Uncertainty analysis in matched-field geoacoustic inversions

Chen-Fen Huang,[a] Peter Gerstoft, and William S. Hodgkiss
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238

(Received 29 April 2005; revised 18 October 2005; accepted 19 October 2005)

Quantifying uncertainty for parameter estimates obtained from matched-field geoacoustic inversions using a Bayesian approach requires estimation of the uncertainties in the data due to ambient noise as well as modeling errors. In this study, the variance parameter of the Gaussian error model, hereafter called error variance, is assumed to describe the data uncertainty. In practice, this parameter is not known *a priori*, and choosing a particular value is often difficult. Hence, to account for the uncertainty in error variance, several methods are introduced for implementing both the full and empirical Bayesian approaches. A full Bayesian approach that permits uncertainty of the error variance to propagate through the parameter estimation processes is a natural way of incorporating the uncertainty of error variance. Due to the large number of unknown parameters in the full Bayesian uncertainty analysis, an alternative, the empirical Bayesian approach, is developed, in which the posterior distributions of model parameters are conditioned on a point estimate of the error variance. Comparisons between the full and empirical Bayesian inferences of model parameters are presented using both synthetic and experimental data. [DOI: 10.1121/1.2139075]

PACS number(s): 43.30.Pc, 43.60.Pt [AIT] Pages: 197–207

I. INTRODUCTION

Ocean acoustic data inversions typically have focused just on inverting the parameters for one environmental model,[1–5] but some researchers have also considered selecting the best environmental parametrization over a family of geoacoustic models.[6,7] Under the Bayesian framework, all inferences are based on the posterior distribution $p(\mathbf{m}|\mathbf{d}, \boldsymbol{\eta}_0)$ given by

$$p(\mathbf{m}|\mathbf{d}, \boldsymbol{\eta}_0) \propto p(\mathbf{d}|\mathbf{m}, \boldsymbol{\eta}_0)p(\mathbf{m}|\boldsymbol{\eta}_0), \tag{1}$$

where $\mathbf{m}$ represents the geoacoustic model parameter vector, $\mathbf{d}$ represents the data, and $p(\mathbf{d}|\mathbf{m}, \boldsymbol{\eta}_0)$ and $p(\mathbf{m}|\boldsymbol{\eta}_0)$ represent the likelihood and prior distribution conditioned on $\boldsymbol{\eta}_0$, respectively. The symbol $\boldsymbol{\eta}$ refers to other possible unknown quantities in our mathematical model, such as uncertainty in signal characteristics as well as uncertainty of other parameters not included in $\mathbf{m}$ (e.g., ocean water column sound speed parameters). As indicated above, it is customary to keep these quantities at fixed values $\boldsymbol{\eta}_0$.

Under the Bayesian approach, unless there is absolute certainty regarding the value of $\boldsymbol{\eta}$, inference of $\mathbf{m}$ should be made by integrating out the effect of $\boldsymbol{\eta}$ from the joint posterior probability $p(\mathbf{m}, \boldsymbol{\eta}|\mathbf{d})$:

$$p(\mathbf{m}|\mathbf{d}) = \int p(\mathbf{m}, \boldsymbol{\eta}|\mathbf{d})\, d\boldsymbol{\eta} \tag{2}$$

$$= \int p(\mathbf{m}|\mathbf{d}, \boldsymbol{\eta})p(\boldsymbol{\eta}|\mathbf{d})\, d\boldsymbol{\eta}. \tag{3}$$

The second representation shows that the posterior distribution of interest, $p(\mathbf{m}|\mathbf{d})$, is a mixture of the conditional posterior distributions as shown in Eq. (1) given a fixed $\boldsymbol{\eta}$ where $p(\boldsymbol{\eta}|\mathbf{d})$ is a weighting function for the different possible values of $\boldsymbol{\eta}$. This is referred to as a full Bayesian approach. A major problem with this approach is that the number of possible parameters to include in the uncertainty analysis might be quite large.

An alternative, the empirical Bayesian approach,[8] is to replace $\boldsymbol{\eta}$ by a single estimate $\hat{\boldsymbol{\eta}}$ obtained from the data. Inference of $\mathbf{m}$ is now based on the estimated posterior distribution

$$p(\mathbf{m}|\mathbf{d}, \hat{\boldsymbol{\eta}}). \tag{4}$$

This simplified approach essentially replaces the integration in Eq. (2) by an estimation step. Since the full Bayesian approach accounts explicitly for the uncertainty in $\boldsymbol{\eta}$, the inference of $\mathbf{m}$ based on Eq. (2) should produce a more correct distribution than that based on Eq. (4). For the linear forward model case, adjustments to account for the uncertainty induced by estimating $\boldsymbol{\eta}$, especially to produce valid parameter variances, can be found in Ref. 8.

This study discusses several methods for implementing both the full and empirical Bayesian approaches with a focus on one important parameter usually not included in the Bayesian analysis: the error variance ν in a Gaussian error model. However, our methods are applicable to any nuisance parameter. The error variance is influenced by both errors in the data and systematic errors in modeling the data. While error

[a]To whom correspondence should be addressed. E-mail: chenfen@mpl.ucsd.edu

in the data, also known as noise, usually can be determined directly from the data (e.g., in the absence of signal), the systematic error is more difficult to assess. The error variance is important because incorrect choices for this parameter can seriously skew the posterior probability density (PPD) for the model parameters of interest.

The full Bayesian approach requires integrating out the nuisance parameters in Eq. (2) and either numerical or analytical *integration* can be used. Numerical integration is the most general approach as it can be carried out for any likelihood or prior distribution (see Sec. III A). Analytical integration is only possible for certain parameters with specific likelihood functions and prior distributions. For the error variance parameter in a Gaussian model, integrating out the error variance analytically (Sec. III B) makes this an attractive approach from both a computational and an analytical point of view.

For the empirical Bayesian approach, inferences are conditional on point estimates of the nuisance parameters in Eq. (4) and these can be estimated by either numerical or analytical *optimization* (as opposed to the integration used in the full Bayesian approach). Numerical optimization of the posterior probability with respect to both nuisance parameters and model parameters can easily be applied to most parameters and likelihood functions using standard optimization procedures[9,10] (Sec. III C). Analytic optimization is only feasible for certain combinations of likelihood functions and prior distributions. Assuming a Gaussian error model, it is possible to estimate the error variance analytically[2] (Sec. III D) and thus it is not necessary to use numerical optimization.

When estimating the error variance an interesting alternative to the point estimate (fixing the error variance at some specified value) is to use the analytic estimator of the error variance for each value of the model parameter vector.[11] This gives the same form of the posterior distribution as the full Bayesian approach (Sec. III D).

An objective of this study is the analysis of error variance. Since the computational expenses are of little concern for the example, an exhaustive evaluation of $p(\mathbf{m}|\mathbf{d})$ over a grid of parameter space combined with ordinary numerical integration is employed. This is a robust and accurate approach and is recommended for inverse problems with only a few parameters (e.g., less than eight parameters). However, if the number of parameters is large, Monte Carlo methods of numerical integration[4,7] should be used.

For the exhaustive integration, it is easier to assess the convergence than for the complex Monte Carlo methods. The convergence was assured by running the exhaustive search with a certain discretization and then comparing the result to a down sampled result. The integration is done by simply summing the enumerated values over the grid, since the parameters near the edge of the parameter space usually have less contribution to the integral.

The motivation of developing the full Bayesian approach is to avoid under/overestimating the data error variance and to understand how the data error uncertainty influences the uncertainty of the parameters of interest. The obtained parameter distributions can be used to make statistical predictions of various quantities of interest (e.g., transmission loss, as in Ref. 12).

The remainder of this paper is organized as follows. In the next section, the formulation of the inverse problem using the Bayesian approach is reviewed briefly. Section III outlines the approaches for handling error variance. Section IV provides an analytic expression for posterior probability distribution (PPD) of error variance. Section V presents the results and compares the model parameter posterior probability distributions using both synthetic and experimental data. Lastly, a few concluding remarks are made in Sec. VI.

