Bayesian model selection applied to self-noise geoacoustic inversion

David J. Battle,[a] Peter Gerstoft, William S. Hodgkiss, and W. A. Kuperman
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238

Peter L. Nielsen
NATO Undersea Research Centre, 19138 La Spezia, Italy

(Received 15 March 2004; revised 25 June 2004; accepted 28 June 2004)

Self-noise geoacoustic inversion involves the estimation of bottom parameters such as sound speeds and densities by analyzing towed-array signals whose origin is the tow platform itself. As well as forming inputs to more detailed assessments of seabed geology, these parameters enable performance predictions for sonar systems operating in shallow-water environments. In this paper, Gibbs sampling is used to obtain joint and marginal posterior probability distributions for seabed parameters. The advantages of viewing parameter estimation problems from such a probabilistic perspective include better quantified uncertainties for inverted parameters as well as the ability to compute Bayesian evidence for a range of competing geoacoustic models in order to judge which model explains the data most efficiently.
[DOI: 10.1121/1.1785671]

PACS numbers: 43.30.Pc, 43.60.Pt [AIT] Pages: 2043–2056

I. INTRODUCTION

In self-noise geoacoustic inversion, plant and hydrodynamic noise generated by the tow ship as a by-product of its normal operation is used to interrogate the ocean environment. Because of its inherent mobility, reduced complexity, and low environmental impact, self-noise inversion using a towed array is a very promising modality of geoacoustic exploration.[1]

In our previous paper,[2] data acquired during the joint NATO/Marine Physical Laboratory experiment—MAPEX2000—was analyzed using maximum-likelihood (ML) methods to evaluate the feasibility of self-noise inversion from towed-array data. The major conclusion drawn from this preliminary work was that matched-field processing (MFP), in conjunction with global search procedures such as genetic algorithms (GA), was sufficiently sensitive in the near field to permit robust first-order inversion of parameters such as *p*-wave velocity for a range-independent bottom environment known to be reasonably well characterized as a fluid half-space. This was despite low to moderate signal-to-noise ratios (SNR) during the experiment and considerable uncertainty in relation to several important geometric parameters, such as water depth, source range, and array shape.

In this paper, we again direct our attention to the near-field inversion problem using towed-array data, applying a different paradigm to its solution and addressing two important questions: First, can we better quantify the sensitivity limits of near-field inversion? Second, is there a consistent way of ranking our success in modeling an unknown environment using multiple parametrizations? To address the first question, we focus on characterizing the posterior probability densities (PPDs) associated with ensembles of parameter samples generated by a procedure known as Gibbs sampling. Assuming that the model is correct, the PPD then summarizes our complete state of knowledge about the estimated parameters including their mean (expected) values, maximum *a posteriori* (MAP) values, and variances. In answer to the second question, we find as a consequence of knowing the PPD that Bayesian probability theory embodies a natural ranking for competing models, known as *evidence*. In Sec. II, we briefly reiterate a few salient aspects of Bayesian probability theory relevant to the current analysis. For completeness, we restate Bayes' rule and underline the interpretation of likelihoods, priors, and evidence in the context of inverse problems. In Sec. III, we describe in some detail our implementation of the Gibbs sampler used to obtain the results in later sections. While our development parallels that of Dosso[3–5] as recently applied to the analysis of synthetic and experimental geoacoustic data, we offer our own insights into the workings of the algorithm and some further suggestions for improving its efficiency. In Sec. IV we revisit the MAPEX2000 experiment and various aspects of the signal processing and modeling requirements of near-field inversion. Whereas in Ref. 2 we concentrated on demonstrating inversion consistency throughout an extensive portion of the dataset, in Sec. V we focus on the *methodology* of Bayesian model selection—analyzing a single 10-s frame and deriving PPDs corresponding to a variety of geoacoustic models with increasing levels of complexity. In addition to presenting marginal distributions for the inverted parameters, we evaluate the Bayesian evidence for each model using both analytic Gaussian approximations and a more accurate method known as reverse importance sampling. Our conclusions are given in Sec. VI.

---

[a] Now with the Ocean Engineering Dept., Massachusetts Institute of Technology, Cambridge, MA. Electronic mail: dbattle@mit.edu

II. BAYESIAN INFERENCE

The single aspect that most popularly distinguishes Bayesian probability and its application in inverse problems—Bayesian *inference*—from its so-called *orthodox* or *frequentist* alternatives is the use of prior (before-data) probabilities to modulate posterior (after-data) probabilities according to Bayes' rule

$$p(\mathbf{m}|\mathbf{d},\mathcal{M}_k)=\frac{p(\mathbf{d}|\mathbf{m},\mathcal{M}_k)p(\mathbf{m}|\mathcal{M}_k)}{p(\mathbf{d}|\mathcal{M}_k)}, \tag{1}$$

where $\mathbf{m}$ is a vector of model parameters $m_i\,(i=1,...,P)$, $\mathbf{d}$ is a vector of measured data $d_n\,(n=1,...,N)$, and the *models* $\mathcal{M}_k\,(k=1,...,M)$ embody various parametrizations considered plausible in explaining $\mathbf{d}$. $p(a|b)$ denotes conditional probability, meaning the probability of some outcome a given a previous outcome b. In words, Eq. (1) reads

$$\text{posterior}=\frac{\text{likelihood}\times\text{prior}}{\text{evidence}}.$$

Specializing to the geoacoustic problem at hand, we regard the m_i as representing acoustic quantities that we wish to infer from data $\mathbf{d}_f$ acquired at discrete frequencies $\omega_f\,(f=1,...,F)$ from an array of N hydrophones. The likelihood $p(\mathbf{d}|\mathbf{m},\mathcal{M}_k)$ quantifies the error in matching the $\mathbf{d}_f$ with *replica* fields generated by a continuous wave (cw) propagation model, while the prior $p(\mathbf{m}|\mathcal{M}_k)$ defines and weights plausible ranges for the parameters m_i. The evidence appearing in the denominator can, at first, be thought of as an overall normalization for the PPD, which is independent of the parameter vector $\mathbf{m}$.

A. First level of inference

As defined by MacKay,[6] the first of the two levels of Bayesian inference is concerned with parameter estimation problems in which each proposed model is assumed to be correct, i.e., all conceivable possibilities are encompassed within its prior parameter space. This assumption is responsible for the aforementioned normalization

$$\int p(\mathbf{m}|\mathbf{d},\mathcal{M}_k)d\mathbf{m}=1. \tag{2}$$

Aside from this technicality, it is the *shape* of the PPD, and hence only the likelihood and prior terms, that influence the first level of inference.

It is often a criticism of Bayesian probability that, although it offers clear guidance on how to use prior probabilities through Eq. (1), it is neutral as to how they are derived in the first place. As is the case with inference in general, priors have an inescapable subjective element, which highlights the fact that all inference is subjective to a point. In this work, flat priors have been assumed for all parameters, with lower bounds l_i and upper bounds u_i, such that the normalized prior density for each parameter m_i is given by

$$p(m_i|\mathcal{M}_k)=\frac{1}{(u_i-l_i)}=\frac{1}{w_i}. \tag{3}$$

In such cases, provided the bulk of the posterior probability is bounded within the assumed w_i, the prior is said to be *diffuse* and does not significantly influence the shape of the PPD. Even diffuse priors, however, can impact the *relative* probabilities of proposed models, thus influencing the second level of inference, as described next.

B. Second level of inference

Whereas at the first level of inference only the likelihood and (usually to a lesser extent) the prior are important, the second level involves the denominator, or *evidence* term, which is the normalizing constant for the PPD, *viz.*

$$p(\mathbf{d}|\mathcal{M}_k)=\int p(\mathbf{d}|\mathbf{m},\mathcal{M}_k)p(\mathbf{m}|\mathcal{M}_k)d\mathbf{m}. \tag{4}$$

At the second level of inference, the intention is to rank each model $\mathcal{M}_k$ in terms of its ability to explain the data. As is well known, such a ranking cannot be based solely on the likelihood, as arbitrarily complex models can match the data arbitrarily closely. Paralleling Eq. (1), a posterior probability can also be attributed to each $\mathcal{M}_k$, independent of its parameters. However, in this case, normalization is not meaningful, leaving the proportionality

$$p(\mathcal{M}_k|\mathbf{d})\propto p(\mathbf{d}|\mathcal{M}_k)p(\mathcal{M}_k). \tag{5}$$

The reason that Eq. (5) cannot be normalized is simply that the universe of models $\mathcal{M}_k$ is subjective and infinite. Usually, the best that can be done is to enumerate M models for comparison, and—assuming no initial preference—the resulting prior probabilities are given by

$$p(\mathcal{M}_k)=1/M. \tag{6}$$

It then follows, for the purposes of model comparison, that the posterior probability of each model is simply proportional to its evidence, which for reasons that will become apparent later in this section is sometimes referred to as the *marginal* or integrated likelihood. According to Bayesian theory, the model associated with the highest numerical evidence as given by Eq. (4) is to be preferred.

What may not be immediately apparent from the preceding discussion is how Bayesian evidence automatically discriminates against models that are overly complex, thereby expressing a preference for *parsimonious* parametrizations.[7] This preference is also known as Occam's razor[6] and becomes more visible after assuming a *hypothetical* Gaussian form for the parameter likelihood around the maximum likelihood (ML) point $\hat{\mathbf{m}}$ [with an ML value $p(\mathbf{d}|\hat{\mathbf{m}},\mathcal{M}_k)$ and a parameter covariance matrix $\mathbf{C}_m$]

$$p(\mathbf{d}|\mathbf{m},\mathcal{M}_k)=p(\mathbf{d}|\hat{\mathbf{m}},\mathcal{M}_k)\exp\left(-\frac{\Delta\mathbf{m}^T\mathbf{C}_m^{-1}\Delta\mathbf{m}}{2}\right). \tag{7}$$

In Secs. V B and V C, it will be shown that Eq. (7) is often *not* a good approximation to real likelihood functions, which require numerical integration (Sec. III F). However, such approximations are common in probability theory,[8,6] principally because the integral or *hypervolume* of a Gaussian likelihood function in P dimensions has the analytic form

$$\int p(\mathbf{d}|\mathbf{m},\mathcal{M}_k)d\mathbf{m}=p(\mathbf{d}|\hat{\mathbf{m}},\mathcal{M}_k)(2\pi)^{P/2}\sqrt{\det\mathbf{C}_m}. \tag{8}$$

Making the further simplifying assumptions that the prior is flat and that the integral in Eq. (4) is not strongly affected by including the tails of $p(\mathbf{d}|\mathbf{m},\mathcal{M}_k)$ which lie outside the prior ranges u_i and l_i, Eq. (4) can be rewritten

$$\begin{aligned} p(\mathbf{d}|\mathcal{M}_k) &= p(\mathbf{d}|\hat{\mathbf{m}},\mathcal{M}_k)(2\pi)^{P/2}\sqrt{\det \mathbf{C}_m}\, p(\mathbf{m}|\mathcal{M}_k) \\ &= \underbrace{p(\mathbf{d}|\hat{\mathbf{m}},\mathcal{M}_k)}_{\text{likelihood}}\ \times\ \underbrace{(2\pi)^{P/2}\sqrt{\det \mathbf{C}_m}\prod_{i=1}^{P} w_i}_{\text{Occam factor}}, \\ \text{evidence} &= \text{likelihood} \times \text{Occam factor} \end{aligned} \tag{9}$$

from which it can be seen that the maximum likelihood achieved by a particular model does, as expected, weigh positively toward its ranking. Weighing against the likelihood, however, is the so-called *Occam factor*,[6,7,9] which can be interpreted as an integral of the prior probability density weighted according to the distribution of the likelihood. In other words, the Occam factor is a measure of the concentration of prior probability within the high-likelihood region of a parameter space, and it displays the following general properties:[6–9]

(i) The Occam factor penalizes models that incorporate large numbers of free parameters through the growth of the prior hypervolume $\Pi_{i=1}^{P} w_i$.

