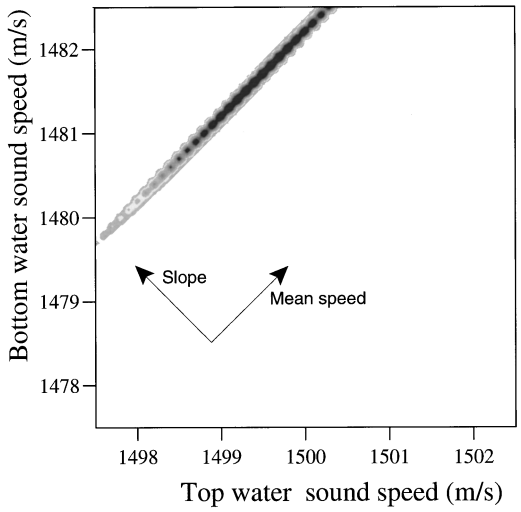

FIG. 4. The estimated marginal 2-D *a posteriori* distribution between the upper and lower ocean sound speed using a full search for a SNR of 10 dB. It is based on the same data as Fig. 3(a). The marginal 1-D distribution for each parameter is displayed on the top and to the right. By reparametrizing the sound speed as slope and mean a better resolution is obtained.

distribution is shown. From this figure it is clear that a better parametrization would be the mean sound speed and the gradient of the sound speed, as indicated in the figure and also discussed in Ref. 12. The difference in ambiguity for the sound speeds estimated at SNR of 40 dB (Fig. 2) and 10 dB (Fig. 3) is due to the fact that at 40 dB, slope and mean sound speed are well resolved, whereas at 10 dB SNR only the slope is well resolved.

2. Optimizing coupled parameters

Coupled parameters usually render an optimization problem slightly more difficult. Parameter coupling has been observed by several researchers.[12,33–35] Coupling can be detected by plotting either the ambiguity function or, alternatively, the 2-D marginal *a posteriori* distribution of the parameters. The advantage of the second approach is that it provides an integrated (global) value across the remaining parameters. However, both approaches have limitations when several parameters are strongly coupled.

For a gradient method coupled parameters do not pose a major problem. A difficulty with gradient methods is the numerical computation of the gradient. When using finite differences for computing, the gradient methods tend to be unstable due to a too small or large step size. Fortunately, it is possible to compute the gradient analytically for a wavenumber integration approach[30] and for a normal-mode approach.[36,37]

Some global search methods additionally exploit gradient information. In ocean acoustics this has been proven successful in Refs. 30 and 34. In Ref. 30, the optimization is a hybrid method combining the global genetic algorithm (GA) with the local Gauss–Newton method. This is implemented by taking several gradient steps between each update of the object function for each individual in the GA population. This approach is quite general but requires a careful analysis of the gradient computation, which was done analytically (in order to avoid huge numerical errors). In Ref. 34, a parameter rotation approach was suggested. The eigenvectors of the *a priori* second moment of the objective function gradient define the transformation for rotating the parameter space. The second moment is defined by implicitly assigning a uniform *a priori* density to the parameters. The computed eigenvectors provide some insight into the geometry of parameter space. After reparametrization, the search proceeds using SA. This approach is efficient if the parameter space is characterized by a few local minima with prominent features in one direction. In cases where the gradient information averages out, this will not provide an improvement. Such cases include circular shaped valleys, landscapes with several valleys, or landscapes with several valleys, or landscapes that are hilly without trend.

3. Including the a priori distribution

A priori knowledge is incorporated using Eq. (28) and based on the former example. In this case the *a priori* knowledge has a maximum at the true value. Initially only prior information of the top sound speed is used, Fig. 5(b). It is seen that this increases the peak of both the upper and lower