II. FORMULATION OF THE INVERSE PROBLEM

In a Bayesian approach for geoacoustic inversions, inferences about the model parameter vector $\mathbf{m}$ based upon an observed data vector $\mathbf{d}$ are made in terms of probability density functions (pdf's). The basic formula for Bayesian parameter estimation is represented by the posterior probability density function, $p(\mathbf{m}|\mathbf{d})$, which by Bayes' theorem is given by

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d})}, \tag{5}$$

where $p(\mathbf{m})$ is the pdf associated with our *a priori* understanding of $\mathbf{m}$ before having access to the data $\mathbf{d}$.

The posterior probability density provides the full description of the state of knowledge about model parameters after observing the data. To interpret the multidimensional PPD, marginalization is used to summarize the PPD for a single parameter m_i by integrating over the remaining parameters $\mathbf{m}'$:

$$p(m_i|\mathbf{d}) = \int p(m_i,\mathbf{m}'|\mathbf{d})\, d\mathbf{m}'. \tag{6}$$

Also, 2-D marginal probability distributions of paired parameters can be obtained in a similar way. Further, the structure of the marginal posterior distribution is captured by the highest posterior density (HPD) interval (or region in the 2-D marginal) at a specified level of probability[13] (Sec. V A).

A. Single-frequency matched-field likelihood function

For matched-field geoacoustic inversions, the relationship between the observed complex-valued pressure field at a single frequency sampled at an N-element array and the predicted pressure field, at the frequency of interest, is described by the data model:

$$\mathbf{d} = \mathbf{D}(\mathbf{m}) + \mathbf{n}, \tag{7}$$

where $\mathbf{d}$ is the observed data and $\mathbf{D}(\mathbf{m})$ is the modeled data based upon a parametrized environmental model. In general, the modeled data is nonlinear with respect to the model parameter vector $\mathbf{m}$. The residual vector $\mathbf{n}$ represents the error terms. Typically, the residual vector is ambient noise but here the interpretation of $\mathbf{n}$ is broadened to include modeling errors.

If we assume that the error vector $\mathbf{n}$ is zero-mean complex Gaussian with covariance matrix $\mathbf{C}_D$, i.e., $\mathbf{n} \sim \mathcal{CN}(0, \mathbf{C}_D)$, then the likelihood function $p(\mathbf{d}|\mathbf{m})$ may be expressed as

$$p(\mathbf{d}|\mathbf{m}, \mathbf{C}_D) = \pi^{-N}|\mathbf{C}_D|^{-1} \times \exp[-(\mathbf{d}-\mathbf{D}(\mathbf{m}))^{\dagger}\mathbf{C}_D^{-1}(\mathbf{d}-\mathbf{D}(\mathbf{m}))], \quad (8)$$

where N is the number of elements in the array and superscript † denotes the complex conjugate transpose. Here, for simplicity, we also assume that the error terms may be described by independent and identically distributed (IID) complex Gaussian random variables with common variance ν, i.e., $\mathbf{C}_D = \nu\mathbf{I}$. This IID type of assumption is useful for convenience, but it may not completely model all the errors of interest. In what follows we shall always refer to the variable ν as the variance of the data errors.

The likelihood of the model parameter vector $\mathbf{m}$ for a given set of data may be written as

$$\mathcal{L}(\mathbf{m}, \nu, s) \equiv p(\mathbf{d}|\mathbf{m}, \nu, s) = \frac{1}{\pi^N \nu^N}\exp\left(-\frac{\|\mathbf{d}-\mathbf{d}(\mathbf{m})s\|^2}{\nu}\right), \quad (9)$$

in which the modeled data $\mathbf{D}(\mathbf{m})$ is represented by $\mathbf{D}(\mathbf{m}) = \mathbf{d}(\mathbf{m})s$, where $\mathbf{d}(\mathbf{m})$ is the replica field vector (or normalized signal field) computed using an acoustic propagation model for the model parameters $\mathbf{m}$, and s is the complex-valued source signature at the frequency of interest.

The source signature can be estimated either by the maximum-likelihood (ML) estimator, i.e., finding the value of s that maximizes the likelihood function,[2] or, should we have no interest in its value, by treating s as a nuisance parameter and eliminating it by integration (as will be discussed in Sec. III). Here we adopt the former method and obtain the ML estimate of the source parameter s as $s_{ML} = \mathbf{d}^{\dagger}(\mathbf{m})\mathbf{d}/\|\mathbf{d}(\mathbf{m})\|^2$. Substituting this relationship into Eq. (9) yields[2]

$$\mathcal{L}(\mathbf{m}, \nu) = \frac{1}{\pi^N \nu^N}\exp\left(-\frac{\phi(\mathbf{m})}{\nu}\right), \quad (10)$$

where $\phi(\mathbf{m})$ denotes an objective function defined as

$$\phi(\mathbf{m}) = \|\mathbf{d}\|^2\left[1 - \frac{|\mathbf{d}(\mathbf{m})^{\dagger}\mathbf{d}|^2}{\|\mathbf{d}\|^2\|\mathbf{d}(\mathbf{m})\|^2}\right] \quad (11)$$

in which the second term in the bracket is the normalized Bartlett power objective function[14] measuring the correlation between the data and the replica vectors. The objective function in Eq. (11) can be generalized[2] when multiple data snapshots are available.

B. Multi-frequency matched-field likelihood function

Assuming that the data errors are statistically independent across frequencies, then the multi-frequency matched-field likelihood function is the product of the single frequency counterparts:

$$\mathcal{L}(\mathbf{m}, \nu_1, \ldots, \nu_J) = \prod_j^J \mathcal{L}_j(\mathbf{m}, \nu_j), \quad (12)$$

where J indicates the number of the processed frequencies and $\mathcal{L}_j(\mathbf{m}, \nu_j)$ is the jth frequency likelihood with the error variance denoted by ν_j as in Eq. (10). To illuminate the significance of ν_j, Eq. (10) is rewritten as

$$\mathcal{L}_j(\mathbf{m}, \nu_j) \propto \exp\left(-\frac{\phi_j(\mathbf{m})}{\nu_j} - N\ln\nu_j\right) \quad (13)$$

in which $1/\nu_j^N$ has been expressed as $\exp(-N\ln\nu_j)$ and the constant π^{-N} is omitted.

1. Frequency-dependent error variance

Error variance is frequency dependent. Not only does ambient noise vary across frequency but the error due to modeling mismatch also varies across frequency. With the assumption that errors are independent across frequencies, as in Eq. (12), the likelihood of $\mathbf{m}$ for multi-frequency cases is the product of the marginal likelihoods of $\mathbf{m}$ for each frequency, with ν_j being integrated out:

$$\mathcal{L}(\mathbf{m}) = \prod_j^J \int_{\nu_j} \mathcal{L}_j(\mathbf{m}, \nu_j)p(\nu_j)\,d\nu_j, \quad (14)$$

where $p(\nu_j)$ is the prior distribution of ν_j which will be specified in Sec. II C.

2. A single global error variance

A common approach is to assume the variation of the data error variances ν_j over the selected frequencies is negligible and model them by a single variable ν_0, i.e.,

$$\nu_j = \nu_0, \quad \text{for } j = 1, \ldots, J. \quad (15)$$

The likelihood for the selected frequencies with a common error variance ν_0 can be written as

$$\mathcal{L}(\mathbf{m}, \nu_0) \propto \exp\left(-\frac{J}{\nu_0}\bar{\phi}^a(\mathbf{m}) - JN\ln\nu_0\right), \quad (16)$$

where $\bar{\phi}^a(\mathbf{m}) = (1/J)\Sigma\,\phi_j(\mathbf{m})$ is the arithmetic mean of the objective function over frequencies.

Equations (14) and (16) hold under the assumption that the errors are independent for each frequency. However, when the errors due to frequency-dependent modeling mismatch are the dominant source of error, the modeling error may not be independent across the frequencies used. Therefore, the number of frequencies J must be replaced with an effective number of frequencies J_{eff}.

C. Noninformative priors

Before applying Bayes' theorem to make inference of parameters, one needs to specify their prior pdf's. The natural choice for a prior pdf is the distribution that allows for the greatest uncertainty while obeying the constraints imposed by prior knowledge. We treat here the case where very little is known about $\mathbf{m}$ and ν *a priori*.

Starting from the model parameters $\mathbf{m}$, all one knows is that the values of the parameters are within lower bounds l_i

TABLE I. Summary of the approaches (single frequency[a]).