(ii) For the same reason, the Occam factor penalizes models with wide *a priori* parameter bounds w_i.

(iii) The Occam factor penalizes models that have to be finely tuned to fit the data, as these have concentrated PPDs and correspondingly small posterior volumes.

In geophysical theory, the first to describe Bayesian evidence in model selection was Jeffreys in 1939.[10] Only very recently, however, new applications have been reported.[7,9,11] The reasons for this no doubt stem from the complexity of real-world problems and the difficulties in deriving and manipulating the necessary probability distributions. Although Eq. (4) is a simple prescription, such integrations require specialized numerical methods that have only recently become practical.

Whereas alternative concepts such as minimum descriptive lengths (MDL), Akaike information criteria (AIC), and likelihood ratio tests have also been applied to model selection,[12] proponents of Bayesian inference point to the fact that the concept of evidence flows naturally and consistently from the basic *desiderata* of probability theory. In fact, most other methods can be viewed as being closely related to, or approximations of the full Bayesian framework described here.[6] Recent analyses have indicated that Bayesian model selection, which has higher computational demands than other approximate methods, is capable of superior performance, particularly in cases of low SNR and/or small amounts of data.[13]

C. Marginal inference

As discussed above, the calculation of (usually nonanalytic) PPDs is central to numerical applications of the first and second levels of Bayesian inference. When summarizing *a posteriori* knowledge of parameter values, however, it is desirable to treat each m_i (or perhaps pairs of m_i in the 2D marginal probability distribution case) separately, while integrating over the range of influence of the other parameters $m_{i\neq j}$, giving *marginal* probability distributions of the form

$$p(m_j|\mathbf{d},\mathcal{M}_k) = \int p(m_j, m_{i\neq j}|\mathbf{d},\mathcal{M}_k)\, dm_{i\neq j}. \tag{10}$$

Parameter estimation based on the properties of marginal distributions is known as marginal inference, and will be applied extensively in Sec. V, in which both one- and two-dimensional marginal distributions are computed from samples drawn from the joint parameter PPD.

III. GIBBS SAMPLING

The basic inputs to Bayesian analysis are probability distributions, and in practice, these can be difficult to estimate given the dimensionality of real-world problems. Gibbs sampling is an iterative Markov chain Monte Carlo (MCMC) procedure designed to sample from joint posterior distributions using only samples from conditional distributions. In the case of large-scale problems with many parameters possessing unknown correlations, joint distributions are usually not available, whereas conditional distributions often are.

The origins of the Gibbs sampler date back to Hastings[14] in statistical analysis and subsequently Geman and Geman,[15] who applied the idea to large-scale image reconstruction problems. Comprehensive accounts of the theory and application of Gibbs sampling to inverse problems—particularly those in geophysics—as well as references to the original literature include Gelfand and Smith,[16] Smith and Roberts,[17] Sen and Stoffa,[18] and Mosegaard and Sambridge.[11] In relation to ocean-acoustic problems of the type of interest here, Dosso[3–5] recently analyzed synthetic and experimental data and concluded that Gibbs sampling is a powerful and robust means of estimating geoacoustic parameters and their associated errors.

In this section, we briefly discuss issues relevant to the integration of high-dimensional PPDs in geoacoustic inversion. These include the Metropolis–Hastings approach to sample generation, definition of a matched-field energy function, Gibbs sampler initialization, coordinate rotation, convergence criteria, and finally, one approach to the numerical estimation of Bayesian evidence.

A. Sample generation

In essence, Gibbs sampling consists of generating successive samples from P conditional distributions, such that at the completion of the qth iteration, the parameter vector $(m_1^q, m_2^q, m_3^q, \ldots, m_{P-1}^q, m_P^q)$ can be considered to have been drawn from the joint PPD. For clarity, it will be implicitly assumed that posterior probabilities are conditional on the data $\mathbf{d}$ and model $\mathcal{M}_k$ from this point. Starting with the initial parameter vector $(m_1^0, m_2^0, m_3^0, \ldots, m_{P-1}^0, m_P^0)$ and denoting the conditional distribution of parameter m_1 with respect to parameters $m_{2\rightarrow P}$ in the first iteration as $p(m_1|m_2^0, m_3^0, \ldots, m_{P-1}^0, m_P^0)$, the sampling proceeds with each successive conditional distribution immediately incorporating the previously selected parameter value. Using the "←" symbol to denote the drawing of a sample, the first iteration can be illustrated as[8]

$$\begin{aligned}
&m_1^1 \leftarrow p(m_1|m_2^0,m_3^0,\ldots,m_{P-1}^0,m_P^0)\\
&m_2^1 \leftarrow p(m_2|m_3^0,m_4^0,\ldots,m_P^0,m_1^1)\\
&m_3^1 \leftarrow p(m_3|m_4^0,m_5^0,\ldots,m_1^1,m_2^1)\\
&\vdots \quad \vdots \quad \vdots\\
&m_P^1 \leftarrow p(m_P|m_1^1,m_2^1,\ldots,m_{P-2}^1,m_{P-1}^1),
\end{aligned}$$

with the qth iteration given by

$$\begin{aligned}
&m_1^q \leftarrow p(m_1|m_2^{q-1},m_3^{q-1},\ldots,m_{P-1}^{q-1},m_P^{q-1})\\
&m_2^q \leftarrow p(m_2|m_3^{q-1},m_4^{q-1},\ldots,m_P^{q-1},m_1^q)\\
&m_3^q \leftarrow p(m_3|m_4^{q-1},m_5^{q-1},\ldots,m_1^q,m_2^q)\\
&\vdots \quad \vdots \quad \vdots\\
&m_P^q \leftarrow p(m_P|m_1^q,m_2^q,\ldots,m_{P-2}^q,m_{P-1}^q).
\end{aligned}$$

In the Metropolis–Hastings variant of the Gibbs sampler used here, samples are drawn from each conditional distribution according to the Metropolis rule, wherein uniformly distributed perturbations, resulting in modified parameter vectors $\mathbf{m}'$, are accepted with probability

$$p_{\text{accept}} = \begin{cases} \dfrac{p(\mathbf{m}')}{p(\mathbf{m})} & \text{for } p(\mathbf{m}') \leqslant p(\mathbf{m}), \\ 1 & \text{for } p(\mathbf{m}') > p(\mathbf{m}). \end{cases} \tag{11}$$

As in Eq. (1), the $p(\mathbf{m})$ are posterior probabilities which, given simplifying Gaussian assumptions regarding the noise and modeling errors, take the exponential form

$$p(\mathbf{m}) = \exp[-E(\mathbf{m})], \tag{12}$$

where $E(\mathbf{m})$ is an *energy function* to be discussed shortly. The above combination of sequential perturbation, acceptance, and rejection establishes a Markov chain whose sampling density can be shown to converge to the Gibbs–Boltzmann distribution from thermodynamics[11,18]

$$p(\mathbf{m}) = \frac{1}{Z}\exp\left[-\frac{E(\mathbf{m})}{T}\right], \tag{13}$$

with the normalization or *partition* function Z ignored, and a particular choice of unity for the temperature T. This is the same equilibrium distribution associated with simulated annealing (SA) algorithms, and hence Gibbs sampling has the interpretation of being SA conducted at a constant temperature of $T=1$.[19,20]

When evaluating the multidimensional integrals [Eq. (10)] with importance sampling, the variance of the integral estimates is reduced if the generating distribution is proportional to the integrand, and this is actually obtained with a Gibbs sampler.[3,8,18] Generally, when evaluating the multidimensional integrals [Eq. (10)] with importance sampling, the sampling distribution should be concentrated in regions of the parameter space where the PPD is most significant. Ideally, if the PPD is perfectly mirrored by the sampling distribution, then the error in estimating the PPD and its associated marginal integrals tends to zero—a situation never achieved in practice, but approximated with a given amount of computation such that Gibbs sampling is usually far more attractive than either less adaptive or more exhaustive sampling techniques.[3,8,18] While other methods of importance sampling have also been applied to PPD estimation, including, for example, GA in ocean acoustics,[21–26] the great advantage of Gibbs sampling is that the sampling distribution is *known* to be given by Eq. (13). In the case of GA and related methods, the sampling distribution is usually not known and consequently the results may be difficult to interpret.[4]

B. Energy function

In Ref. 2, parameter estimation was carried out by directly maximizing normalized Bartlett power objective functions of the form

$$B_f(\mathbf{m}) = \left[\frac{\mathbf{w}_f^\dagger(\mathbf{m})\hat{\mathbf{R}}_f\mathbf{w}_f(\mathbf{m})}{\mathrm{tr}[\hat{\mathbf{R}}_f]\|\mathbf{w}_f(\mathbf{m})\|^2}\right], \tag{14}$$

where the $\mathbf{w}_f$ were replica vectors calculated from a continuous wave acoustic propagation model and the $\hat{\mathbf{R}}_f$ $(f=1,\ldots,F)$ were cross-spectral density matrices (CSDMs) averaged from N_s snapshot vectors $\mathbf{d}_{n,f}$ of the array data at frequencies ω_f according to

$$\hat{\mathbf{R}}_f = \frac{1}{N_s}\sum_{n=1}^{N_s} \mathbf{d}_{n,f}\mathbf{d}_{n,f}^\dagger. \tag{15}$$

To formulate the energy function $E(\mathbf{m})$ required by Eq. (12), Eq. (14) can be transformed into the *mismatch* or error function[27]

$$\phi_f(\mathbf{m}) = 1 - B_f(\mathbf{m}), \tag{16}$$

following which[3,25]

$$E(\mathbf{m}) = \sum_{f=1}^{F} \frac{\phi_f(\mathbf{m})|\mathbf{d}_f|^2}{\hat{\sigma}_f^2}, \tag{17}$$

where $|\mathbf{d}_f|^2$ is the total power seen by the array in frequency band f and $\hat{\sigma}_f^2$ is a corresponding estimate of the variance associated with the combination of model mismatch, ocean, and instrumental noise.