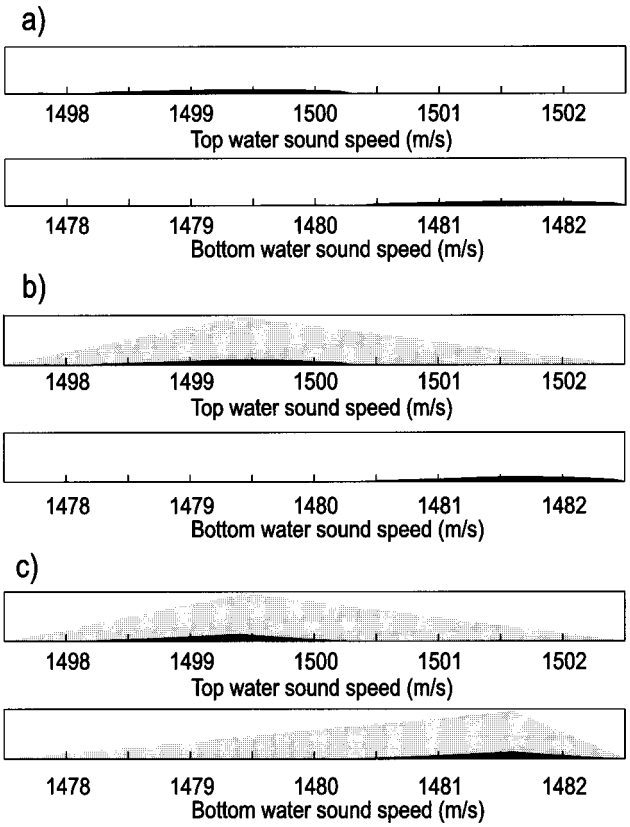

FIG. 5. The estimated *a posteriori* distribution using GA for SNR=10 dB. Only the two sound-speed parameters are displayed. (a): no *a priori* information is used [similar to Fig. 3(b)] in (b) *a priori* knowledge (grey area) about the top sound speed is included and in (c) *a priori* knowledge (grey area) about both top and bottom sound speeds are included. The use of *a priori* knowledge has sharpened the peaks marginally.

sound-speed point. This is because *a priori* knowledge is multiplied on the full *a posteriori* distribution. When using prior information for the lower and upper sound speed, the peak in the distribution becomes more pronounced, Fig. 5(b). It is clear that wrong prior information must bias the estimate.

Prior information about the environment, i.e., measured sound speeds, and the main parameters in terms of their search interval is important for obtaining good inversion results. Incorporation of smooth *a priori* distributions such as Eq. (28) does not seem significant, as the *a posteriori* distribution does not change much.

B. Yellow Shark data

This example is based on the SACLANTCEN Yellow Shark 94 (YS-94) experiment. YS-94 was a carefully designed major experiment in shallow water (100 m) south of Elba in the Mediterranian Sea. A fixed source–receiver geometry was used and a comprehensive environmental data set was available: Sea surface temperature, sea surface motion, currents, 2D temperature/salinity structure along transect, cores, and high resolution seismics. For a detailed description see Ref. 38. In the data used here the source was located at a 9-km range from a vertical array extending the complete depth of the water column, it transmitted energy at 7 frequencies: 200, 250, 315, 400, 500, 630, and 800 Hz. The

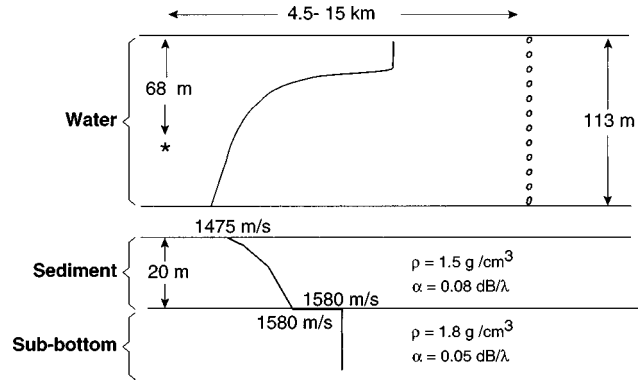

FIG. 6. **ysmodels**. The baseline environmental model used in the inversion of the YS-94 data Ref. 38.

vertical array consisted of 32 hydrophones. The SNR was estimated to about 30 dB.[38] The environment is shown in Fig. 6. The covariance matrix estimate was based on $K=17$ time frames and $P=4$ orthogonal windows, as described in Sec. 1 C 3.

1. Single frequency

First it will be discussed how the *a posteriori* distribution is constructed. This was done from a theoretical point of view in Sec. I, but here a more practical approach is taken by (a) discussing of the convergence of the objective function, (b) constructing the likelihood function from the objective function and, (c) construction of *a posteriori* distribution from the likelihood function.