Section: Approach	PPD of $\mathbf{m}$	Error variance	Remark
III A: Full Bayesian (numerical)	$\int p(\mathbf{m},\nu\|\mathbf{d})\,d\nu$	$\int p(\mathbf{m},\nu\|\mathbf{d})\,d\mathbf{m}$	Theoretically preferred but computationally expensive
III B: Full Bayesian (analytic)	$\dfrac{1}{\phi^N(\mathbf{m})}$	$\nu^{-[(2N-M^*)/2+1]}\exp\left(-\dfrac{\phi(\hat{\mathbf{m}})}{\nu}\right)$[b]	Theoretically & computationally preferred
III C: Empirical Bayesian (numerical)	$p(\mathbf{m}\|\mathbf{d},\nu_{\mathrm{MAP}})$	$\nu_{\mathrm{MAP}}=\arg\max_{\nu} p(\mathbf{m},\nu\|\mathbf{d})$	Computationally efficient
III D: Empirical Bayesian (analytic)	$p(\mathbf{m}\|\mathbf{d},\nu_{\mathrm{ML}})$	$\nu_{\mathrm{ML}}=\dfrac{\phi(\hat{\mathbf{m}})}{N}$	Computationally preferred

[a]As for the multi-frequency data, see Eq. (24) for the details.
[b]The formula is derived in Sec. IV.

and upper bounds u_i based upon the measurements independent of the acoustic data. We assume that there is no prior preference for any value over any other; then a uniform distribution over that range is a practical choice for the prior,

$$p(m_i)=\frac{1}{u_i-l_i},\quad l_i<m_i<u_i. \tag{17}$$

For the error variance parameter ν, all one knows about this parameter *a priori* is that it is always positive and one is equally ignorant about its value or any of its powers (e.g., a standard deviation or a variance). The appropriate distribution for this parameter is a uniform pdf on $\ln\nu$.[15] In practice, the limits of this prior do not go all the way to zero and infinity. We usually know in advance that ν cannot be much less than the ambient noise variance estimate (ν_{AN}) and cannot be much greater than the maximum likelihood estimate (ν_{ML}). Thus,

$$p(\nu)\propto\frac{1}{\nu},\quad 0.1\nu_{\mathrm{AN}}<\nu<10\nu_{\mathrm{ML}}. \tag{18}$$

Note that using broader limits for the error variance is to obtain a full distribution where the upper and lower ends have zero probability.

With the additional assumption that all model parameters $(m_1,\ldots,m_M)$ and ν are mutually independent, the prior distribution is the product of the prior distributions for each parameter:

$$p(\mathbf{m},\nu)=p(\mathbf{m})p(\nu)=\prod_i^M p(m_i)p(\nu)\propto\frac{1}{\nu} \tag{19}$$

over the interval where prior probability of $\mathbf{m}$ is nonzero. Then, based upon Bayes' theorem, the posterior distribution is as follows:

$$p(\mathbf{m},\nu|\mathbf{d})\propto p(\mathbf{d}|\mathbf{m},\nu)p(\mathbf{m},\nu)\propto\mathcal{L}(\mathbf{m},\nu)\frac{1}{\nu}$$
$$\propto\exp\left(-\frac{\phi(\mathbf{m})}{\nu}-(N+1)\ln\nu\right) \tag{20}$$

with the scale factor that makes the posterior distribution integrate to one being omitted. The final representation of Eq. (20) will be extensively used in this analysis.

Note that the posterior distribution for a uniform prior on ν or for a uniform prior on $\ln\nu$ differs only in an N or $N+1$ in front of $\ln\nu$ [Eq. (20)]. This suggests that for reasonably large N their respective PPDs are similar.

III. ERROR VARIANCE AS A NUISANCE PARAMETER

In this section, we shall treat the error variance in a Gaussian error model as a nuisance parameter and discuss both the full and empirical Bayesian methods from an implementation perspective. For convenience of comparison, the approaches discussed below are summarized in Table I.

A. Full Bayesian estimation—numerical integration

The full Bayesian approach is a natural way of incorporating the uncertainty of error variance in the analysis. The approach does not assume the error variance at a particular value, rather, it regards the error variance as an unknown in the parameter space. In this way, the approach allows the data uncertainty to propagate through the parameter estimation processes and, at the end, reflect uncertainty in the error variance in the resulting parameter estimation.

Therefore, the true posterior distribution of the model parameters is obtained by integrating out ν from the joint posterior distribution of $\mathbf{m}$ and ν:

$$p(\mathbf{m}|\mathbf{d})=\int p(\mathbf{m},\nu|\mathbf{d})\,d\nu. \tag{21}$$

B. Full Bayesian estimation—analytic integration

In this approach, the error variance is considered as a nuisance parameter and is eliminated by integrating the likelihood function weighted by the prior distribution of ν over the entire range:[16]

$$p(\mathbf{m}|\mathbf{d}) = \int_0^\infty p(\mathbf{m}, \nu|\mathbf{d})\, d\nu$$

$$\propto p(\mathbf{m}) \underbrace{\int_0^\infty p(\mathbf{d}|\mathbf{m}, \nu) p(\nu)\, d\nu}_{p(\mathbf{d}|\mathbf{m})}. \tag{22}$$

Incorporating the noninformative prior of ν in Eq. (18), the exact expression of $p(\mathbf{d}|\mathbf{m})$ can be shown to be of the form[17]

$$p(\mathbf{d}|\mathbf{m}) = \frac{1}{\pi^N} \frac{(N-1)!}{\phi^N(\mathbf{m})}. \tag{23}$$

This likelihood function is preferable in estimating PPD of model parameters for two reasons. Theoretically, this formula is derived based on a full Bayesian methodology. Computationally, this method is faster than a computer-based numerical integration of the full Bayesian procedure.

The above analytic solution, Eq. (23), can be extended straightforwardly to the multi-frequency data set. From Eq. (14), the multi-frequency likelihood function can be written in a concise form:

$$\mathcal{L}(\mathbf{m}) \propto \left(\frac{1}{\bar{\phi}^g(\mathbf{m})} \right)^{NJ} \tag{24}$$

where $\bar{\phi}^g(\mathbf{m}) = \sqrt[J]{\Pi \phi_j(\mathbf{m})}$ is the geometric mean of the objective function over frequency when the error variance is frequency dependent. However, for the case where the error variance is assumed to be constant over the processed frequencies, the arithmetic mean $\bar{\phi}^a(\mathbf{m})$ is used instead of $\bar{\phi}^g(\mathbf{m})$ in Eq. (24) in a manner analogous to that used by Ref. 11.

C. Empirical Bayesian estimation—optimizing error variance jointly with model parameters

The main idea behind this approach is to consider the error variance on the same level as the model parameters and optimize the joint posterior probability $p(\mathbf{m}, \nu|\mathbf{d})$ to find the estimate of ν:

$$\nu_{\mathrm{MAP}} = \arg\max_{\nu} p(\mathbf{m}, \nu|\mathbf{d}). \tag{25}$$

Then, the posterior distribution of the model parameters is approximated by fixing the error variance at ν_{MAP}:

$$p(\mathbf{m}|\mathbf{d}) \approx p(\mathbf{m}|\mathbf{d}, \nu_{\mathrm{MAP}}) \propto \exp\left[-\frac{\phi(\mathbf{m})}{\nu_{\mathrm{MAP}}} \right]. \tag{26}$$

This approach is easier to implement numerically compared to the full Bayesian approach because the value of the error variance can be found by any efficient optimization procedure, such as simulated annealing or genetic algorithms.[18]

D. Empirical Bayesian estimation—maximum likelihood estimate

The maximum likelihood estimator for the error variance can be obtained analytically

$$\nu_{\mathrm{ML}}(\mathbf{m}) = \frac{\phi(\mathbf{m})}{N}. \tag{27}$$

Two approaches for implementing this estimator in ocean acoustic inversions have been proposed by Mecklenbräuker and Gerstoft.[11]

Following an empirical Bayesian methodology, one requires only an estimate of the error variance. First, the ML solution of model parameter vector, $\hat{\mathbf{m}}$, is found [by minimizing the objective function, Eq. (11), over all $\mathbf{m}$]. Second, an estimate of error variance is obtained:[2]

$$\nu_{\mathrm{ML}}(\hat{\mathbf{m}}) = \frac{\phi(\hat{\mathbf{m}})}{N}. \tag{28}$$

Then, the PPD of $\mathbf{m}$ is approximated by fixing ν at $\nu_{\mathrm{ML}}(\hat{\mathbf{m}})$:

$$p(\mathbf{m}|\mathbf{d}, \nu_{\mathrm{ML}}(\hat{\mathbf{m}})) \propto \exp\left[-\frac{\phi(\mathbf{m})}{\nu_{\mathrm{ML}}} \right]. \tag{29}$$

The error variance estimated from either ν_{MAP} [Eq. (25)] or ν_{ML} [Eq. (28)] often results in overly optimistic posterior distributions of the model parameters since a single value of the error variance may not be representative.