To estimate $\hat{\sigma}_f^2$, two simplifying assumptions have been made here. First, considering the single frame of data to be discussed in Sec. IV, and the three frequencies selected for inversion, it is reasonable to take the SNR as approximately constant across frequency. Second, the spatial distribution of noise along the array has been assumed constant despite the short range of the experiment which, in reality, would have resulted in higher SNRs at the end of the array nearer the source. Following the maximum-likelihood arguments of Gerstoft and Mecklenbräuker in relation to the noise variance expected after averaging N_s snapshots, as in Eq. (15), the average variance can be estimated as[3,25,27]

$$\hat{\sigma}_{\text{ave}}^2 = \frac{\phi_{\text{ave}}(\hat{\mathbf{m}})|\mathbf{d}_{\text{ave}}|^2}{N_e}, \tag{18}$$

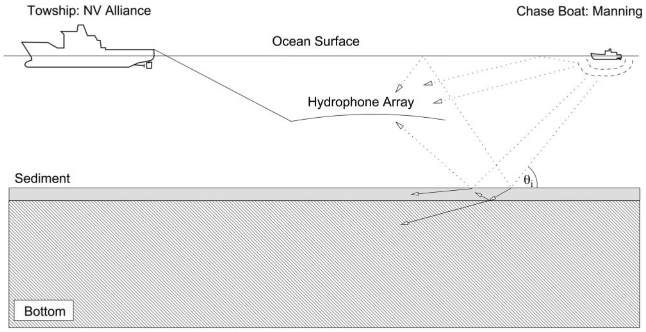

FIG. 1. The configuration of the ALLIANCE and the MANNING during the MAPEX2000 self-noise experiment.

where $\phi_{\mathrm{ave}}(\hat{\mathbf{m}})$ is the average normalized Bartlett mismatch obtained using a global search (at the first level of inference), $|\mathbf{d}_{\mathrm{ave}}|^2$ is the average measured power, and N_e is an estimate of the number of degrees of freedom associated with the noise.

As discussed by Gerstoft and Mecklenbräuker[25,27] and Dosso,[3,4] the selection of appropriate values for N_e (and hence $\hat{\sigma}^2_{\mathrm{ave}}$) in geoacoustic inversion is problematic due to the fact that the structure of the true model and hence also that of the true signal are generally unknowable. Because of this, *modeling* errors actually tend to dominate even at moderate SNR values—in most cases swamping the impact of noise completely. Faced with this difficulty, we adopted a heuristic approach and chose $N_e=5$ (Refs. 28, 25, 3, 4) noting that: (1) This was approximately the number (six) of independent snapshots used, and (2) from experience, acceptance ratios (proportions of accepted Metropolis moves) on the order of 25% to 50% are typically indicative of Gibbs sampler algorithms performing correctly.[11,13] Too large an estimate of N_e leads to artificially narrow PPDs *relative to the prior bounds* through the exponentiation in Eq. (12), and correspondingly poor acceptance ratios.

To summarize, the final energy function used in our Gibbs sampler had the form

$$E(\mathbf{m})=S\sum_{f=1}^{F}\phi_f(\mathbf{m}), \tag{19}$$

with $S=N_e/\phi_{\mathrm{ave}}(\hat{\mathbf{m}})$. Given that the highest average Bartlett power (correlation coefficient squared) obtained here was approximately 0.8, the overall energy scaling factor S was

$$S=\frac{N_e}{\phi_{\mathrm{ave}}(\hat{\mathbf{m}})}\approx\frac{5}{0.2}=25. \tag{20}$$

C. Initialization

As Gibbs samplers essentially sample around the modes of PPDs, they must be correctly initialized to regions of high posterior probability. In many reported applications, an initialization or *burn-in* phase has preceded to the equilibrium $(T=1)$ sampling phase such that an initial quenching from some high temperature takes place.[8,3,4] In this work, we found that SA initialization did perform satisfactorily provided the starting temperature and annealing schedule were correctly chosen—usually by trial and error. However, substituting a genetic algorithm (GA) in place of SA ultimately proved faster in locating the main concentrations of probability. This was probably due to the insensitivity of GA to parameter correlations that can seriously affect the acceptance ratios obtained during conventional SA.

D. Coordinate rotation

As noted by Collins and Fishman,[29] algorithms that generate univariate parameter perturbations, such as SA and Gibbs sampling, are particularly susceptible to poor acceptance ratios when sampling strongly correlated parameter spaces. In ocean acoustics, such parameter correlations are frequently encountered,[30] and coordinate rotation, whereby the parameter covariance matrix **C**, or alternatively, a covariance matrix computed from field derivatives[29] is effectively orthogonalized, is a standard solution that will not be discussed here in detail. Following Dosso,[3] our Gibbs sampler

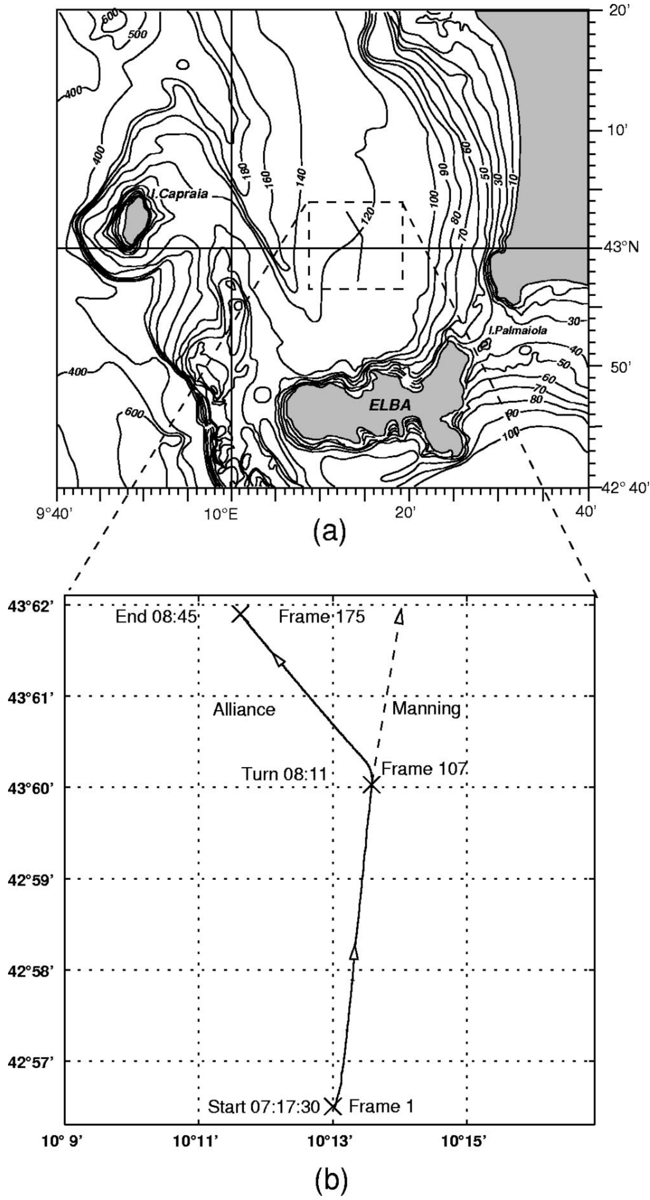

FIG. 2. The tracks of the NRV ALLIANCE and MANNING during the MAPEX2000 self-noise experiment. All times are UTC. Each frame represents a 30-s increment.

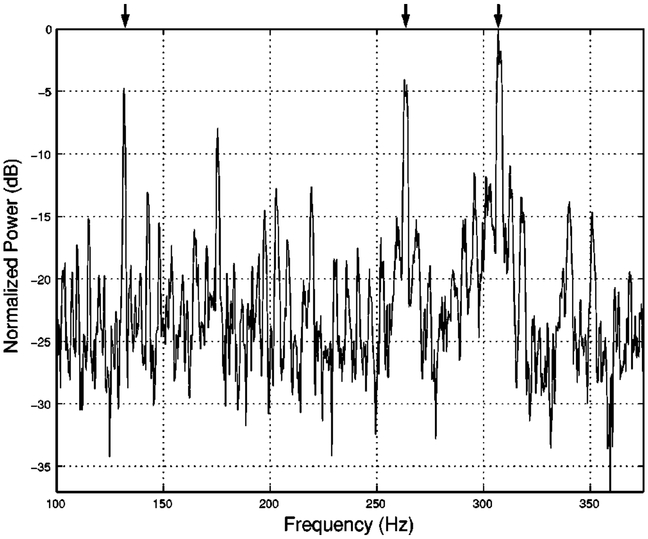

FIG. 3. Normalized power spectral density computed from time-series data beamformed in the approximate direct-path direction of the support ship MANNING. The three frequencies used for inversion were 131.47, 262.94, and 306.88 Hz, as indicated by the top markers.

originally included a covariance estimation phase conducted at $T=1$ following which the parameter covariance was estimated from sampled vectors $\mathbf{m}$ as

$$\mathbf{C}_m \approx \langle \mathbf{m}\mathbf{m}^\dagger \rangle - \langle \mathbf{m} \rangle \langle \mathbf{m} \rangle^\dagger. \qquad (21)$$

Collins and Fishman[29] originally suggested using the covariance matrix of the derivatives of the field. Instead of sampling the covariance matrix $\mathbf{C}_m$, the Cramer–Rao lower bound matrix $\mathbf{C}_{\mathrm{CRLB}}$ which contains information about parameter coupling, can be used to estimate the required parameter rotations. In this work the Cramer–Rao formulation due to Baggeroer *et al.* extended to the multifrequency case[31] is used

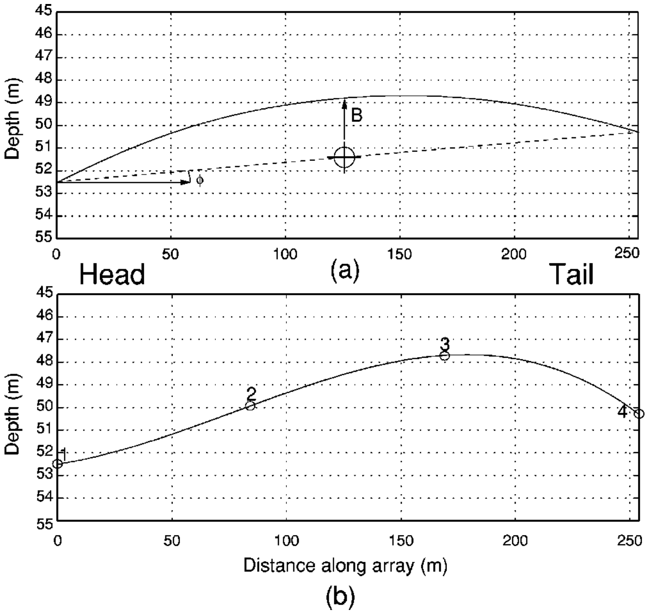

FIG. 4. Towed-array shape models. (a) Simple parabolic model with depth, bow and tilt. (b) Cubic spline model fitting a variable number of depth points—in this case z_{a1}, z_{a2}, z_{a3}, and z_{a4}.