To investigate the above, a single frequency inversion for the model in Fig. 6 is performed as it is much faster than the more accurate broadband inversion. The unknown part of the environment is represented by either 5, 7, or 10 parameters, as indicated in Table I. Clearly, which parameter sets are found depends on the number of iterations in a search. The search is performed with either 5000 forward modeling runs (split into ten independent populations each with 500

TABLE I. Parameter search bound for the YS-94 case. Each parameter was discretized into 64 values. The 5, 7, and 10 parameters refers to the number of parameters used in the inversion. The bottom sound-speed profile was modeled using the increase from the previous sound-speed point, as is common in SAGA. The receiver depth is the depth of the deepest hydrophone, this controls the vertical position of the entire vertical array in the water column.

Parameter	Lower	Upper
5, 7, and 10 parameters:		
Source range (m)	7	11
Source depth (m)	65	75
Tilt (m)	−3	3
Water depth (m)	110	118
Receiver depth (m)	96	104
7 and 10 parameters:		
Bottom sound speed at interface (m/s)	1460	1500
Bottom sound-speed increase at 5 m (m/s)	10	50
10 parameters:		
Bottom sound-speed increase at 10 m (m/s)	10	50
Bottom sound-speed increase at 20 m (m/s)	10	50
Bottom attenuation (dB/λ)	0	0.4

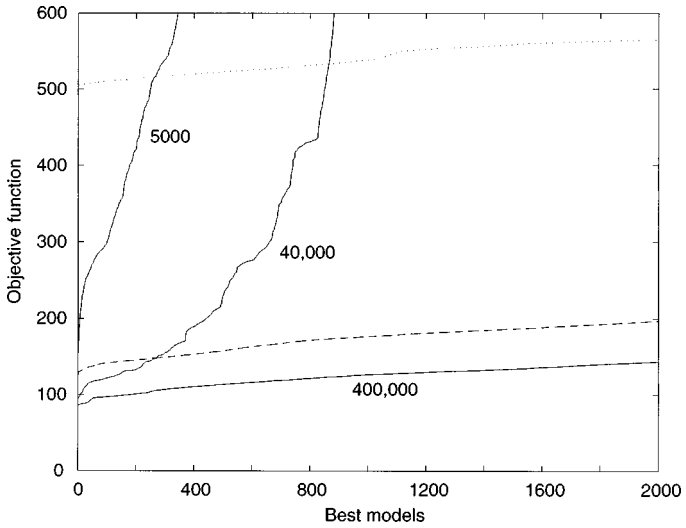

FIG. 7. Sorted values of the objective function, Eq. (17), for the best estimated models for each optimization. The best models are the ones that gave the lowest value of the objective function. Solid lines: The best values of the objective function based on either 4×10^5, 4×10^4, or 5×10^3 evaluations of the forward model with ten unknown parameters. Dashed line: based on 400 000 evaluations with seven unknown parameters. Dotted line: based on 400 000 evaluations with five unknown parameters.

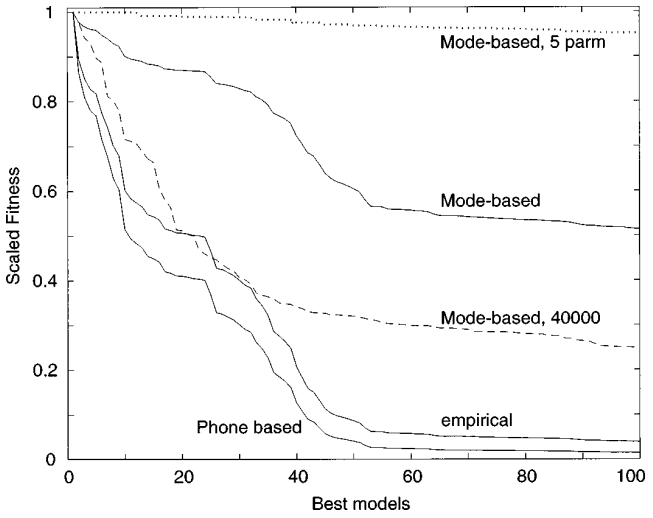

FIG. 8. Sorted values of likelihood functions using either an empirical, Eq. (11), a phone-based, Eq. (18), or a mode-based, Eq. (25) with $J=5$, likelihood function. Solid lines: Based on 4×10^5 forward model evaluations and estimation of ten parameters with either empirical, phone-based, or mode-based likelihood function. The dashed line is based on 4×10^4 forward model evaluations and estimation of ten parameters and mode-based likelihood function. The dotted line is based on 4×10^5 forward model evaluations, estimation of five parameters and mode-based likelihood function.

forward modeling runs [10×500]), 40 000 forward modeling runs [20×2000], or 400 000 forward modeling runs [200×2000].