The other approach proposed in Ref. 11 is to substitute Eq. (27) into the likelihood formula [Eq. (10)] without fixing a value for ν. With the noninformative prior for $\mathbf{m}$, the PPD of $\mathbf{m}$ is proportional to the likelihood function

$$p(\mathbf{m}|\mathbf{d}, \nu_{\mathrm{ML}}(\mathbf{m})) \propto p(\mathbf{m}) \mathcal{L}(\mathbf{m}, \nu_{\mathrm{ML}}(\mathbf{m})) \propto \mathcal{L}(\mathbf{m}, \nu_{\mathrm{ML}}(\mathbf{m}))$$

$$\propto \frac{1}{\phi^N(\mathbf{m})}. \tag{30}$$

Note that the likelihood formula derived in Sec. III B [Eq. (23)] and the result derived above [the third representation of Eq. (30)] possess the same functional form.

IV. PPD OF THE ERROR VARIANCE

The posterior distribution of ν is obtained by integrating the joint PPD over the model parameter vector:

$$p(\nu|\mathbf{d}) = \int p(\mathbf{m}, \nu|\mathbf{d})\, d\mathbf{m}. \tag{31}$$

One could always evaluate the integral in Eq. (31) numerically, but an analytic expression can be obtained under the simplifying Gaussian approximation for the PPD of the model parameters. This approach is known as Laplace's method, a family of asymptotic techniques used to approximate integrals.[19]

Inspired by Malinverno,[20] let us assume that, for any value of ν, the PPD of the model parameter vector is approximated by a Gaussian function of $\mathbf{m}$ centered on the MAP solution $\hat{\mathbf{m}}$ (Ref. 21) with the posterior covariance matrix of model parameters $\hat{\mathbf{C}}_{\mathrm{M}}$ (the hat is used to denote the quantity inferred *a posteriori*):

$$p(\mathbf{m}, \nu|\mathbf{d}) \approx p(\hat{\mathbf{m}}, \nu|\mathbf{d}) \exp\left(-\frac{(\mathbf{m}-\hat{\mathbf{m}})^T \hat{\mathbf{C}}_{\mathrm{M}}^{-1} (\mathbf{m}-\hat{\mathbf{m}})}{2} \right). \tag{32}$$

Since an unnormalized Gaussian pdf has the following constant,

$$\int \exp\left(-\frac{\mathbf{m}^T\hat{\mathbf{C}}_{\mathrm{M}}^{-1}\mathbf{m}}{2}\right) d\mathbf{m} = (2\pi)^{M/2}\sqrt{|\hat{\mathbf{C}}_{\mathrm{M}}|}, \tag{33}$$

substituting Eq. (20) into $p(\hat{\mathbf{m}}, \nu|\mathbf{d})$ yields the approximation to the marginal PPD of ν:

$$p(\nu|\mathbf{d}) \approx \frac{a}{\nu^{N+1}} \exp\left(-\frac{\phi(\hat{\mathbf{m}})}{\nu}\right)\sqrt{|\hat{\mathbf{C}}_{\mathrm{M}}|}, \tag{34}$$

where a is a scale factor that makes $p(\nu|\mathbf{d})$ integrate to one.

To evaluate the posterior covariance matrix of model parameters $\hat{\mathbf{C}}_{\mathrm{M}}$, the forward model is linearized with respect to the best-fit model vector

$$\mathbf{D}(\mathbf{m}) \approx \mathbf{K}(\mathbf{m}-\hat{\mathbf{m}}) + \mathbf{D}(\hat{\mathbf{m}}), \tag{35}$$

where $\mathbf{K} = [\partial\mathbf{D}(\mathbf{m})/\partial\mathbf{m}]|_{\mathbf{m}=\hat{\mathbf{m}}}$ is an $N \times M$ matrix of Fréchet derivatives evaluated at the best-fit model solution $\hat{\mathbf{m}}$. Comparing the exponent of Eq. (32) with Eq. (8) in which the nonlinear forward model is substituted by Eq. (35) gives

$$\hat{\mathbf{C}}_{\mathrm{M}} = (2\mathbf{K}^{\dagger}\mathbf{C}_{\mathrm{D}}^{-1}\mathbf{K})^{-1}. \tag{36}$$

Since the data errors are expected *a priori* to be IID with variance ν ($\mathbf{C}_{\mathrm{D}} = \nu\mathbf{I}$) and with the further assumption that the model parameters are resolved by the data set, the determinant of $\hat{\mathbf{C}}_{\mathrm{M}}$ is

$$|\hat{\mathbf{C}}_{\mathrm{M}}| = \left|\frac{\nu}{2}(\mathbf{K}^{\dagger}\mathbf{K})^{-1}\right| \propto |\nu\mathbf{I}_{M^*}| \propto \nu^{M^*}, \tag{37}$$

where M^* is the number of model parameters resolved by the data. Generally speaking, M^* is determined by the number of elements in the model parameter vector when $\mathbf{K}$ has full rank (all the model parameters are constrained by $\mathbf{K}$). However, due to the presence of noise, the parameter sensitivity and the possibility of linear dependence between the model parameters, not all of these may be estimated to a useful accuracy. Therefore, $M^* \leq M$. Then, substituting Eq. (37) into Eq. (34), we obtain the following approximation

$$p(\nu|\mathbf{d}) \propto \nu^{-[(2N-M^*)/2+1]} \exp\left(-\frac{\phi(\hat{\mathbf{m}})}{\nu}\right). \tag{38}$$

This approximation to the marginal PPD of ν is the so-called inverse chi-square distribution used in Bayesian analysis (see Ref. 13, Sec. 2.3.1). However, the inverse chi-square distribution presented in Ref. 13 is based on real-valued Gaussian random variables. For the complex-valued Gaussian random variables, the scaled error variance $\nu/(2\phi(\hat{\mathbf{m}}))$ has the inverse chi-square probability density with $2N-M^*$ degrees of freedom.

Notice that M^* can be found from comparing the analytic result with the PPD of ν obtained from the numerical integration of the full Bayesian, or it can be estimated from the rank of the posterior covariance matrix of the model parameters $\hat{\mathbf{C}}_{\mathrm{M}}$.

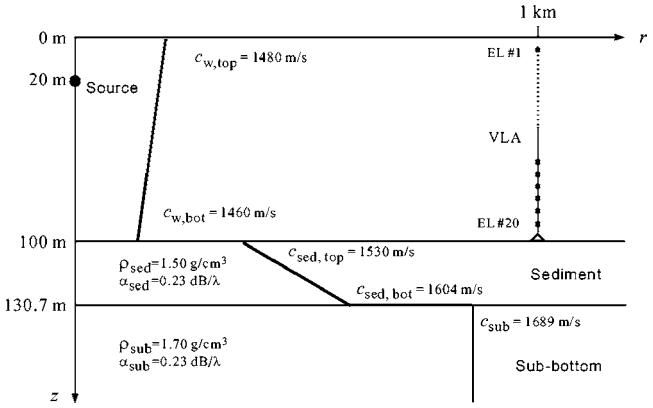

FIG. 1. The **sdc** environment from the Geo-Acoustic Inversion [22] shown here for the parameters corresponding to ground true.

V. RESULTS AND DISCUSSION

A. Synthetic data

To illustrate the various approaches presented in Sec. III, a data set is synthesized using the environmental model employed in the Geo-Acoustic Inversion Workshop 1997.[22] Figure 1 shows the baseline model that consists of a downward refracting sound speed profile overlying a positive-gradient sediment layer atop of a homogeneous subbottom layer. The vertical array consists of 20 hydrophones equally spaced over a 95-m interval with the first phone at 5-m depth, and the source located at 1-km range and 20-m depth, transmitting CW tones at 100 and 200 Hz. The calculations of acoustic fields are performed by the normal-mode propagation model ORCA.[23] The interaction of acoustic fields with the baseline environment for these two frequencies is plotted in Fig. 2.