$$[\mathbf{C}_{\mathrm{CRLB}}^{-1}]_{ij} = \sum_{f=1}^{F} \mathrm{Tr}\left[\mathbf{K}_f^{-1} \frac{\partial \mathbf{K}_f}{\partial m_j} \mathbf{K}_f^{-1} \frac{\partial \mathbf{K}_f}{\partial m_i} \right], \qquad (22)$$

where the $\mathbf{K}_f$ are outer products of replica fields $\mathbf{w}_f$ calculated at the ML point $\hat{\mathbf{m}}$ and loaded uniformly on their diagonals to approximate the estimated SNR of the data. That Eq. (22) can be used in this way *independently of the data* is less surprising when it is considered that coordinate rotation is simply another kind of importance sampling designed to concentrate samples in regions of high posterior probability. While it is conceivable that a proposed ocean model might be sufficiently far from reality for Eq. (22) to fail as the basis of such a sampling distribution, in all cases reported here we observed substantial improvements in Gibbs sampler acceptance ratios when $\mathbf{C}$ was estimated semianalytically using Eq. (22) (up to 50%) as opposed to Eq. (21) (≈25%).

E. Convergence criteria

As we were mainly interested in *marginal* PPDs, we based our Gibbs sampler convergence criteria on the maximum fractional change undergone by any 50-bin parameter histogram in 100 cycles through the Gibbs sampler. While not as rigorous as the dual-population convergence criteria used by others,[3] we checked for convergence every 100 steps and found that histogram changes below 10% were a reliable sign of convergence. All results presented in Sec. IV achieved this criterion.

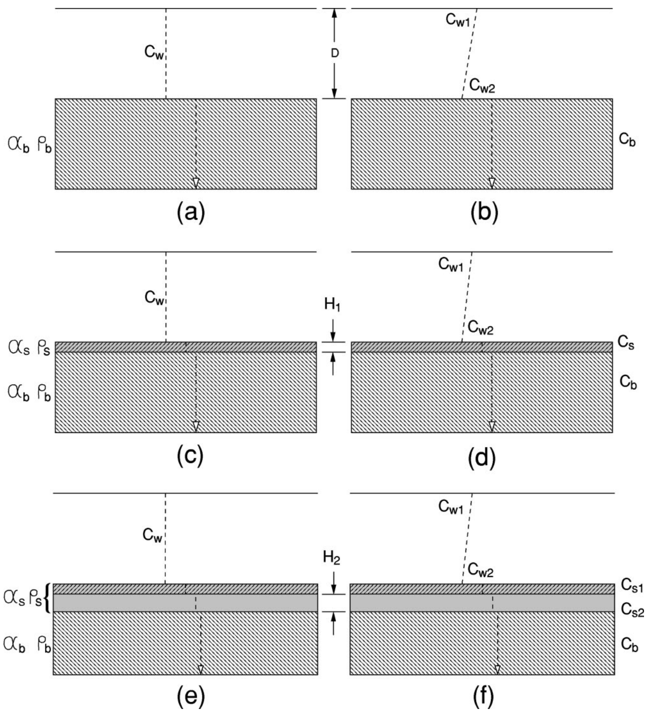

FIG. 5. Six acoustic waveguide models with varying degrees of complexity and their associated parameters. (a) Iso-velocity water over half-space. (b) Sound-speed gradient over half-space. (c) Iso-velocity water over sediment and half-space. (d) Sound-speed gradient over sediment and half-space. (e) Iso-velocity water over two-layer sediment and half-space. (f) Sound-speed gradient over two-layer sediment and half-space. Note the distinct surface and bottom water column velocities C_{w1} and C_{w2} in models (b), (d), and (f). Also note that while the two sediment layer velocities C_{s1} and C_{s1} are distinct in models (e) and (f), the densities and attenuations of these layers (ρ_s and α_s) have been assumed equal for simplicity.

TABLE I. Parameters and inversion bounds for the first case involving a parabolic array and Pekeris environment.

Parameter	Symbol	Unit	Min	Max
Water sound speed	C_w	m/s	1510	1530
Water depth	D	m	115	125
Bottom sound speed	C_b	m/s	1450	1650
Bottom density	ρ_b	g/cm^3	1	3
Bottom attenuation	α_b	dB/λ	0	2
Source depth	Z_s	m	0	5
Source range	R_s	m	275	325
Array depth	Z_a	m	45	55
Array tilt	ϕ	degrees	−2	2
Array bow	B	m	−5	5

F. Reverse importance sampling

In Sec. II B, Bayesian evidence was defined in terms of an integral that, given a Gaussian approximation to the parameter likelihood and flat priors, had a simple analytic form. Unfortunately, as will be apparent in Sec. V, such approximations can result in errors of many orders of magnitude and sometimes compromise the outcome of Bayesian model selection. More accurate calculation of evidence calls for a numerical integration approach.

Ó Ruanaidh and Fitzgerald[8] describe such an approach that is closely related to the concept of importance sampling, but works in reverse to extract the probabilistic evidence of a model from samples $\mathbf{m}_q$ already available from its joint PPD. In the present case, these samples will have been generated at the first level of inference, i.e., in estimating the PPD of the parameters via Gibbs sampling. The task then is to invoke a Gaussian approximation, not for the integral itself, but to be *reverse sampled* by the existing ensemble, thereby calibrating the unknown evidence integral against that of the Gaussian, which is known. Taking $g(\mathbf{m})$ as a normalized Gaussian

$$\int g(\mathbf{m})d\mathbf{m}=1, \tag{23}$$

and $\mathcal{E}$ as the unknown evidence integral

$$\mathcal{E}=\int f(\mathbf{m})d\mathbf{m}, \tag{24}$$

$\mathcal{E}$ can be shown to be approximated by[8]

$$\frac{1}{\mathcal{E}}\approx\frac{1}{Q}\sum_{q=1}^{Q}\frac{g(\mathbf{m}_q)}{f(\mathbf{m}_q)}. \tag{25}$$

Calculating Bayesian evidence by reverse importance sampling (RIS) is straightforward, requiring negligible further computation once the PPD samples $\mathbf{m}_q$ and their functional values $f(\mathbf{m}_q)$ have been generated by Gibbs sampling.

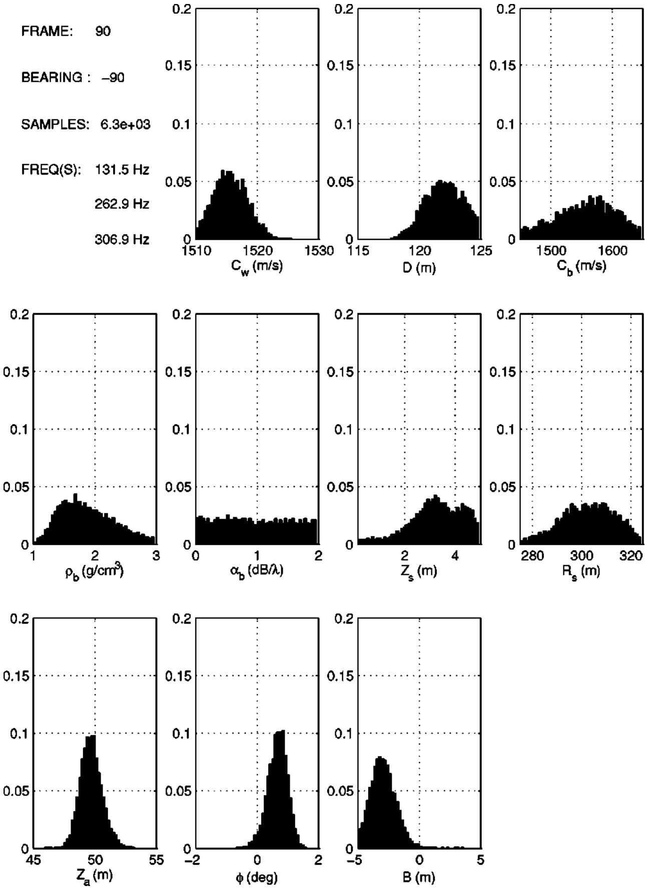

FIG. 6. Marginal PPD histograms for frame 90. Environment model (a), parabolic array and MANNING source.

IV. SELF-NOISE INVERSION

A. MAPEX2000

The 90-min experiment discussed here was conducted by the NATO Undersea Research Center and the Marine Physical Laboratory as part of MAPEX2000, which was specifically directed at validating a range of array processing and geoacoustic inversion techniques. The data were acquired north of the island of Elba, off the Italian west coast on the 29th of November, 2000 between the times of 07:17:30 and 08:45:00 UTC.[2] The array used consisted of 128 hydrophones evenly spaced at 2 m. Half-wavelength sampling therefore occurred at approximately 375 Hz. Depth control with this array proved less accurate than hoped, although this turned out to be of little consequence, as the array shape was included in the vector $\mathbf{m}$ along with the other parameters to be optimized.

The self-noise inversion dataset comprised 175 frames sampled at 30-s intervals. Only the first 10 s of each interval, or 60 000 samples at a sampling rate of 6000 samples per second, were recorded. One novel aspect of this self-noise experiment was that two research vessels were involved. From about frame 60 to frame 107, the NRV ALLIANCE, with its array towed approximately 330 m behind, was followed by a smaller vessel, the MANNING, at a range of approximately 900 m. The horizontal distance of the MANNING from the tail of the array was therefore approximately 300 m.

As in our earlier analysis,[2] we have made the assumption here that the source ship was well approximated as a point source over the range of frequencies considered. While realizing that this may not generally be the case with larger vessels or at higher frequencies, the point-source assumption seems to have been borne out at least in the case of the MANNING by consistently sharp matched-field processor (MFP) peaks.

TABLE II. Gibbs sampler inversion results for the first case involving a parabolic array and Pekeris environment.

