On the effect of error correlation on matched-field geoacoustic inversion

Chen-Fen Huang, Peter Gerstoft, and William S. Hodgkiss
Marine Physical Laboratory, Scripps Institution of Oceanography, 9500 Gilman Drive, La Jolla, California 92093-0238
chenfen@mpl.ucsd.edu, gerstoft@ucsd.edu, wsh@ucsd.edu

Abstract: The effect of correlated data errors on matched-field geoacoustic inversion for vertical array data is examined. The correlated errors stem from the inability to model the inhomogeneities in the environment resulting in an additional error term beyond ambient noise. Simulated data with these correlated errors are generated and then inverted with or without using the proper covariance matrix. Results show that the correlated error has a negative impact on geoacoustic parameter estimation if not accounted for properly.

PACS numbers: 43.30.Pc, 43.60.Pt [GD]
Date Received: August 13, 2006 **Date Accepted:** November 13, 2006

1. Introduction

In experimental data with high signal-to-noise ratio, error in modeling the parametrized environment is expected to be the dominant source of error in the inversion procedure. Because the modeling error may be correlated across receivers, the assumption of independent and identically distributed (IID) complex Gaussian error[1–3] in the likelihood function is no longer appropriate. Instead, a full data error covariance matrix $\mathbf{C}_D$ is needed. Some efforts have been made to estimate the error correlation, e.g., Dosso[4] assumed the errors to be ergodic in space, leading to a Toeplitz structure of $\mathbf{C}_D$.

We studied the effect of error correlation on matched-field geoacoustic inversion using simulated complex-valued pressure fields with correlated errors added. A Bayesian approach is adopted to solve the inversion where uncertainty in the inversion is described by the posterior density (PPD) of the estimated parameters. Since the primary interest is the effect of the error correlation on the geoacoustic inversion results, we assume the error covariance matrix is known and originates from uncertainty in modeling the inhomogeneities in the environment. These inhomogeneities could be due to a random environment such as internal waves (simulated here) or they could result from the chosen parametrization of the environment. For the latter case, the parametrization error must be so small that it can be modeled as a random field component. These environmental errors will cause a correlated error that presents a smooth, though random, shift in the fields.

Inversions were carried out first without, and subsequently with the full covariance matrix. In the former case, only the diagonal elements of the covariance matrix were assumed nonzero. As expected, inversions that did not take account of the full covariance matrix were poor, producing a deviation from the true solution for each environmental parameter. On the other hand, using the proper covariance matrix led to a substantially better result.

2. Theory

For matched-field geoacoustic inversion problems, the relationship between the observed complex-valued pressure field sampled at an N-element array and the predicted pressure field, at the frequency of interest, is described by the signal model

$$\mathbf{d} = \mathbf{D}(\mathbf{m}) + \mathbf{n}, \tag{1}$$

where $\mathbf{d}$ is the observed data, $\mathbf{D}(\mathbf{m})$ is the modeled data based upon a parametrized environmental model, and the vector $\mathbf{n}$ represents the error terms, that is, the residual variations in the data that are not deterministically reproducible.[1]

Let the operator $\mathcal{G}_{\text{th}}$ denote an exact forward modeling operator from model to data domain. The results of measurements that consist of the response of the physical system and the measurement errors can be described by

$$\mathbf{d} = \mathcal{G}_{\text{th}} + \mathbf{n}_{\text{obs}}, \tag{2}$$

where $\mathbf{n}_{\text{obs}}$ is an unknown residual which may result from mechanisms that are not considered in the exact theory, such as ambient noise.