In order to demonstrate the effect of error variance on the parameter estimation, the amount of noise corresponding to 20-dB SNR [equivalent to $\nu_{\mathrm{true}} = 0.0083$, see Eq. (A6) in the Appendix] is purposely added to the data. The parameters to be estimated are the geoacoustic parameters, including sediment thickness d, top and bottom sediment sound speeds $c_{\mathrm{sed,top}}$ and $c_{\mathrm{sed,bot}}$, subbottom sound speed c_{sub}, and the error variance ν. Figure 3 shows the parameter estimate using the full Bayesian approach for a frequency of 100 Hz. The line subplots along the diagonal are the one-dimensional (1-D) marginal PPDs for each parameter, $p(m_i|\mathbf{d})$, and the contour subplots in the upper triangle are the 2-D marginal PPDs corresponding to the paired parameters in the bottom-most and left-most line subplots, $p(m_b, m_l|\mathbf{d})$. In each contour plot, the gray-scale coloring from darkest to lightest represents 50%, 75%, and 95% highest posterior density (HPD). The β% HPD describes a region which contains β% of the total probability.[13] Due to the nonlinearity of the forward model, the PPD of model parameters is no longer Gaussian. Therefore, the best-fit model (cross signs in 2-D contours or arrow lines in 1-D plots) is not necessarily coincident with those from the mode of the marginal (plus signs in 2-D).

The 1-D and 2-D marginal PPDs reveal the uncertainty of the parameter estimation but in addition the 2-D PPDs also show the correlations between the paired parameters. For example, the contour subplot on row 1 and column 3

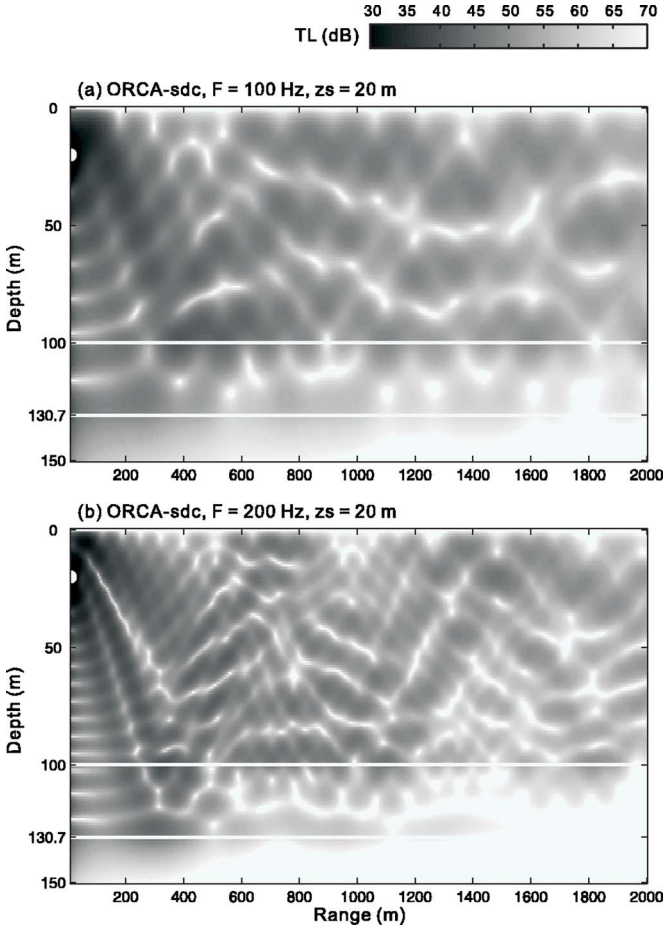

FIG. 2. Range-depth transmission loss for (a) $F=100$ and (b) 200 Hz, respectively, using the **sdc** environment. The two white lines mark the water-sediment (top) and sediment-subbottom (bottom) interfaces.

shows the correlation between bottom sediment sound speed $c_{\text{sed,bot}}$ and sediment thickness d. The result suggests that there is a strong positive coupling between these two parameters. Therefore, the interparameter correlation results in a relatively flat distribution in the 1-D marginal PPDs for the parameters $c_{\text{sed,bot}}$ and d. If more information about one of these two parameters could be obtained, then the 1-D marginal of the other could be sharpened. Likewise, the parameters of $c_{\text{sed,bot}}$ and $c_{\text{sed,top}}$ are strongly correlated in a negative manner. However, the 2-D PPDs for each pair of m_i and ν show that the error variance has little correlation with any other geoacoustic parameters, i.e., $p(m_i,\nu|\mathbf{d})=p(m_i|\mathbf{d})p(\nu|\mathbf{d})$.

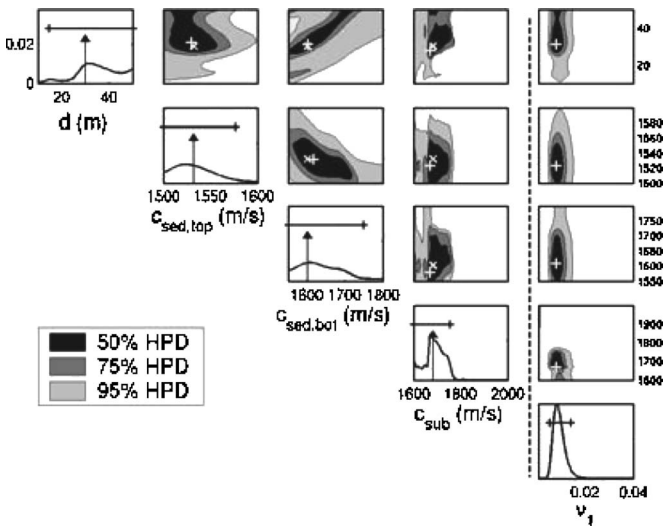

FIG. 3. Marginal posterior probability densities (PPDs) of the geoacoustic parameters as well as the error variance for $F=100$ Hz. In each 1-D marginal, the horizontal error bar shows 95% highest posterior density (HPD) interval.

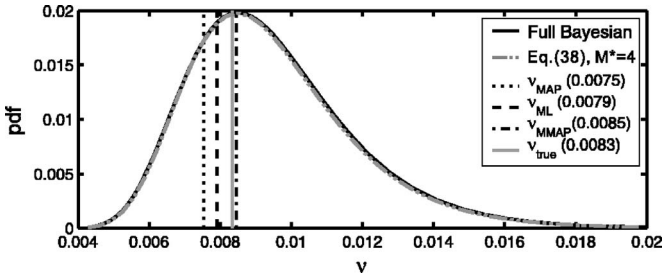

FIG. 4. The enlarged version of the marginal posterior distribution of the error variance in Fig. 3. Note that ν^{MMAP} is the MAP estimate of the marginal PPD of ν; it is not to be confused with the MAP estimate of the multidimensional joint PPD, ν^{MAP}

The error variances estimated from the various approaches are summarized in Fig. 4. First, the PPD of ν obtained from the numerical integration is compared with that from the analytic integration. We find that by setting the parameter $M^*=4$ in the analytic formula of the error variance distribution [Eq. (38)], the analytic result (dash-dot-dot curve) has excellent agreement with the numerical integration result (solid curve). Note that even though the PPDs of the geoacoustic parameters are only very approximately Gaussian, the PPD of ν follows an inverse chi-squared density. In this example, the analytic expression of the posterior distribution of ν agrees well with the full Bayesian result, given the number of well-determined parameters used.

Second, the point estimates of error variance from the empirical Bayesian approaches are summarized in Fig. 4. Since the error variance is distributed *a posteriori* as an inverse chi-square with $(2N-4)$ degrees of freedom, one may consider the peak of this marginal posterior distribution as an estimate of the error variance, named as the marginal MAP (MMAP) value of error variance, ν_{MMAP}. The solid vertical line shows the actual value of the error variance added to the data (0.0083). The ML (dash), MAP (dot), and MMAP (dash-dot) estimates of error variance are also shown. Among the various point estimates of error variance, ν_{MMAP} is the largest since it automatically takes into account the reduction in the degrees of freedom (for the inverse chi-square distribution) in the process of integration over the model parameters. The difference between the MAP and ML estimates of the error variance is due to the $1/\nu$ prior being used. In this synthetic data case, since the only uncertainty is the random error added to the data, there is not much difference in the estimated error variances among the various approaches.