Parameter	Unit	MAP	Mean	Std. dev.
C_w	m/s	1515	1516	3
D	m	121.5	122.1	1.5
C_b	m/s	1620	1560	47
ρ_b	g/cm^3	2.0	1.9	0.4
α_b	dB/λ	2.0	1.0	0.6
Z_s	m	4.5	3.3	1.0
R_s	m	295	303	7
Z_a	m	50.2	49.9	0.9
ϕ	degrees	0.4	0.7	0.3
B	m	−2.3	−2.8	1.0

At frame 107, ALLIANCE started a 45° turn, while MANNING continued on track. Then, at frame 122, MANNING left station and departed the area at increased speed without turning. In this analysis, we consider only the 10 s of data acquired in frame 90, during which the MANNING was in station approximately 300 m behind the Alliance, as illustrated in Fig. 1. We do not consider the ALLIANCE, which posed much lower a source level for the purposes of reliable self-noise inversion.[2]

Figure 2 shows the track of the two ships during the experiment, and gives an indication of the local bathymetry. Over the entire track, the depth sounder recorded a steady increase in depth from approximately 116 to 124 m. Expendable bathy-thermograph casts showed an almost iso-velocity (winter) water column with a velocity of approximately 1520 m/s. The speed of the ALLIANCE was a steady 4 knots and the prevailing sea state was 1.

In the area of the experiment, the seafloor has been fairly well characterized over the last 30 years,[23,32,26,33] and is known to be flat and covered with a thin layer of clay and sand–clay sediments. Below the thin sediment layer is known to exist a reasonable approximation to an acoustic half-space with the approximate parameters $c_b \approx 1600$ m/s, $\rho_b \approx 1.8$ g/cm^3, and $\alpha_b \approx 0.15$ dB/λ.[26] At finer levels of detail, Holland and Osler[34] have noted the existence of some randomly placed thin, high-speed layers within sediment core samples.

In no results to date has there been evidence of significant shear wave propagation in the seafloor in the North Elba area, and hence only compressional wave parameters were considered here. In any case, the assumption of a shear-wave velocity c_s within the plausible range for the bottom type in question ($c_s \ll c_w$) would lead mainly to an increased bottom loss,[35] to which the current near-field inversion technique would be relatively insensitive.

B. Preprocessing

From each 60 000-point frame, 6 Kaiser–Bessel windowed, 16 384-point snapshots were averaged with 50% overlap in order to estimate both the power-spectral densities (PSDs) and cross-spectral density matrices (CSDMs). Elementary frequency tracking was used to follow small variations in the tonal frequencies emitted by the MANNING. The three center frequencies selected for inversions were 131.47, 262.94, and 306.88 Hz, as indicated by the markers at the top of the normalized PSD in Fig. 3, which was computed from time-series data beamformed in the approximate direct-path direction of the MANNING. In this paper, all angles are given with respect to the vertical down direction, with 90° pointing toward the towship ALLIANCE and −90° pointing toward the support ship MANNING.

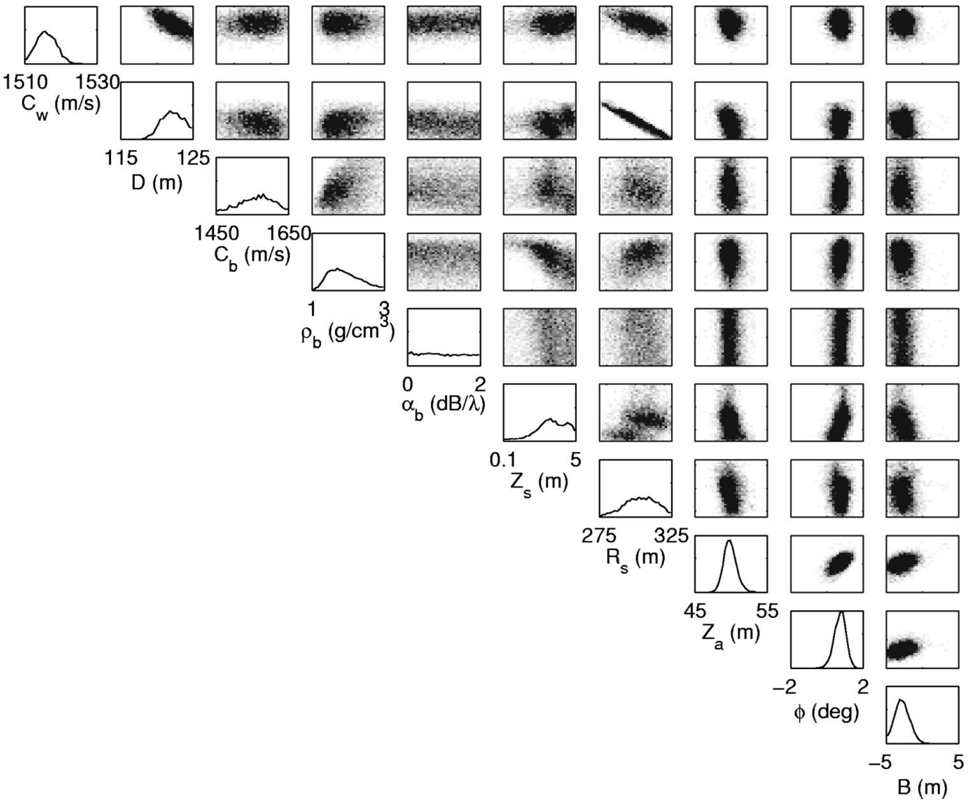

FIG. 7. 1D and 2D Marginal PPDs for frame 90. Environment (a), parabolic array and MANNING source (PPDs below the diagonal are symmetric with those above).

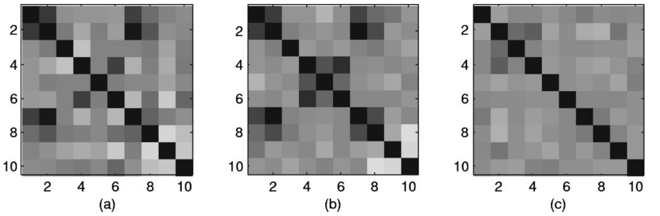

FIG. 8. Parameter correlation matrices for the Pekeris environment and parabolic array (parameter order as per Tables I and II): (a) Gibbs sampler estimate; (b) Estimate via Eq. (22); (c) Estimate via Eq. (21) after sampling in rotated coordinates. The parameter indices now relate to a new set of orthogonal coordinates.

Given the level of interference originating from nearby shipping traffic, it was found useful to incorporate additional spatial filtering of the snapshot data. Preprocessing therefore entailed temporal windowing and 2D FFT transformation of each snapshot, frequency masking to exclude certain ranges of spatial frequencies, and then inverse 1D FFT transformation back into phone-frequency space. In this analysis, the spatial frequency cutoff was set to include arrival angles within 60° of endfire to capture most of the expected multipath arrivals from the MANNING. To retain the normalization of Eq. (14), identical filtering was applied to the replica vectors $\mathbf{w}_f$.

C. Propagation modeling

Accurately modeling the response of acoustic environments is a principal aspect of geoacoustic inversion. In the present case of self-noise inversion, particular attention is required to aspects of short-range propagation that differentiate it from long-range waveguide propagation—namely the leaky or *virtual modes* that result from steep angles of bottom incidence.[1] In view of this, two approaches to propagation modeling capable of accuracy in the near field were used, the first of which—wave number integration—has already been detailed in Ref. 2. The second approach, based on the complex effective depth (CED) ideas of Zhang and Tindle,[36] was motivated by the modeling requirements of Gibbs sampling, which can easily run to the order of 10^5 models for a single inversion.

The first great advantage of CED models is that they are based on normal modes, which are characteristic of the environment and independent of source–receiver geometry. By keeping environmental parameters at the beginning of the Gibbs chain, it is possible to improve efficiency by reusing the mode functions to evaluate the array field for a wide variety of geometric variations. Second, by replacing the single-interface reflection coefficient of Zhang and Tindle with an invariant embedding scheme,[37] the capability of our CED code was extended to arbitrary bottom layering at little additional cost. The disadvantage of our existing CED code is that it cannot handle water column sound-speed gradients, which were instead modeled by wave number integration in this work. Modifications of our CED code along the lines described by Westwood *et al.*[38] are expected to remedy this problem, enabling Gibbs sampling to be applied efficiently to more complex environments for which wave number integration is currently slower by a factor of about 20.

TABLE III. Results for the array model selection. The three-point spline is seen to give the highest evidence. Note that the table displays log evidence and the probability of a model is based on just the evidence; thus, the small differences for each model are important.

		Log evidence	
Model	Min *E*	RIS	Gaussian
Parabola	20.1	−14.4	−14.3
Spline 3	**19.4**	**−14.2**	**−13.8**
Spline 4	19.1	−14.4	−13.8
Spline 5	18.9	−14.8	−14.1
Spline 6	17.9	−15.3	−14.4
Spline 7	17.7	−15.5	−14.6
Spline 8	17.3	−15.7	−15.5

D. Array modeling

From the standpoint of real data analysis, it is important to allow for geometric distortion of the array from its ideal horizontal and straight configuration. Whereas in Ref. 2 we modeled the array as a parabolic curve with bow B meters and tilt ϕ degrees, in this work we extended the description to a cubic spline passing through multiple equally spaced points. Each point has freedom in depth, allowing complicated array shapes to be modeled. Figure 4 is an illustration of both models, as it is our intention in Sec. V to compare each with respect to Bayesian evidence, thereby quantifying the level of complexity most appropriate.

V. RESULTS

In this section, we present inversion results obtained for frame 90 of the MAPEX2000 dataset using the MANNING as a source, and a variety of geoacoustic and array parametrizations. Strictly speaking, the selection of the array and environmental models should have been made concurrently; however, we undertook these analyses separately on the assumption that the array parameters were well determined by the data and hence hierarchically separable from the less

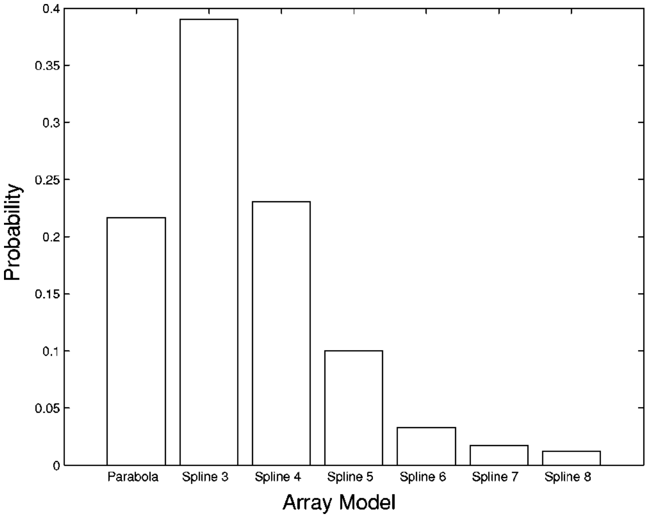

FIG. 9. Normalized probabilities for the towed-array models listed in Table III based on Bayesian evidence.