Another source of errors is that we never have access to $\mathcal{G}_{\text{th}}$. Thus, the simplified parametrization with model parameter $\mathbf{m}$ of the true geoacoustic environment is the other source of error,

$$\mathcal{G}_{\text{th}} = \mathbf{D}(\mathbf{m}) + \mathbf{n}_{\text{mod}}. \tag{3}$$

Thus, from Eqs. (2) and (3), we obtain

$$\mathbf{d} = \mathbf{D}(\mathbf{m}) + \mathbf{n}_{\text{mod}} + \mathbf{n}_{\text{obs}}. \tag{4}$$

Therefore, the total error is the sum of modeling and observational errors, $\mathbf{n}=\mathbf{n}_{\text{mod}}+\mathbf{n}_{\text{obs}}$. The likelihood function $\mathcal{L}(\mathbf{m})$ is the conditional probability density function (pdf) of $\mathbf{d}$ given $\mathbf{m}$ and can be related to the pdf of the total error by

$$\mathcal{L}(\mathbf{m}) = p(\mathbf{d}|\mathbf{m}) \equiv p[\mathbf{d} - \mathbf{D}(\mathbf{m})] = p(\mathbf{n}_{\text{mod}} + \mathbf{n}_{\text{obs}}). \tag{5}$$

Suppose that $\mathbf{n}_{\text{mod}}$ and $\mathbf{n}_{\text{obs}}$ are Gaussians and statistically independent; then, the sum is also a Gaussian, in which the means and covariances add.[5] Thus, the likelihood is

$$\mathcal{L}(\mathbf{m}) = \frac{1}{\pi^N |\mathbf{C}_D|} \exp\{-[\mathbf{d} - \mathbf{D}(\mathbf{m})]^\dagger \mathbf{C}_D^{-1} [\mathbf{d} - \mathbf{D}(\mathbf{m})]\}, \tag{6}$$

where $|\cdot|$ and † denote, respectively, determinant and complex conjugate transpose. The data error covariance matrix, $\mathbf{C}_D$, is the sum of the modeling covariance matrix $\mathbf{C}_{\text{mod}}$ and observational covariance matrix $\mathbf{C}_{\text{obs}}$,

$$\mathbf{C}_D = \mathbf{C}_{\text{mod}} + \mathbf{C}_{\text{obs}}. \tag{7}$$

For a single deterministic signal, $\mathbf{D}(\mathbf{m})$ is represented by $\mathbf{w}(\mathbf{m})s$, where $\mathbf{w}(\mathbf{m})$ is the normalized signal field computed using an acoustic propagation model for a model $\mathbf{m}$ and s is the complex source signature at the frequency of interest. The maximum likelihood (ML) estimate of $\mathbf{m}$ and s is found by minimizing the exponent of Eq. (6).

We factorize the inverse of the data error covariance matrix, via the singular value decomposition expression of $\mathbf{C}_D=\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger$, as follows:

$$\mathbf{C}_D^{-1} = \mathbf{A}^\dagger \mathbf{A}, \tag{8}$$

where $\mathbf{A}=\mathbf{\Lambda}^{-1/2}\mathbf{U}^\dagger$, in which the columns of $\mathbf{U}$ are the eigenvectors of $\mathbf{C}_D$ and $\mathbf{\Lambda}$ has the corresponding eigenvalues along the main diagonal. Prewhitening $\mathbf{d}$ and $\mathbf{w}$, via the transformation: $\tilde{\mathbf{d}}=\mathbf{A}\mathbf{d}$ and $\tilde{\mathbf{w}}=\mathbf{A}\mathbf{w}$, the likelihood function can be rewritten as

$$\mathcal{L}(\mathbf{m},s) = \frac{1}{\pi^N |\mathbf{C}_D|} \exp[-(\tilde{\mathbf{d}} - \tilde{\mathbf{w}}s)^\dagger (\tilde{\mathbf{d}} - \tilde{\mathbf{w}}s)]. \tag{9}$$

The above equation has the same form as one in the IID case, except that the observed and replica vectors are replaced by their prewhitened counterparts. The ML estimate of s can be