In Fig. 7 the sorted values of the objective function are displayed for the best models. The best models are the one that gave the lowest value of the objective function. It is seen that for ten parameters we obtain a lower value of the objective function than when using seven or five parameters. Clearly, if more free parameters are available it is possible to obtain a better fit. The curve with ten parameters gives the best fit. Whether this improved fit is significant or whether the extra parameters are just fitting additive noise can be tested.[39] When using more forward modeling runs, more samples with a high degree of fit are obtained, as can be seen by comparing the curves for 5000, 40 000, and 400 000 forward modeling runs in Fig. 7.

Based on these values, the likelihood function (weighted fitness) is computed for each of these models, see Fig. 8. It is seen from Eqs. (15) and (18) that the noise estimate depends on the best estimated value of the objective function and, therefore, the value of the likelihood function depends on the search. If this best value is much better that the other value found during the optimization, the likelihood function, Eq. (18), will decrease quite rapidly (compare the curve for five parameters with those for ten parameters). How fast the curve decreases also depends on the number of modes used in the objective function. Note that even though the objective function assumes only a few modes the forward model always includes all propagating modes.

Based on the weighted fitness, as displayed in Fig. 8, the integrals for the marginal distributions, Eq. (5), are estimated. It was found that when using the noise estimate, Eq. (15), with the likelihood function, Eq. (18), to estimate the *a posteriori* distribution that the distributions became single peaked. This is probably due to an underestimation of the noise, because the data on each hydrophone is correlated. For long-range propagation, the pressure field can be described as a sum of modes. Thus it is expected that the denominator in Eq. (15) should express the number of propagating modes, see Eq. (24).

Both the semiempirical weighting and the ML weighting using five modes in the noise estimate and estimating five parameters when 400 000 forward models is used, are shown in Fig. 9(a) and (b) when 400 000 forward models are used in the optimization. It should be noted that the objective function as well as the samples used in the estimation of the objective function are identical for both methods. Thus the difference in the plot is entirely due to different weighting of the objective function when constructing the *a posteriori* distributions. Intuitively, the empirical estimates appear to overestimate the resolution, whereas the ML gives a more realistic estimate of the peak. The empirical estimates of the variance depend on the number of forward modeling runs. The variance will not be as small if a smaller number of forward modeling runs was used during the optimization.

For the results of the optimization with 5 parameters, it is seen that the parameter estimate of the source depth reaches the upper bound. This indicates that the optimization has not performed well, probably because the environment has not been well described. When using nine parameters it is seen that the source depth becomes more stable. Again it is seen that the empirical weighting gives a more optimistic estimate of the uncertainties.

Comparing the likelihood based results with five or nine parameters [Fig. 9(b) and (d)], it is seen that the spread of the distributions is about the same. The estimated parameters are, however, not the same; due to additional parameters for the nine parameter problem. As only one frequency is used in the optimization, more stability is probably obtained by increasing the number of frequencies.