The comparison of the marginal PPDs for each of the geoacoustic parameters is given in Fig. 5. The 1-D PPDs are estimated using full Bayesian treatment of error variance via numerical integration (solid; Sec. III A) and these using the

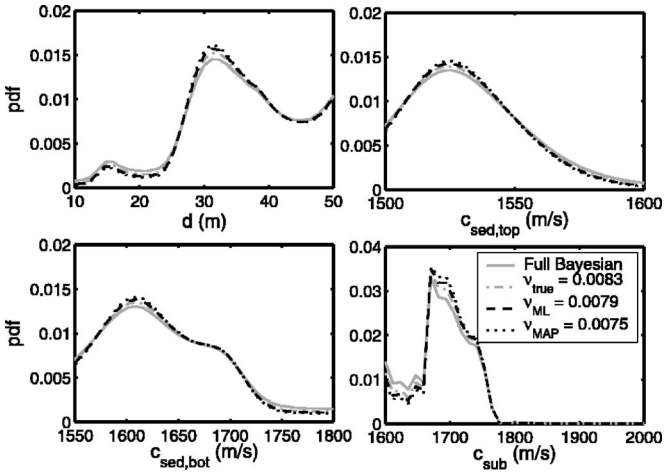

FIG. 5. Comparison of the marginal PPDs for the geoacoustic parameters using different approaches in handling the error variance. $F=100$ Hz).

empirical Bayesian methods based on the true value (dash-dot) and the ML (dash) and MAP (dot) estimates of the error variance. In addition, the PPDs using the analytic integration of the full Bayesian approach (Sec. III B) are identical to these using numerical integration (solid). From the simulations, the empirical Bayesian method using an ML or MAP estimate of the error variance is a good approximation to the full Bayesian approach:

$$p(\mathbf{m}|\mathbf{d}) \approx p(\mathbf{m}|\mathbf{d},\nu). \tag{39}$$

However, the difference would become more distinguishable when the number of inverted parameters is similar to the number of the data points.

Figure 6 shows the marginal PPDs for the same parameter set using 200-Hz frequency data with the additive noise corresponding to 15-dB SNR ($\nu_{\text{true}}=0.0194$) in contrast to the 20-dB SNR noise added to the 100-Hz data. Except for $c_{\text{sed,top}}$, the geoacoustic parameters are poorly estimated in comparison of Fig. 3. Since the higher frequency has higher

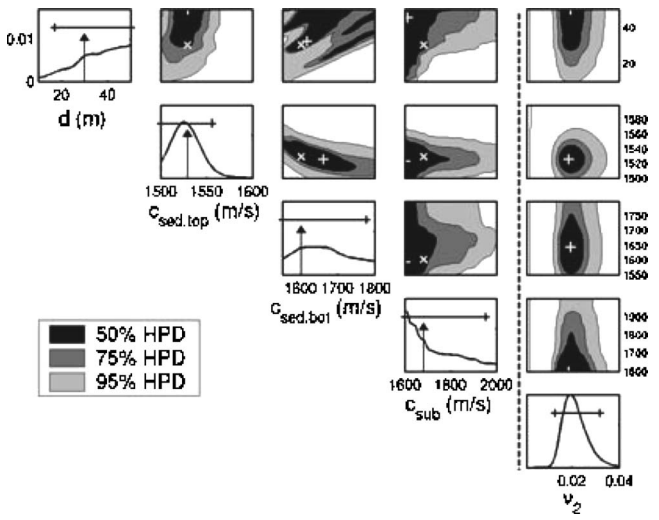

FIG. 6. Full Bayesian approach. 1-D and 2-D marginal PPDs of the geoacoustic parameters as well as the error variance for 200 Hz. The actual value of the error variance added to the data is equivalent to 15–dB SNR ($\nu^{\text{true}}=0.0194$). The format is the same as Fig 3

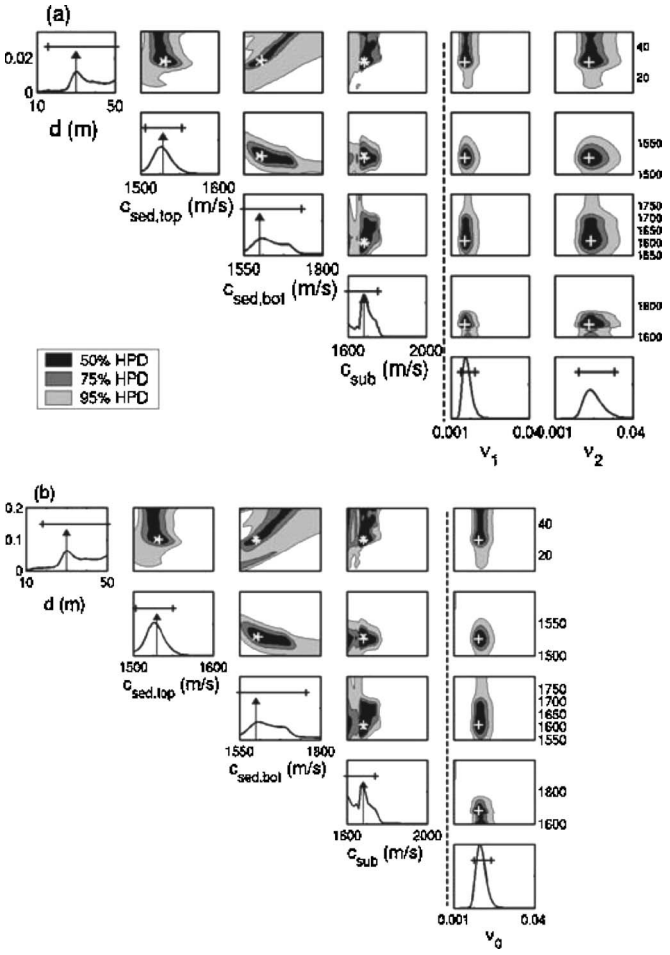

FIG. 7. Full Bayesian approach for the multi-frequency case (a) 1-D and 2-D marginal PPDs of the geoacoustic parameters as well as the error variance for frequency of 100 and 200 Hz. (b) The marginal PPDs of geoacoustic parameters as well as the global error variance. The format is the same as in Fig. 3.

resolution in the upper sediment (more structure in the acoustic field at the water-sediment, as seen in Fig. 2) but shorter penetration depth, only $c_{\text{sed,top}}$ is better resolved at 200-Hz frequency.

Having estimated the marginal PPDs using data at 100- and 200-Hz frequencies separately, we then estimate the PPDs using data from both frequencies which have different error variances. Figure 7 demonstrates the multi-frequency case: (a) the error variance are appropriately accounted for (modeled by ν_1 and ν_2) and (b) the error variance are assumed the same over these two frequencies (modeled by ν_0). In addition, the 1-D PPDs using the analytic integration of full Bayesian method, Eq. (24), are shown in both figures, where the geometric mean is used for Fig. 7(a) and the arithmetic mean for Fig. 7(b). The results are not distinguishable from those using the numerical integration.

Comparing Fig. 7(a) with Fig. 7(b), we see there is a slight difference between the two treatments of error variance. The uncertainties of the geoacoustic parameters using two frequencies (Fig. 7) are reduced significantly in contrast to those using single frequency (Figs. 3 and 6), in agreement with Refs. 1–4.