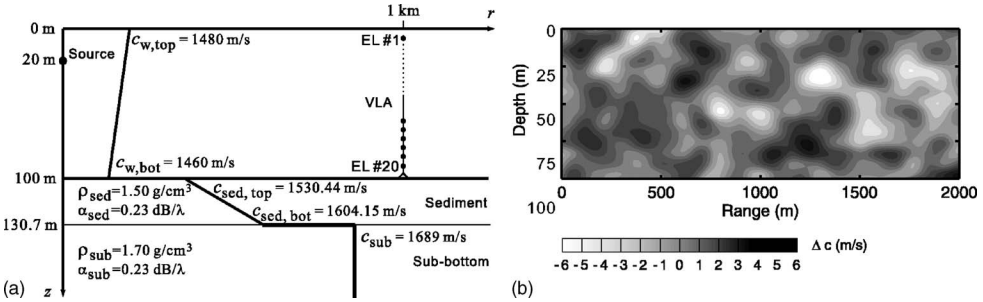

Fig. 1. (a) The sdc baseline environment from the Geo-Acoustic Inversion Workshop 1997 (Ref. 10). (b) One realization of the range-depth ocean sound-speed perturbation.

found by setting $(\partial/\partial s)\log \mathcal{L}(\mathbf{m},s)=0$, and the solution is $s_{\mathrm{ML}}=\tilde{\mathbf{w}}^{\dagger}\tilde{\mathbf{d}}/\|\tilde{\mathbf{w}}\|^2$. After substituting this relationship into Eq. (9), the following equation is obtained:

$$\mathcal{L}(\mathbf{m},\mathbf{C}_D)=\frac{1}{\pi^N|\mathbf{C}_D|}\exp[-\phi(\mathbf{m})], \tag{10}$$

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

$$\phi(\mathbf{m})=\mathbf{d}^{\dagger}\mathbf{C}_D^{-1}\mathbf{d}\left[1-\frac{|\mathbf{w}^{\dagger}\mathbf{C}_D^{-1}\mathbf{d}|^2}{(\mathbf{d}^{\dagger}\mathbf{C}_D^{-1}\mathbf{d})(\mathbf{w}^{\dagger}\mathbf{C}_D^{-1}\mathbf{w})}\right]. \tag{11}$$

The second term in the braces is a generalization of the conventional Bartlett processor. When the error terms at the receivers are assumed IID, the objective function reverts to the Bartlett power objective function[2] for then $\mathbf{C}_D$ is proportional to the identity matrix.

3. Setup: Background and perturbed environments

We use synthetic data so that the true environmental model is known in advance and can be compared with the inversion results. Figure 1(a) 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.

3.1 Modeling uncertainty

An exact theory of the geoacoustic environment (deterministic) is highly complex, and requires unrealistic demand on our ability to model the environment. Stochastic descriptions are needed because many parameters are unknown. For example, the ocean sound-speed field might be better described statistically such as by use of the internal wave spectrum. The variations between acoustic fields generated from ocean sound-speed realizations constitute the modeling uncertainty. Inspired by Gouveia and Scales,[6] the covariance matrix of data error [neglecting $\mathbf{C}_{\mathrm{obs}}$ in Eq. (7)] is assumed to be the sum of various sources of modeling error,

$$\mathbf{C}_D\approx\mathbf{C}_{\mathrm{mod}}=\mathbf{C}_{\mathrm{mod}}^{\mathrm{SSP}}+\mathbf{C}_{\mathrm{mod}}^{\mathrm{WD}}+\mathbf{C}_{\mathrm{mod}}^{\mathrm{SED}}+\cdots, \tag{12}$$

where the superscripts SSP, WD, and SED represent uncertainties not captured in the parametrized environment due to variations in ocean sound speed, water depth, and sediment sound speed, respectively. All the errors are assumed to be additive; the final $\mathbf{C}_D$ matrix is the sum of the individual covariances.