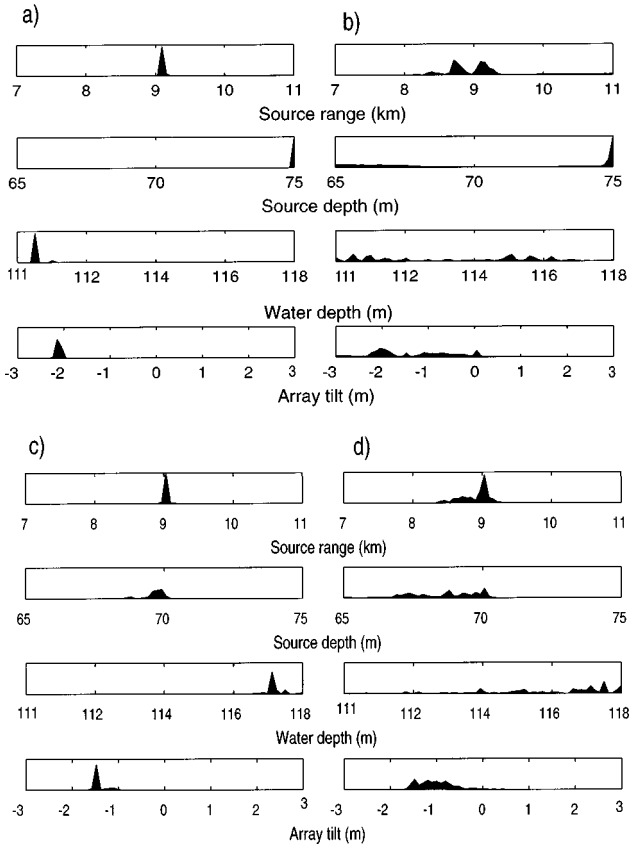

FIG. 9. The estimated *a posteriori* distribution for a search of five or nine parameters. (a) Five parameters, empirical posteriori; (b) five parameters, likelihood based posteriori; (c) nine parameters, empirical posteriori; (d) nine parameters, likelihood based posteriori. Only four parameters are shown in the plot as the fifth parameter was not that well determined.

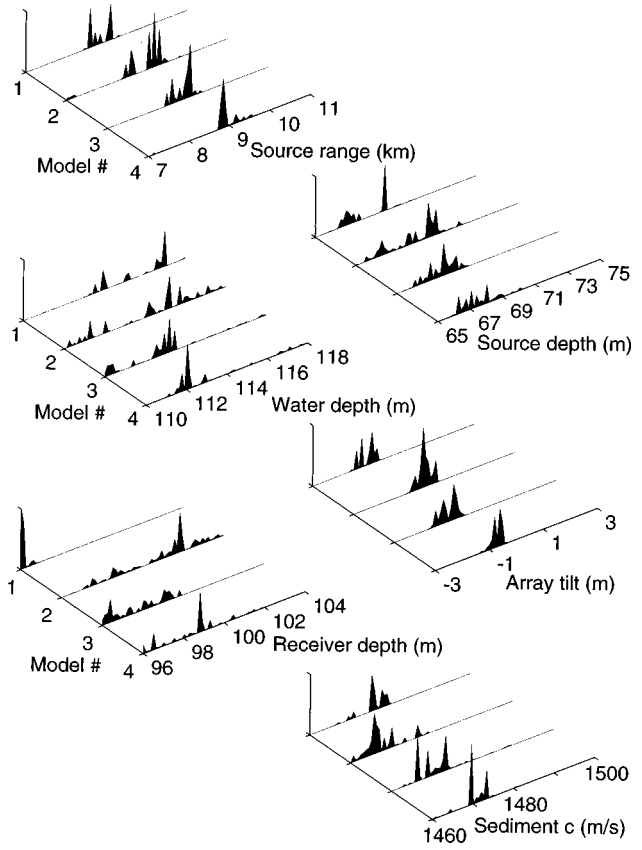

FIG. 10. The estimated *a posteriori* distribution for the YS-94 data. The noise has been estimated using five modes.

2. Multifrequency

By using observations at more frequencies, more information is used and thus a more robust estimation of the underlying parameters is usually obtained. In order to appreciate the value of more frequency observations it is essential to use the ML approach rather than the empirical approach (Sec. I B). The additional information that is gained from using more data is not reflected in the empirically based probability distributions, and thus it cannot be used to study convergence of solutions. The inversion is carried out using the four data models with increasing information:

(1) one frequency at 400 Hz;
(2) three frequencies at 200, 400, and 800 Hz;
(3) five frequencies at 400, 315, 400, 500, and 800 Hz; and
(4) seven frequencies at 400, 250, 315, 400, 500, 630, and 800 Hz.

The corresponding distributions are shown in Fig. 10. In general, as more frequencies are used, the solution seems to converge and the spread of the distributions decreases. One exception is the estimation of receiver depth for one frequency. However, it is clustered at one bound indicating that a solution outside the search bound is preferable. Incoherent averaging over frequency is especially effective if one or two octaves of signal bandwidth are available.

III. CONCLUSIONS

A precise formulation has been given for estimating the *a posteriori* distribution of environmental parameters retrieved from an ocean acoustic experiment. From these distributions, all information about the parameters can be extracted, as mean, higher moments, and marginal distributions. Numerical evaluation of multidimensional integrals over *a posteriori* distributions is required. These integrals can be numerically estimated using samples from global optimization methods. The method is based on importance sampling and requires no additional evaluation of the objective function at the expense of (negligible) bias.

The maximum-likelihood solution to an inversion problem does not provide an estimate of final parameter uncertainty. The likelihood based *a posteriori* distribution shows, however, the improvement in performance when using more frequencies, or differences when using a near or far array. As precise knowledge about the likelihood function is often unavailable for practical problems, the empirical formulation might be preferable, although uncertainty estimates are less accurate.