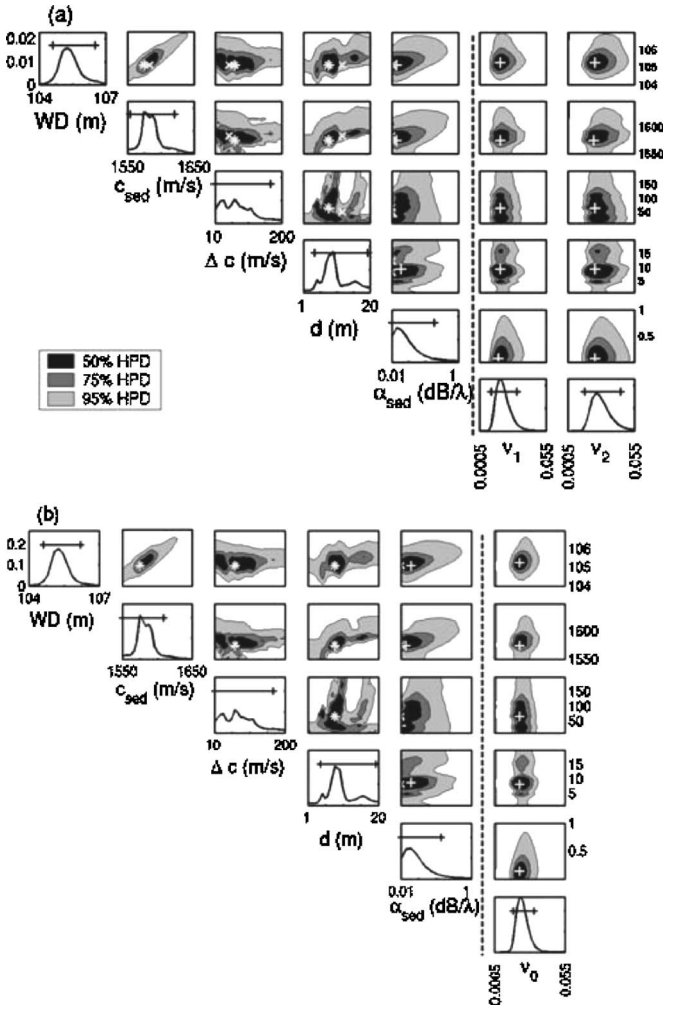

FIG. 8. Full Bayesian approach for the multi-frequency case (a) 1-D and 2-D marginal PPDs fo the geoacoustic parameters as well as the error variances for the freuqenices 195 and 395 Hz. (b) 1-D and 2-D marginal PPDs of the geoacoustic parameters as well as the global error variance. The format is the same as in Fig. 3.

B. Experimental data

Data acquired during the ASIAEX 2001 East China Sea experiment (see Ref. 24) are used to illustrate the approaches for incorporating uncertainty in the error variance. A 16-element vertical line array was deployed in 105-m-deep water (element 4 failed during deployment). The source was towed at a depth of about 48 m. A general bathymetric and geological survey has indicated that in the neighborhood of the experimental site, the environment is nearly range independent. Therefore, the ocean environment is modeled as an ocean layer overlying a uniform sediment layer atop of a basement. All layers are assumed to be range independent.

In our previous study,[24] matched-field geoacoustic inversion using the frequencies 195, 295, and 395 Hz was carried out to invert for the seafloor parameters. Based upon the GPS position of R/V *Melville*, the observed data $\mathbf{d}$ was approximately 1.7 km away from the source. To reduce mismatch in water depth, source position, array geometry, and ocean sound-speed profile, we have inverted for a total of 13 model parameters. In order to estimate the model parameters, a global optimization method, based on genetic algorithms, along

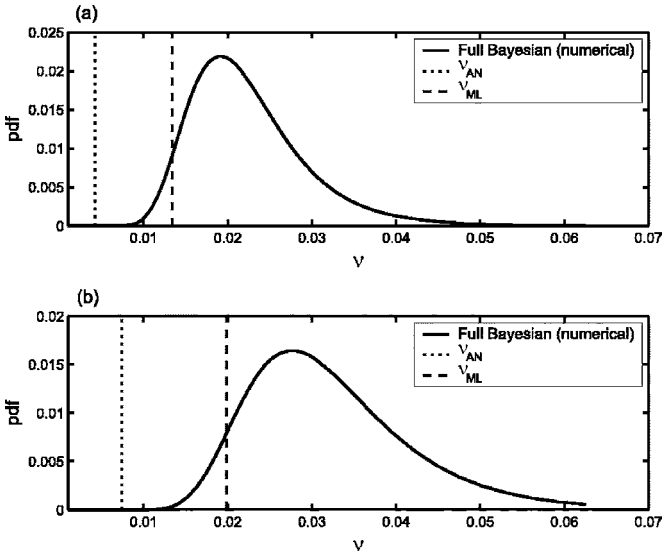

FIG. 9. Posterior marginal distribution of the error variance obtained from the numerical integration. (a) $F=195$ Hz and (b) $F=395$ Hz. The vertical lines show the different estimates of error variance.

with the normal-mode propagation model SNAP[25] was used.

We inspect here the posterior probability densities of the following five model parameters: water depth (WD), sediment sound speed (c_{sed}), basement sound speed increase (Δc), sediment thickness (d), and sediment attenuation (α_{sed}), with all other parameters fixed at their optimal values (using the empirical Bayesian treatment). The same data set is used in this analysis but with the selected frequencies of 195 and 395 Hz. Therefore, we are finding the PPDs of a total of seven parameters (the five model parameters plus the error variances at the two frequencies).

Figure 8 shows the full Bayesian approach for the multi-frequency case: (a) the frequency-dependent error variances are considered and (b) the error variances are assumed constant over frequencies. The error variances do not vary too much across the processed frequencies, therefore, the PPDs of the model parameters do not have a significant difference between the two treatments of error variances.

Figure 9 shows the posterior marginal distributions of error variance obtained from the numerical integration: (a) $F=195$ Hz and (b) $F=395$ Hz. The error estimates from the other examined approaches are superimposed on the PPD of ν. The dotted line represents the ambient noise estimated directly from the data (corresponding to SNR=23 and 21 dB, respectively) and the dashed line denotes the ML estimate of the error variance. The marginal distribution of the error variance captures that the effective error variance may be larger than the ambient noise estimate.

In the experimental data, with the high SNR, the modeling error in the parametrized environment is the dominant source of error in the estimation procedure. Because the modeling error may not be independent across the receivers, the IID assumption in the likelihood function is no longer appropriate; a full data uncertainty covariance matrix $\mathbf{C}_{\text{D}}$ is needed. Therefore, to describe the data uncertainty for N complex-valued measurements, a huge number of quantities, N^2, needs to

be estimated. A way to fix the defect of the likelihood function is to down sample the observations as adopted by Ref. 26. In the experimental data used here, due to the large separation of the array elements (5-m interelement spacing and 15 hydrophones in total), it was not necessary to down sample the data.

Finally, the comparison of PPDs for the model parameters using different error variance estimates is shown in Fig. 10. The PPD of the model parameters using the ambient noise variance (dotted) yields too optimistic an uncertainty estimate. The PPD based on the ML estimate of error variance (dash) is similar to the one obtained by the full Bayesian approach (solid). It is noteworthy that, in the 1-D marginal PPDs of WD and c_{sed}, the location of the peaks varies with different values of error variance. The reason is that there exists nonsymmetrically distributed HPD contours in the joint marginal PPD, and when using a lower value of the error variance, the lower probability density in the 2-D marginal contributes more into the 1-D marginal distribution.

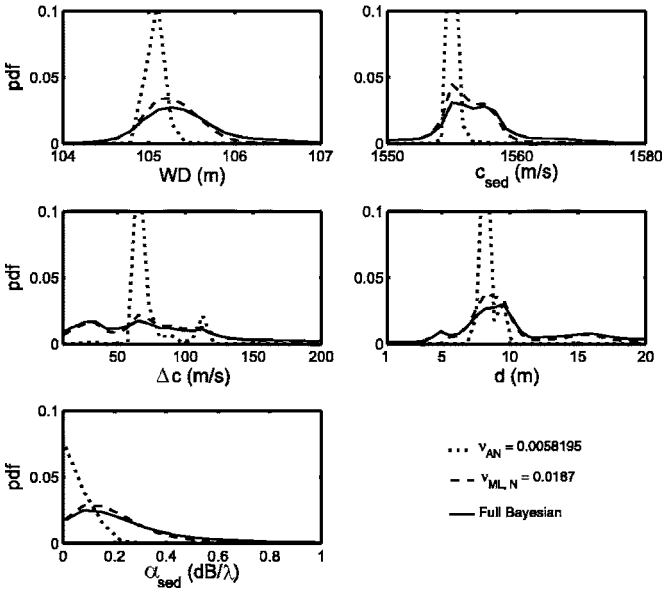

FIG. 10. Comparison of the marginal PPDs for the model parameters using different approaches in the error variance.