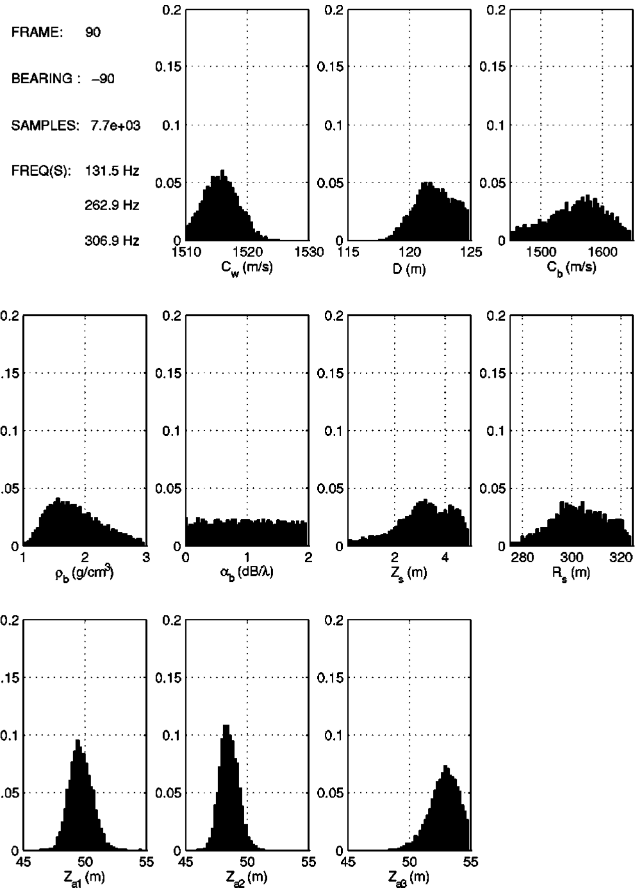

FIG. 10. Marginal PPD histograms for frame 90. Environment model (a), three-point array and MANNING source.

strongly determined environmental parameters. MacKay[6] remarks that this approach can often emerge as a reasonable approximation.

The structure of this section, then, is first in Sec. V A to revisit the baseline case of the parabolic array and Pekeris waveguide environment. Second, in Sec. V B, the behavior of Bayesian evidence with respect to array model complexity is analyzed assuming the Pekeris environment. Finally, in Sec. V C we fix the array model at the most probable configuration determined previously and compare the six environmental models illustrated in Fig. 5 with respect to their Bayesian evidence.

A. Parabolic array and Pekeris environment

The first result assumes the same simple Pekeris environment and parabolic array model as previously employed in Ref. 2, except that in this case, we have retained the bottom density ρ_b and attenuation α_b parameters previously neglected due to their perceived insensitivity. Table I details the search bounds used in the first inversion, which converged after 63 000 CED forward models (each at three frequencies) in approximately 1 h.

In Fig. 6, histograms generated by Gibbs sampling proportional to the marginal PPDs of the ten inverted parameters have been smoothly resolved into 50 bins, indicating good convergence. Table II summarizes the results of the first inversion in terms of MAP and mean parameter estimates in addition to standard deviations computed from the ensemble of samples. In most respects, the results obtained here agree well with those obtained earlier using a GA global search.[2] In particular, the MAP estimate of the bottom velocity C_p of 1580 m/s is identical to that obtained previously. In this case, however, at least the relative parameter sensitivities are clearly evident from the widths and curvatures of the marginal PPDs.

It is noteworthy that the bottom density ρ_b in Fig. 6 appears to have been well resolved. In fact, ρ_b appears as well resolved here as the bottom velocity C_b. As the bottom reflected arrivals are relatively steep, the magnitude of these arrivals is related to the impedance contrast at the bottom boundary. Physically, the estimated density of 1.9 ± 0.4 g/cm^3 agrees well with the value of 1.8 g/cm^3 commonly assumed for sediment densities in the North Elba area.[32,26,33]

Typifying the characteristics of a poorly resolved parameter, the variance of the *p*-wave bottom attenuation α_b is a large proportion of its prior range, and the marginal PPD is correspondingly flat. This confirms earlier observations that bottom attenuation is not a sensitive parameter in near-field inversion.

In Fig. 7, 1D and 2D marginal probability plots indicate relatively homogeneous PPDs for most parameter pairs, except in the case of water depth D and source range R_s, which shows a very strong positive correlation of the type commonly found in shallow-water acoustics.[21,30] This correlation is also apparent in the parameter covariance matrices, estimated via Eqs. (21) and (22) and plotted as *correlation* matrices ρ_{ij} in Figs. 8(a) and (b), respectively, such that

$$\rho_{ij}=C_{ij}/\sqrt{C_{ii}C_{jj}}. \tag{26}$$

As discussed in Sec. III D, the estimates obtained using the two methods should agree given sufficient samples and an accurate enough ML estimate. Visually, Figs. 8(a) and (b) bear a close resemblance, with high positive correlations between water depth D (parameter 2) and source range (parameter 7), and some negative correlations elsewhere. In subsequent sampling, these correlations were effectively removed through coordinate rotation as described previously, with the result being a more diagonally dominant correlation matrix for the orthogonalized parameter set, as illustrated in Fig. 8(c). Estimating **C** using Eq. (21) typically required ten thousand or so forward models to achieve 5% relative convergence, whereas using Eq. (22) took almost negligible time.

B. Array model selection

In this section, we isolate the issue of array selection from that of environment selection by holding the latter constant. Taking the simple Pekeris waveguide as a baseline environment, Table III summarizes the outcomes of Bayesian evidence calculations for seven array models, including the original parabolic array with three parameters, and six additional cubic spline arrays with up to eight parameters (equally spaced depth points). The second column of the table gives the minimum error energies achieved with each array, illustrating the point in Sec. II B that arbitrarily complex models can achieve arbitrarily good fits to the data. The third column gives the logarithms (to base 10) of the evi-

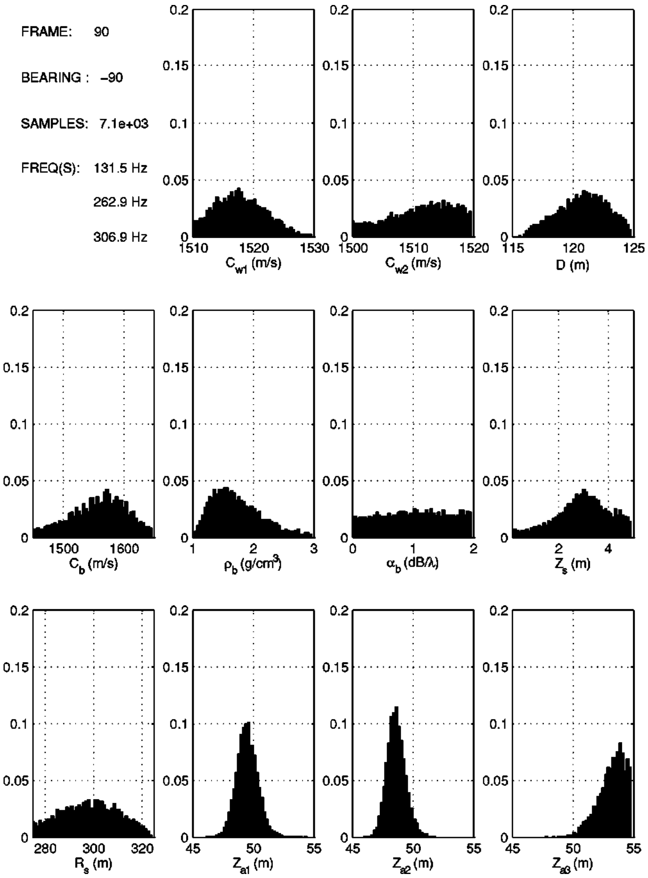

FIG. 11. Marginal PPD histograms for frame 90. Environment model (b), three-point array and MANNING source.

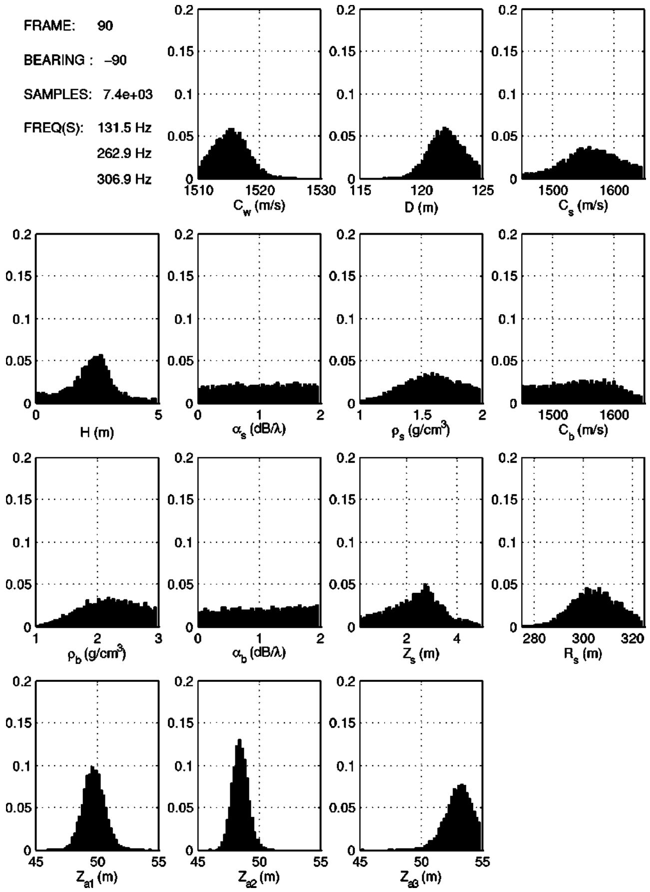

FIG. 12. Marginal PPD histograms for frame 90. Environment model (c), three-point array and MANNING source.

dence values as calculated by reverse importance sampling—the most accurate method used here. These log-evidence values can be compared with those in the fourth column, which were calculated by a fast approximate method based on the Gaussian assumption of Eq. (9). As discussed in Sec. II B, in the absence of prior preferences toward particular models, that with the highest evidence (or log evidence) is preferred, and the array model with the highest evidence in this case is the three-point spline. While array models of higher complexity achieved lower energies, the Occam factor implicit in the evidence formulation automatically discriminated against them, leading to a monotonic decrease in probability, as illustrated by the normalized probability plot of Fig. 9.

Interestingly, both the parabola and three-point spline possess the same number of parameters, though the geometric mapping of these parameters into the misfit function was different for each case. While an argument could be made that the particular prior bounds used weighed unreasonably against the parabolic array (or vice versa), these bounds were believed to be reasonably representative of the state of prior uncertainty. For this reason, the three-point spline array was selected here as the basis for subsequent inversions.

C. Environment model selection

Having established that the three-point spline had the highest Bayesian evidence of the seven array models evaluated, we now turn to evaluating evidence values for various environmental models. Figure 5 depicts schematically the six ocean-acoustic environments considered, including the simple Pekeris waveguide used as the baseline case above. Again, the choice of models was subjective—in this case based on previous geoacoustic inversions in the North Elba region[23,32,26,33] and also recent high-resolution surveys,[34,39] which have indicated near-surface stratification that might reasonably be detected by the present self-noise technique. The addition of sound-speed gradients to the water column, while seemingly unnecessary given the almost iso-velocity profile measured at sparse locations during the experiment, was of interest here because the impact of such profiles needs to be better understood in relation to near-field inversion.