It is well known that SA and GA have superior performance for ocean geo-acoustic parameter estimations over local deterministic solution strategies due to the large number of secondary local optima of the objective function. Now, uncertainty studies can also be enhanced by a global approach.

The examples illustrate the use of this approach for simulated and real data on a vertical array. Array geometry is

arbitrary to the approach and both frequency and time domain data can be used.

APPENDIX A: OPTIMAL IMPORTANCE SAMPLING

Consider the evaluation of the multidimensional integral Eq. (6). First, the integral is rewritten as an expectation

$$\theta=\int_{\mathscr{M}}\frac{f(\mathbf{m})\sigma(\mathbf{m})}{g(\mathbf{m})}\,g(\mathbf{m})d\mathbf{m}\equiv\mathsf{E}_g\left[\frac{f(\mathbf{M})\sigma(\mathbf{M})}{g(\mathbf{M})}\right],\quad\text{(A1)}$$

where the random parameter vector $\mathbf{M}$ is selected from the generating distribution g. This expectation is estimated by the arithmetic mean $\hat{\theta}$ from Eq. (7). The variance is given by

$$\mathrm{Var}_g\,\hat{\theta}=\left[\mathsf{E}_g\left(\frac{f^2(\mathbf{M})\sigma^2(\mathbf{M})}{g^2(\mathbf{M})}\right)-\theta^2\right]\Big/\,N_{\mathrm{obs}}.\quad\text{(A2)}$$

Notice that the variance decreases as $O(N_{\mathrm{obs}}^{-1})$ with increasing number of samples N_{obs}. Using a variational procedure (with the constraint that g be a probability density) it can be shown that this variance is minimized for

$$g^{\mathrm{MV}}(\mathbf{m})=\frac{|f(\mathbf{m})|\sigma(\mathbf{m})}{\int_{\mathscr{M}}|f(\mathbf{m})|\sigma(\mathbf{m})d\mathbf{m}}.\quad\text{(A3)}$$

APPENDIX B: MATCHED FIELD LIKELIHOOD

Starting from Eq. (12) and using the Gaussianity of $\mathbf{e}(\omega_l)$ as stated in Sec. I C 1, the probability density (for a single time frame $K=1$) given the signal S_l and the noise power spectral density ν_l is given by

$$\mathscr{L}_1(\mathbf{m},S,\nu)=\prod_{l=1}^{L}(\pi\nu_l)^{-N}\exp\left[-\frac{|\mathbf{q}_l-\mathbf{w}_l(\mathbf{m})S_l|^2}{\nu_l}\right].\quad\text{(B1)}$$

Errors $\mathbf{e}_1$, $\mathbf{e}_2$ at differing frequencies $\omega_1\neq\omega_2$ are assumed uncorrelated. For large observation times it is a good approximation for the noise in the data (but might be violated for deterministic modeling errors, cf. Sec. I C). Measurement data $\mathbf{q}_{l,k}$ from multiple time frames $k=1,...,K$ is incorporated by multiplying the corresponding probability densities (B1) for each single time frame. This gives

$$\mathscr{L}_1=\prod_{k=1}^{K}\prod_{l=1}^{L}(\pi\nu_l)^{-N}\exp\left[-\frac{|\mathbf{q}_{l,k}-\mathbf{w}_l(\mathbf{m})S_{l,k}|^2}{\nu_l}\right].\quad\text{(B2)}$$

The ML estimate $\hat{\mathbf{m}}^{\mathrm{ML}}$ for $\mathbf{m}$ is obtained by jointly maximizing over the signal and noise parameters $(S_{l,k},\nu_l\forall l,k)$ and the model parameter vector $\mathbf{m}$. The maximization w.r.t. $S_{l,k}$ is obtained by requiring $\partial\mathscr{L}_1/\partial S_{l,k}=0$ in closed form: $\hat{S}_{l,k}=\mathbf{w}_l^{\dagger}(\mathbf{m})\mathbf{q}_{l,k}/|\mathbf{w}_l(\mathbf{m})|^2$. It is seen that $\hat{S}_{l,k}$ depends on $\mathbf{m}$ but not on ν. Inserting this into (B1) yields

$$\mathscr{L}_2(\mathbf{m},\nu)=\prod_{l=1}^{L}(\pi\nu_l)^{-NK}\exp\left[-\frac{\phi_l(\mathbf{m})}{\nu_l}\right]\quad\text{(B3)}$$

with

$$\phi_l=\mathrm{tr}\,\hat{\mathbf{R}}_l-\frac{\mathbf{w}_l^{\dagger}(\mathbf{m})\hat{\mathbf{R}}_l\mathbf{w}_l(\mathbf{m})}{\mathbf{w}_l^{\dagger}(\mathbf{m})\mathbf{w}_l(\mathbf{m})}.\quad\text{(B4)}$$