VI. CONCLUSIONS

This paper describes several methods for handling nuisance parameters based on both the full and empirical Bayesian approaches. In a full Bayesian approach, the inference is made from the joint posterior probability distribution (PPD) of the model parameters and the nuisance parameters, whereas in an empirical Bayesian approach the PPD of the model parameters is conditioned on a point estimate of the nuisance parameters. The full Bayesian approach takes more complete accounting of uncertainty in the nuisance parameters, but it is computationally expensive. The applications of the approaches to the error variance parameters in a Gaussian error model were examined.

Following a full Bayesian methodology, the analytic expression of PPD of the model parameters was derived for both single frequency [Eq. (23)] and multi-frequency data [Eq. (24)]. The results (Figs. 3, 6, and 7) show that the PPD of the model parameters using this analytic formula cannot be distinguished from that using numerical integration of the full Bayesian approach. Therefore, the analytic integration of the full Bayesian approach is theoretically and computationally preferred.

The analytic expression for the PPD of the error variance was derived and follows the so-called inverse chi-square distribution. This analytic result agrees well with the distribution obtained using numerical integration of the full Bayesian approach.

The empirical Bayesian approach using either the maximum likelihood or the maximum *a posteriori* estimate of the error variance was implemented. For the examples presented here, the 1-D PPDs of the model parameters using both the empirical and full Bayesian approaches yield similar results, but this is most likely not true in general.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research under Grant No. N00014-05-1-0264.

APPENDIX: ADDITIVE NOISE

For the case that the observed data are written in the form of the cross-spectral density matrix (CSDM), the noise-contaminated data with error variance ν_0 are synthesized, based on true data $\mathbf{d}$, by

$$\mathbf{R} = \mathbf{d}\mathbf{d}^{\dagger} + \nu_0 \mathbf{I}. \tag{A1}$$

The *array* signal-to-noise ratio (SNR) is the ratio of signal and noise powers:

$$\text{SNR} = 10 \log \frac{\mathbf{d}^{\dagger}\mathbf{d}}{\nu_0}. \tag{A2}$$

Equation (11) can be generalized as

$$\phi_0(\mathbf{m}) = \text{Tr}\,\mathbf{R}\left[1 - \frac{\mathbf{d}(\mathbf{m})^{\dagger}\mathbf{R}\mathbf{d}(\mathbf{m})}{\text{Tr}\,\mathbf{R}}\right]. \tag{A3}$$

Note that the replica field vector $\mathbf{d}(\mathbf{m})$ is computed from an acoustic model for the vector of unknown parameters $\mathbf{m}$ and is normalized to have unit length. If we normalize the objective function by the trace of the CSDM, denoted by Tr $\mathbf{R}$ (the total intensity of the acoustic field recorded at the receivers),

$$\phi_n(\mathbf{m}) = \frac{\phi_0(\mathbf{m})}{\text{Tr}\,\mathbf{R}}, \tag{A4}$$

the noise estimate needs to be scaled by Tr $\mathbf{R}$,

$$\nu_n = \frac{\nu_0}{\text{Tr}\,\mathbf{R}} = \frac{\nu_0}{\mathbf{d}^{\dagger}\mathbf{d} + N\nu_0}, \tag{A5}$$

and is written in terms of the *array* SNR

$$\nu_n = \frac{1}{10^{\text{SNR}/10} + N}. \tag{A6}$$

[1]J. P. Hermand and P. Gerstoft, "Inversion of broadband multi-tone acoustic data from the YELLOW SHARK summer experiment," IEEE J. Ocean. Eng. **21**, 324–346 (1996).
[2]P. Gerstoft and C. F. Mecklenbräuker, "Ocean acoustic inversion with estimation of *a posteriori* probability distributions," J. Acoust. Soc. Am. **104**, 808–819 (1998).
[3]N. R. Chapman and C. E. Lindsay, "Matched-field inversion for geoacoustic model parameters in shallow water," IEEE J. Ocean. Eng. **21**(4), 347–354 (1996).
[4]S. E. Dosso, "Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach," J. Acoust. Soc. Am. **111**, 129–142 (2002).
[5]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).
[6]C. F. Mecklenbräuker, P. Gerstoft, J. Bohme, and P. Chung, "Hypothesis testing for geoacoustic environmental models using likelihood ratio," J. Acoust. Soc. Am. **105**, 1738–1748 (1999).
[7]D. Battle, P. Gerstoft, W. S. Hodgkiss, W. A. Kuperman, and P. Nielsen, "Bayesian model selection applied to self-noise geoacoustic inversion," J. Acoust. Soc. Am. **116**, 2043–2056 (2004).
[8]B. P. Carlin and T. A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd ed. (Chapman and Hall, London, 2000).
[9]Z. -H. Michalopoulou and M. Picarelli, "A Gibbs sampling approach to maximum *a posteriori* time delay and amplitude estimation," in *Proceedings of IEEE ICASSP '02* (IEEE, New York, 2002), Vol. **3**, pp. 3001–3004.
[10]S. E. Dosso, "Probabilistic geoacoustic inversion," J. Acoust. Soc. Am. **113**, 2189–2190 (2003).
[11]C. F. Mecklenbräuker and P. Gerstoft, "Objective functions for ocean acoustic inversion derived by likelihood methods," J. Comput. Acoust. **8**, 259–270 (2000).
[12]P. Gerstoft, C. -F. Huang, and W. S. Hodgkiss, "Estimation of transmission loss in the presence of geoacoustic inversion uncertainty," IEEE J. Ocean. Eng. (in press 2005).
[13]G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis* (Addison-Wesley, Reading, MA, 1992).
[14]A. B. Baggeroer, W. A. Kuperman, and H. Schmidt, "Matched field processing: Source localization in correlated noise as an optimum estimation problem," J. Acoust. Soc. Am. **83**, 571–587 (1988).
[15]H. Jeffreys, *Theory of Probability* (Oxford U.P., Oxford, 1939).
[16]J. J. K. Ó Ruanaidh and W. J. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing*, Statistics and Computing Series (Springer, New York, 1996).
[17]The derivation of the following formula Eq. (23) requires substituting Eqs. (10) and (18) into Eq. (22) and completing the integrand by the use of the gamma integral of the form $\int_0^\infty x^{m-1}\exp(-ax)\,dx = a^{-m}\Gamma(m)$. Strictly speaking, the limits of the integration do not have to go from zero to infinity, since this prior is multiplied by a Gaussian likelihood function which dies away rapidly as $\nu \to 0$ and $\nu \to \infty$ (e.g., Refs. 16 and 27).
[18]P. Gerstoft, *SAGA Users Guide 5.0, an Inversion Software Package*, An updated version of "SAGA Users Guide 2.0, an inversion software package," SACLANT Undersea Research Centre, SM–333, La Spezia, Italy, 1997.
[19]C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers: Asymptotic Methods and Perturbation Theory* (Springer-Verlag, New York, 1999).
[20]A. Malinverno, "A Bayesian criterion for simplicity in inverse problem parametrization," Geophys. J. Int. **140**, 267–285 (2000).
[21]Note that the symbol $\hat{\mathbf{m}}$ refers to both MAP and ML estimates of $\mathbf{m}$. In the absence of prior information on $\mathbf{m}$, the $\mathbf{m}$ that maximizes the likelihood function, the ML solution, is the same as the MAP solution that maximizes the posterior pdf $p(\mathbf{m}|\mathbf{d})$.
[22]A. Tolstoy, N. R. Chapman, and G. Brooke, "Workshop'97: Benchmarking for geoacoustic inversion in shallow water," J. Comput. Acoust. **6**(1&2), 1–28 (1998).
[23]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).
[24]C. -F. Huang and W. S. Hodgkiss, "Matched field geoacoustic inversion of low frequency source tow data from the ASIAEX East China Sea experiment," IEEE J. Ocean. Eng. **29**, 952–963 (2004).
[25]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.
[26]M. J. Wilmut, S. E. Dosso, and J. Dettmer, "Data error estimation for matched-field geoacoustic inversion," J. Acoust. Soc. Am. **115**(5), 2408–2409 (2004).
[27]G. L. Bretthorst, *Bayesian Spectrum Analysis and Parameter Estimation* (Springer-Verlag, New York, 1988).