Marginal PPDs generated by Gibbs sampling for each of the six environments—assuming the three-point spline array previously selected—appear in Figs. 10 to 15. In the case of model (a) in Fig. 10, it can be observed that the change in array parametrization has not significantly affected either the shapes of the PPDs or the parameter estimates obtained previously using a parabolic array. Similarly to Fig. 6, the source depth Z_s is observed to be multimodal, but becomes less so following the addition of a water column sound-speed gradient. For model (b) in Fig. 11, the velocity estimate for the lower water column was 1512±6 m/s relative to the surface velocity estimate of 1518±4 m/s, suggesting a significant deviation from the iso-velocity assumption made in Ref. 2.

Model (c) is again an iso-velocity model, with the bulk water sound-speed estimate C_w being 1516±3 m/s. How-

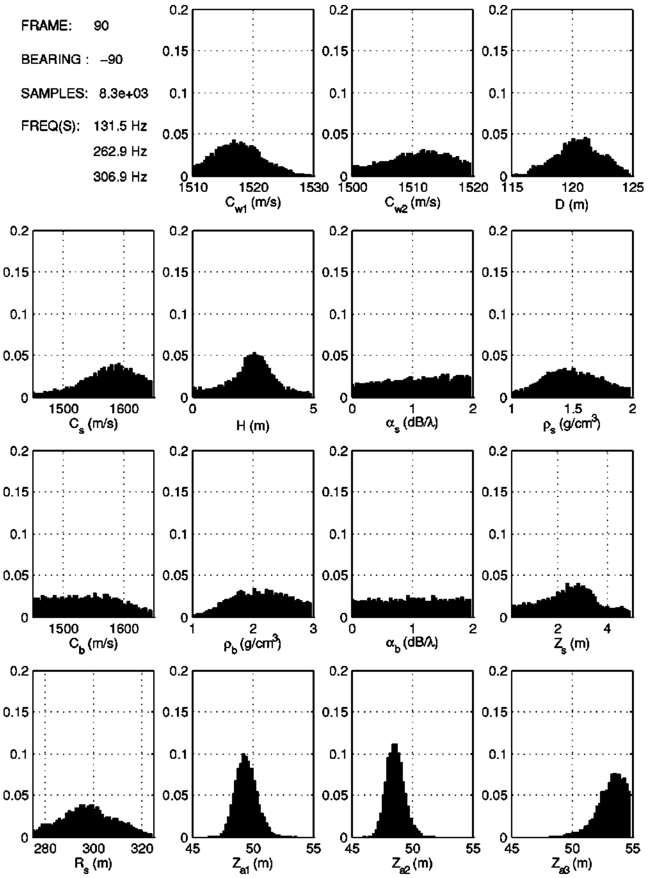

FIG. 13. Marginal PPD histograms for frame 90. Environment model (d), three-point array and MANNING source.

ever, in this case, the possibility of a sediment layer with its own density and p-wave velocity has been included. As is apparent in Fig. 12, the existence of such a sediment layer down to 2.3±1.0 m is suggested by the well-resolved peak in the sediment thickness PPD. This feature was subsequently confirmed to be common to most of the MAPEX2000 dataset between frames 90 and 117, and its estimated depth correlates very well with bottom structures proposed by Gingras and Gerstoft[26] for the North Elba seabed, as does the estimated sediment density of 1.6±0.2 g/cm^3. With the reintroduction of a water column sound-speed gradient, this peak remains unchanged; however, the sediment sound-speed estimate increases from 1565 to 1576 m/s while the sediment density estimate decreases to 1.5±0.2 g/cm^3. It should also be noted that in comparison with the sediment parameters, those of the half-space below are poorly resolved, with the only meaningful observation being that the half-space density is likely to be in excess of 2 g/cm^3.

Figures 14 and 15 illustrate the effect of introducing still further structure to the bottom acoustic model in the form of an additional iso-velocity fluid sediment layer. The prior bound of 20 m for the thickness of this layer was intended to reflect what was believed to be the overall sensitivity limit of the MAPEX2000 inversion geometry over the frequency range used—though this itself is a function of the actual bottom structure. Whereas the PPD of the H_1 layer thickness parameter becomes substantially flatter after adding a second layer, the H_2 PPD appears distinctly multimodal, with a sec-

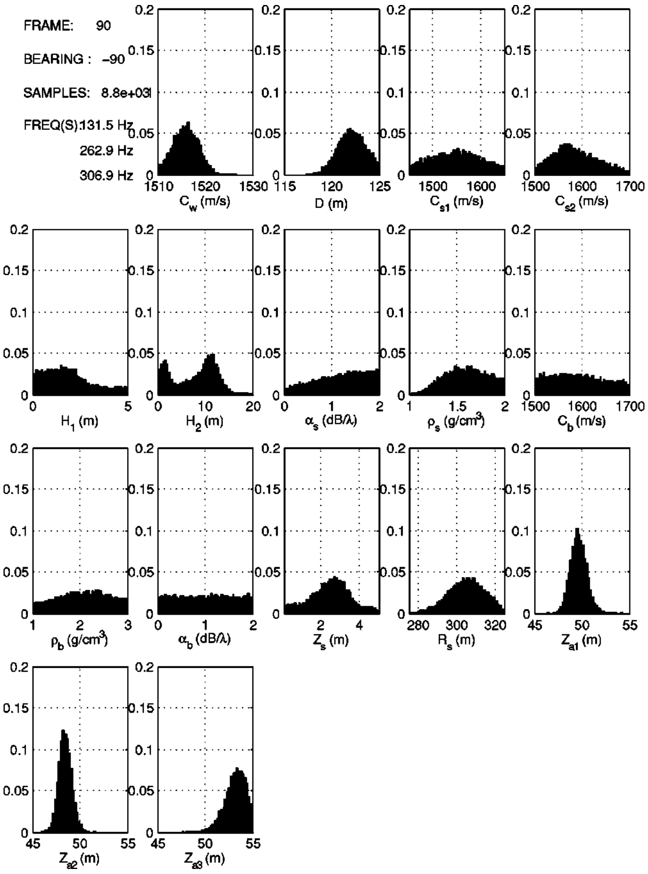

FIG. 14. Marginal PPD histograms for frame 90. Environment model (e), three-point array and MANNING source.

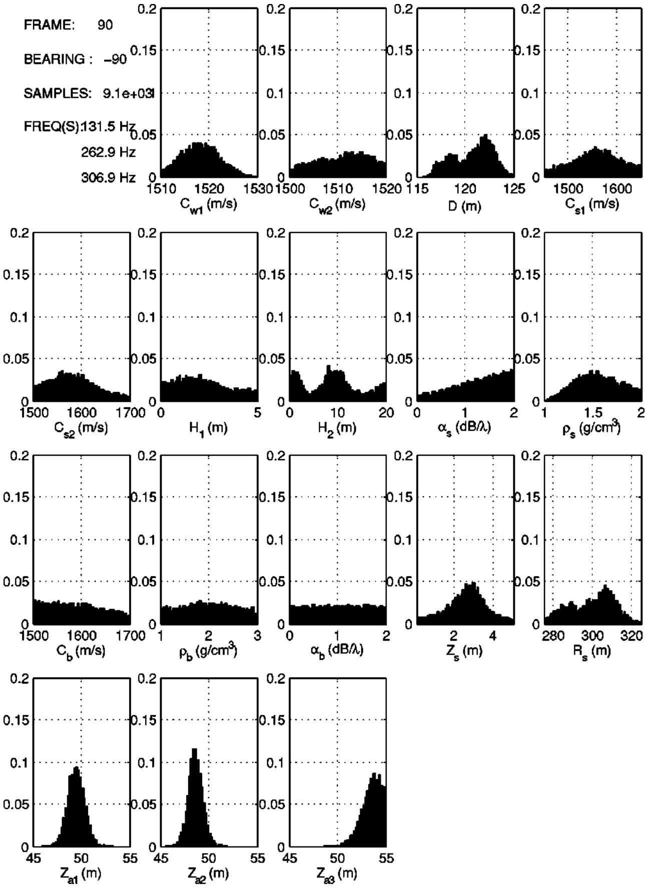

FIG. 15. Marginal PPD histograms for frame 90. Environment model (f), three-point array and MANNING source.

TABLE IV. Results for environment model selection. Model (d) is seen to give the highest evidence. Note that the table displays log evidence and the probability of a model is based on just the evidence; thus, the small differences for each model are important.

Model	Min E	Log evidence	
		RIS	Gaussian
(a)	19.4	-14.2	-13.8
(b)	19.1	-14.1	-13.7
(c)	19.7	-15.2	-14.7
(**d**)	**14.7**	$\mathbf{-13.6}$	$\mathbf{-12.7}$
(e)	15.0	-13.9	-13.1
(f)	14.9	-13.7	-13.1

ond peak at approximately 10 m. This is possibly on account of two actual sediment layers in the bottom, or also possibly just a resonant condition for the bottom interface reflectivity.

As done in Sec. V B for array model selection, the quantitative results of environment model selection are summarized in Table IV and Fig. 16. Unlike the previous case, it is difficult to rank competing environmental models *a priori* in relation to their parametric complexity, and hence the evidence does not behave monotonically either side of the optimal selection—model (d). The MAP and mean parameter estimates for this model, along with their associated standard deviations, represent the final output of the two levels of Bayesian inference employed in this paper, and are summarized in Table V.

Also unlike the previous case, the environment model found to have the highest Bayesian evidence was also that which achieved the best overall fit. Given the hierarchical relationship between models (d) and (f), it is at first surprising that the latter could not achieve a lower energy, though it is possible that the initial GA search simply failed to find the globally optimal solution. This is testimony to why it is more desirable to characterize PPDs in inverse problems rather than just search for the best fit. Further, it can be noted from Eq. (25) that the reverse importance sampling procedure used for accurate evidence estimation—unlike the Gaussian approximation of Eq. (9)—does not depend on the maximum likelihood estimate $p(\mathbf{d}|\mathcal{M}_k)$. The evidence values in the second column of Table IV, therefore, are unbiased by suboptimal ML solutions, unlike those in the third column—although in any case, both methods identified model (d) as the most probable.