Optimizing w.r.t. ν_l yields

$$\hat{\nu}_l^{\mathrm{ML}}=\frac{1}{N}\,\phi_l.\quad\text{(B5)}$$

This gives

$$\mathscr{L}_3(\mathbf{m})=\left(\frac{N^K}{e\,\pi^K}\right)^{NL}\left(\frac{1}{\phi(\mathbf{m})}\right)^{N}.\quad\text{(B6)}$$

The ML solution $\hat{m}^{\mathrm{ML}}$ is obtained by maximizing $\mathscr{L}_3$ over all $\mathbf{m}\in\mathscr{M}$. Finally, an estimate for the noise power spectral density (which is assumed independent of $\mathbf{m}$) is obtained from Eq. (B5) and the ML solution $\hat{\nu}_l^{\mathrm{ML}}$ at $\hat{m}^{\mathrm{ML}}$ into the likelihood function, Eq. (B3). From now on, we consider the noise spectral density as known and only keep the free argument $\mathbf{m}$ of the objective function ϕ_l. This approach leads to

$$\mathscr{L}(\mathbf{m})=\prod_{l=1}^{L}(\pi\hat{\nu}_l^{\mathrm{ML}})^{-N}\exp\left[-\frac{\phi_l(\mathbf{m})}{\hat{\nu}_l^{\mathrm{MD}}}\right],\quad\text{(B7)}$$

which results in the definition of Eq. (18).

[1] N. Metropolis and S. Ulam, ''The Monte Carlo method,'' J. Am. Stat. Assoc. **44**, 335–341 (1948).
[2] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, ''Equation of states done by fast computing machines,'' J. Chem. Phys. **1**, 1087–1092 (1953).
[3] Y. Bard, *Nonlinear Parameter Estimation* (Academic, San Diego, 1974).
[4] W. Menke, *Geophysical Data Analysis: Discrete Inverse Theory* (Academic, San Diego, 1989).
[5] S. D. Rajan, J. F. Lynch, and G. V. Frisk, ''Perturbative inversion methods for obtaining bottom geoacoustic parameters in shallow water,'' J. Acoust. Soc. Am. **82**, 998–1017 (1987).
[6] A. Tarantola, *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation* (Elsevier, Amsterdam, 1987).
[7] P. W. Cary and C. H. Chapman, ''Automatic 1-D waveform inversion of marine seismic reflection data,'' Geophys. J. **93**, 527–546 (1988).
[8] M. K. Sen and P. L. Stoffa, ''Nonlinear one-dimensional seismic waveform inversion using simulated annealing,'' Geophysics **56**, 1624–1638 (1991).
[9] K. Mosegaard and A. Tarantola, ''Monte Carlo sampling of solutions to inverse problems,'' J. Geophys. Res. **100**, 12431–12447 (1995).
[10] M. K. Sen and P. L. Stoffa, ''Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion,'' Geophysical Prospecting **44**, 313–350 (1996).
[11] M. K. Sen and P. L. Stoffa, *Global Optimization in Geophysical Inversion* (Elsevier, Amsterdam, 1995).
[12] P. Gerstoft, ''Inversion of seismoacoustic data using genetic algorithms and *a posteriori* probability distributions,'' J. Acoust. Soc. Am. **95**, 770–782 (1994).
[13] P. Gerstoft and D. F. Gingras, ''Parameter estimation using multi-frequency range-dependent acoustic data in shallow water,'' J. Acoust. Soc. Am. **99**, 2839–2850 (1996).
[14] M. D. Collins, W. A. Kuperman, and H. Schmidt, ''Nonlinear inversion for ocean-bottom properties,'' J. Acoust. Soc. Am. **92**, 2770–2783 (1992).
[15] C. E. Lindsay and N. R. Chapman, ''Matched field inversion for geophysical parameters using adaptive simulated annealing,'' IEEE J. Ocean Eng. **18**, 224–231 (1993).
[16] 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 J. Ocean Eng. **18**, 232–239 (1993).
[17] P. Gerstoft and A. Caiti, ''Acoustic estimation of bottom parameters: error bounds by local and global methods,'' in *Second European Conference on Underwater Acoustics*, edited by L. Bjørnø (European Commission, Luxembourg, 1994), pp. 887–892.
[18] J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods* (Wiley, New York, 1964).
[19] A. J. W. Duijndam, ''Bayesian estimation in seismic inversion. Part I: Principles,'' Geophysical Prospecting **36**, 878–898 (1988).
[20] J. Lefebvre, H. Roussel, E. Walter, D. Lecointe, and W. Tabbara, ''Pre-