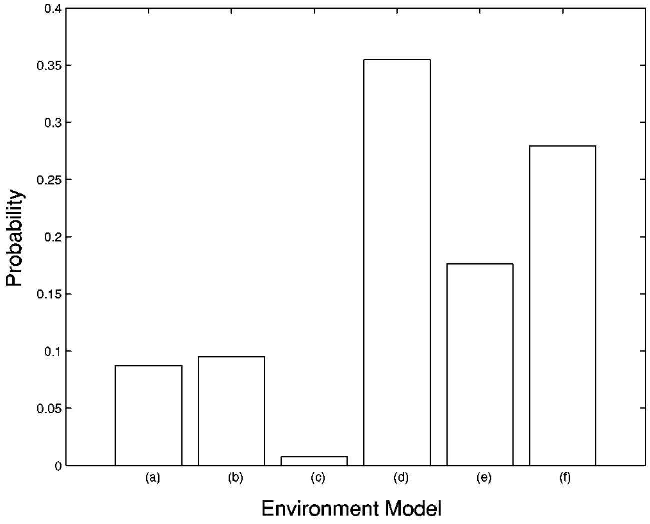

FIG. 16. Normalized probabilities for the environment models listed in Table IV based on Bayesian evidence.

TABLE V. Results of final geoacoustic inversion.

Parameter	Unit	MAP	Mean	Std. dev.
C_{w1}	m/s	1518	1518	4
C_{w2}	m/s	1512	1511	5
D	m	121	121	1.8
C_s	m/s	1598	1577	45
H	m	2.7	2.5	1
α_s	dB/λ	2.0	1.1	0.6
ρ_s	g/cm^3	1.5	1.5	0.2
C_b	m/s	1539	1540	51
ρ_b	g/cm^3	2.5	2.2	0.5
α_b	dB/λ	1.4	1.0	0.6
Z_s	m	2.0	2.5	1
R_s	m	301	300	10
z_{a1}	m	49.6	49.7	0.9
z_{a2}	m	48.7	48.8	0.7
z_{a3}	m	53.8	53.4	1.0

In relation to Fig. 16, there appears to be a consistent demarcation between environments with iso-velocity water columns as opposed to sound-speed gradients—the left and right columns of Fig. 5. The latter are always more probable than the former, with the difference between models (c) and (d) being almost anomalous. This suggests that even at the short ranges necessary for self-noise inversion, it is worthwhile modeling the water column as possessing at least a first-order sound-speed gradient.

Finally, on the basis of histogram plots such as those in Figs. 10 to 15, it is not easy to see the reasons for particular model preferences, but in numerous reruns, the essential pattern of Fig. 16 was fairly consistently reproduced. In any case, we stated in the Introduction that the focus of this paper was on the *methodology* of Bayesian model selection rather than exhaustive demonstration, which would need consistency checks across a more extensive portion of our data set than the single frame analyzed here, and also possibly a wider range of realistic ocean models.

VI. CONCLUSIONS

A Bayesian approach to solving self-noise inversion problems has been presented. This approach provides consistent methods for both parameter estimation at the first level of inference, and model selection at the second level of inference.

To implement the calculations necessary—namely PPD estimation, marginalization, and integration—we used a Metropolis–Hastings variant of the popular Gibbs sampler algorithm, which, combined with fast acoustic modeling, made the tasks of parameter estimation and model selection tractable. Variations on conventional Gibbs sampler design, such as an initial fast GA optimization, and coordinate rotation according to semianalytic covariance estimates proved successful in accelerating overall algorithm performance.

The second level of inference is concerned with model selection, i.e., the problem of selecting a model that best explains a data set at a given signal to noise ratio. Bayesian model selection, as discussed here, is based on calculating evidence, which is the integral of the product of likelihood and prior probability over all model parameters. This approach penalizes models that are more complex than warranted by the data by virtue of the Occam factor.

ACKNOWLEDGMENTS

This work was supported by NAVSEA/ASTO under Contract N00014-01-D-0043-D003 and by ONR under Grant N00014-03-1-0393.

[1] W. A. Kuperman, M. F. Werby, K. E. Gilbert, and G. J. Tango, ''Beam forming on bottom-interacting tow-ship noise,'' IEEE J. Ocean. Eng. **OE-10**(3), 290–298 (1985).

[2] D. J. Battle, P. Gerstoft, W. A. Kuperman, W. S. Hodgkiss, and M. Siderius, ''Geoacoustic inversion of tow-ship noise via near-field-matched-field processing,'' IEEE J. Ocean. Eng. **OE-28**(3), 454–467 (2003).

[3] S. E. Dosso, ''Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach,'' J. Acoust. Soc. Am. **111**, 129–142 (2002).

[4] S. E. Dosso and P. L. Nielsen, ''Quantifying uncertainty in geoacoustic inversion. II. Application to broadband, shallow-water data,'' J. Acoust. Soc. Am. **111**, 143–159 (2002).

[5] S. E. Dosso, ''Environmental uncertainty in environmental source localization,'' Inverse Probl. **19**, 419–431 (2003).

[6] D. J. C. MacKay, ''Bayesian interpolation,'' Neural Comput. **4**, 415–447 (1992).

[7] A. Malinverno, ''Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem,'' Geophys. J. Int. **151**, 675–688 (2002).

[8] J. J. K. Ó Ruanaidh and W. J. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing* (Springer, New York, 1996).

[9] A. Malinverno, ''A Bayesian criterion for simplicity in inverse problem parametrization,'' Geophys. J. Int. **140**, 267–285 (2000).

[10] H. Jeffreys, *Theory of Probability* (Oxford University Press, Oxford, 1939).

[11] K. Mosegaard and M. Sambridge, ''Monte Carlo analysis of inverse problems,'' Inverse Probl. **18**, 29–54 (2002).

[12] C. F. Mecklenbräuker, P. Gerstoft, J. F. Bohme, and P. Chung, ''Hypothesis testing for geoacoustic environmental models using likelihood ratio,'' J. Acoust. Soc. Am. **105**, 1738–1748 (1999).

[13] C. Andrieu and A. Doucet, ''Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC,'' IEEE Trans. Signal Process. **47**(10), 2667–2676 (1999).

[14] W. K. Hastings, ''Monte Carlo sampling methods using Markov chains and their applications,'' Biometrika **87**, 97–109 (1970).

[15] S. Geman and D. Geman, ''Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images,'' IEEE Trans. Pattern Anal. Mach. Intell. **6**, 721–741 (1984).

[16] A. E. Gelfand and A. F. M. Smith, ''Sampling-based approaches to calculating marginal densities,'' J. Am. Stat. Assoc. **85**, 398–409 (1990).

[17] A. F. M. Smith and G. O. Roberts, ''Bayesian computation via the gibbs sampler and related Markov chain Monte Carlo methods,'' J. R. Stat. Soc. Ser. B. Methodol. **55**, 3–23 (1993).

[18] M. K. Sen and P. L. Stoffa, ''Bayesian inference, Gibbs sampler and uncertainty estimation in geophysical inversion,'' Geophys. Prospect. **44**, 313–350 (1996).

[19] W. A. Kuperman, M. D. Colins, J. S. Perkins, and N. R. Davis, ''Optimal time-domain beamforming with simulated annealing including application of *a priori* information,'' J. Acoust. Soc. Am. **88**(4), 1802–1810 (1990).

[20] M. D. Collins and W. A. Kuperman, ''Focalization: Environmental focusing and source localization,'' J. Acoust. Soc. Am. **90**, 1410–1422 (1991).

[21] P. Gerstoft, ''Inversion of seismoacoustic data using genetic algorithms and *a posteriori* probability distributions,'' J. Acoust. Soc. Am. **95**, 770–782 (1994).

[22] 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).

[23] P. Gerstoft and D. F. Gingras, ''Parameter estimation using multifrequency range-dependent acoustic data in shallow water,'' J. Acoust. Soc. Am. **99**, 2839–2850 (1996).

[24] J. Hermand and P. Gerstoft, ''Inversion of broadband multitone acoustic data from the Yellow Shark summer experiments,'' IEEE J. Ocean. Eng. **OE-21**(4), 324–346 (1996).

[25] P. Gerstoft and C. F. Mecklenbräuker, ''Ocean acoustic inversion with estimation of *a posteriori* probability distributions,'' J. Acoust. Soc. Am. **104**, 808–819 (1998).

[26] D. F. Gingras and P. Gerstoft, ''Inversion for geometric and geoacoustic parameters in shallow water: Experimental results,'' J. Acoust. Soc. Am. **97**, 3589–3598 (1995).

[27] C. F. Mecklenbräuker and P. Gerstoft, ''Objective functions for ocean acoustic inversion derived by likelihood methods,'' J. Comput. Acoust. **8**(2), 259–270 (2000).

[28] P. Gerstoft, SAGA Users guide 2.0, an inversion software package, SACLANT Undersea Research Centre, SM-333, 1997. http://www.mpl.ucsd.edu/people/gerstoft/saga

[29] M. D. Collins and L. Fishman, ''Efficient navigation of parameter landscapes,'' J. Acoust. Soc. Am. **98**, 1637–1644 (1995).

[30] G. L. D'Spain, J. J. Murray, W. S. Hodgkiss, N. O. Booth, and P. W. Schey, ''Mirages in shallow water matched field processing,'' J. Acoust. Soc. Am. **105**, 3245–3265 (1999).

[31] A. B. Baggeroer, W. A. Kuperman, and H. Schmidt, ''Matched field processing: Source localization in correlated noise as an optimum parameter estimation problem,'' J. Acoust. Soc. Am. **83**, 571–587 (1988).

[32] F. B. Jensen, ''Comparison of transmission loss data for different shallow water areas with theoretical results provided by a three-fluid normal-mode propagation model,'' in *Sound Propagation in Shallow Water*, edited by O. F. Hastrup and O. V. Olesen, CP-14 (SACLANT ASW Research Centre, La Spezia Italy, 1974), pp. 79–92.

[33] D. D. Ellis and P. Gerstoft, ''Using inversion techniques to extract bottom scattering strengths and sound speeds from shallow-water reverberation data,'' in Proc. ECUA96, 1996, pp. 557–562.

[34] C. W. Holland and J. Osler, ''High-resolution geoacoustic inversion in shallow water: A joint time- and frequency-domain technique,'' J. Acoust. Soc. Am. **107**, 1263–1279 (2000).

[35] D. D. Ellis and D. M. F. Chapman, ''A simple shallow water propagation model including shear wave effects,'' J. Acoust. Soc. Am. **78**, 2087–2095 (1985).

[36] Z. Y. Zhang and C. T. Tindle, ''Complex effective depth of the ocean waveguide,'' J. Acoust. Soc. Am. **93**, 205–213 (1993).

[37] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP, New York, 1994).

[38] E. K. Westwood, C. T. Tindle, and N. R. Chapman, ''A normal mode model for acoustoelastic ocean environments,'' J. Acoust. Soc. Am. **100**, 3631–3645 (1996).

[39] C. W. Holland, ''Regional extension of geoacoustic data,'' in 5th *Europ. Conf. Underwater Acoustics*, edited by Zakharia, Chevret, and Dubail (European Commission, Luxembourg) 2000, pp. 793–796.