diction from wrong models: The Kriging approach,'' IEEE Antennas Propag. Mag. **38**, 35–45 (1996).
[21] F. B. Jensen and M. C. Ferla, ''SNAP-The SACLANTCEN normal mode acoustic propagation model,'' SACLANT Undersea Research Centre, SM-121, La Spezia, Italy, 1979.
[22] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1993).
[23] D. Maiwald and J. F. Böhme, ''Multiple testing for seismic data using bootstrap,'' IEEE Proc. IEEE ICASSP-94 **6**, 89–92 (1994).
[24] A. Tolstoy, *Matched Field Processing for Underwater Acoustics* (World Scientific, Singapore, 1993).
[25] M. L. Hinich, ''Maximum likelihood estimation of a radiating source in a waveguide,'' J. Acoust. Soc. Am. **66**, 480–483 (1977).
[26] E. C. Shang, ''Source depth estimation in wave guides,'' J. Acoust. Soc. Am. **86**, 1960–1964 (1985).
[27] D. J. Thompson, ''Jacknifing multiple-window spectra,'' IEEE Proc. ICASSP-94 **6**, 73–76 (1994).
[28] C. F. Mecklenbräuker, D. Maiwald, and J. F. Böhme, ''F-Test in matched field processing: identifying multimode propagation,'' IEEE Proc. ICASSP-95, Vol. 5, 3123–3126 (1995).
[29] S. Rajan, ''Waveform inversion for the geoacoustic parameters of the ocean bottom,'' J. Acoust. Soc. Am. **91**, 3228–3241 (1992).
[30] P. Gerstoft, ''Inversion of acoustic data using a combination of genetic algorithms and the Gauss-Newton approach,'' J. Acoust. Soc. Am. **97**, 2181–2191 (1995).
[31] P. Gerstoft, ''SAGA Users guide 2.0, an inversion software package,'' SACLANT Undersea Research Centre, SM-333, La Spezia, Italy, 1997.
[32] M. Porter and A. Tolstoy, ''The matched field processing benchmark problems,'' J. Comput. Acoust. **2**, 161–185 (1994).
[33] H. Schmidt and A. B. Baggeroer, ''Physics imposed resolution and robustness issues in seismo-acoustic parameter inversion,'' in *Full Field Inversion Methods in Ocean and Seismic Acoustics*, edited by O. Diachok, A. Caiti, P. Gerstoft, and H. Schmidt (Kluwer, Dordrecht, 1995), pp. 85–90.
[34] M. D. Collins and L. Fishman, ''Efficient navigation of parameter landscapes,'' J. Acoust. Soc. Am. **98**, 1637–1644 (1995).
[35] Z.-H. Michalopoulou, M. B. Porter, and J. Ianniello, ''Broadband source localization in the Gulf of Mexico,'' J. Comput. Acoust. **4**, 361–370 (1996).
[36] R. T. Kessel, ''The variation of modal wavenumbers with geoacoustic parameters in a layered media,'' J. Acoust. Soc. Am. **102**, 2690–2696 (1997).
[37] O. A. Godin, ''Acoustic mode reciprocity in fluid/solid systems: implications on environmental sensitivity and horizontal refraction,'' International Conference on Theoretical and Computational Acoustics, New Jersey, 1997.
[38] J-P. Hermand and P. Gerstoft, ''Inversion of broadband multitone acoustic data from the Yellow shark summer experiments,'' IEEE J. Ocean Eng. **21**, 324–346 (1996).
[39] C. F. Mecklenbräuker, P. Gerstoft, J. F. Böhme, and Pei-Jung Chung, ''Hypothesis testing for acoustic environmental models using likelihood ratio,'' J. Acoust. Soc. Am. (submitted).