Covariances are computed for each source of error by Monte Carlo simulation,

$$\mathbf{C}_{\mathrm{mod}}=\langle(\mathbf{p}-\langle\mathbf{p}\rangle)(\mathbf{p}-\langle\mathbf{p}\rangle)^{\dagger}\rangle, \tag{13}$$

where $\mathbf{p}$ is the acoustic field from a realization of the range-dependent environment.

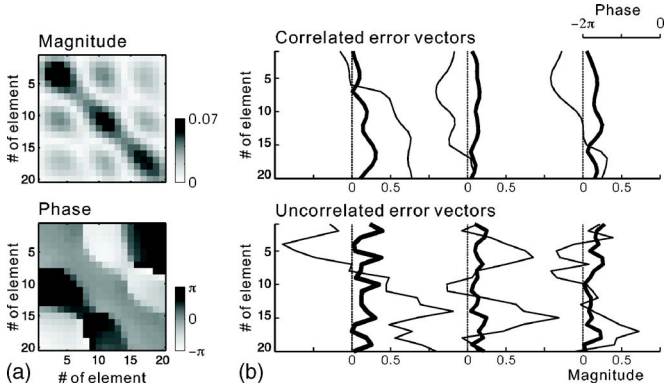

Fig. 2. (a) Magnitude (normalized by its trace) and phase for the error covariance matrix $\mathbf{C}_{\text{mod}}^{\text{SSP}}$ at the source-receiver range of 1 km for frequency of 100 Hz. (b) The correlated error vectors (top) resulting from the ocean sound-speed error for three realizations of the environment and the equivalent error vectors made IID (bottom). Heavy and light lines indicate the magnitude and phase, respectively.

To describe ocean sound-speed variability, we use the root mean square (rms) to measure the magnitude of deviation from its reference value and the correlation lengths to control the rate of change of deviation along the range (L_r) and depth (L_z). Figure 1(b) shows a realization of the range-depth ocean sound-speed perturbation, idealized from the internal wave spectrum, with an rms speed 2 m/s and a Gaussian correlation function of L_r=100 m and L_z =10 m.

3.2 Data uncertainty

Uncertainty due to a lack of knowledge of the sound speed is then mapped into the acoustic field. Data uncertainty induced from the random ocean sound speed is quantified via Eq. (13) with 1000 Monte Carlo runs. To propagate the acoustic field through the range-dependent environment, we use Collins' higher-order energy-conserving parabolic equation[7] with 2.5-m range and 0.15-m depth grid spacing, and four Padé terms. The grid spacing is verified by using a finer grid spacing with the same result obtained. The acoustic fields are simulated on a vertical array with 20 hydrophones equally spaced over 95 m with first phone at 5-m depth, and the source located at 1-km range and 20-m depth, transmitting a 100-Hz tone.

Figure 2(a) shows the structure of data error covariance resulting from ocean sound-speed modeling uncertainty. The error-free signal to modeling error power ratio (SER) averaged across the array is 15 dB. Since the error covariance matrix is Hermitian, the magnitude of the covariance matrix is symmetric (top) about its diagonal elements and the phase is antisymmetric (bottom). Note that the error covariance matrix does not have a Toeplitz structure as assumed in Ref. 4.

Correlation in errors may result in a smoother (although random) pattern in the errors across the array. The top panel of Fig. 2(b) shows three realizations of the error vector resulting from ocean sound-speed modeling error. The bottom panel shows the same total error power distributed independently and identically across the array. The patterns in the top panel appear to be smooth, are intrinsically random, and could be interpreted as a component of the signal.

4. Impact of data error correlation on inversion results

To investigate the influence of data error covariance matrix on inversion in the presence of correlated errors, we consider three approaches for $\mathbf{C}_D$ in Eq. (10).

(1) diagonal matrix with equal variances ($\nu\mathbf{I}$);
(2) diagonal matrix with unequal variances (diag($\mathbf{C}_D$));
(3) full covariance matrix ($\mathbf{C}_D$).

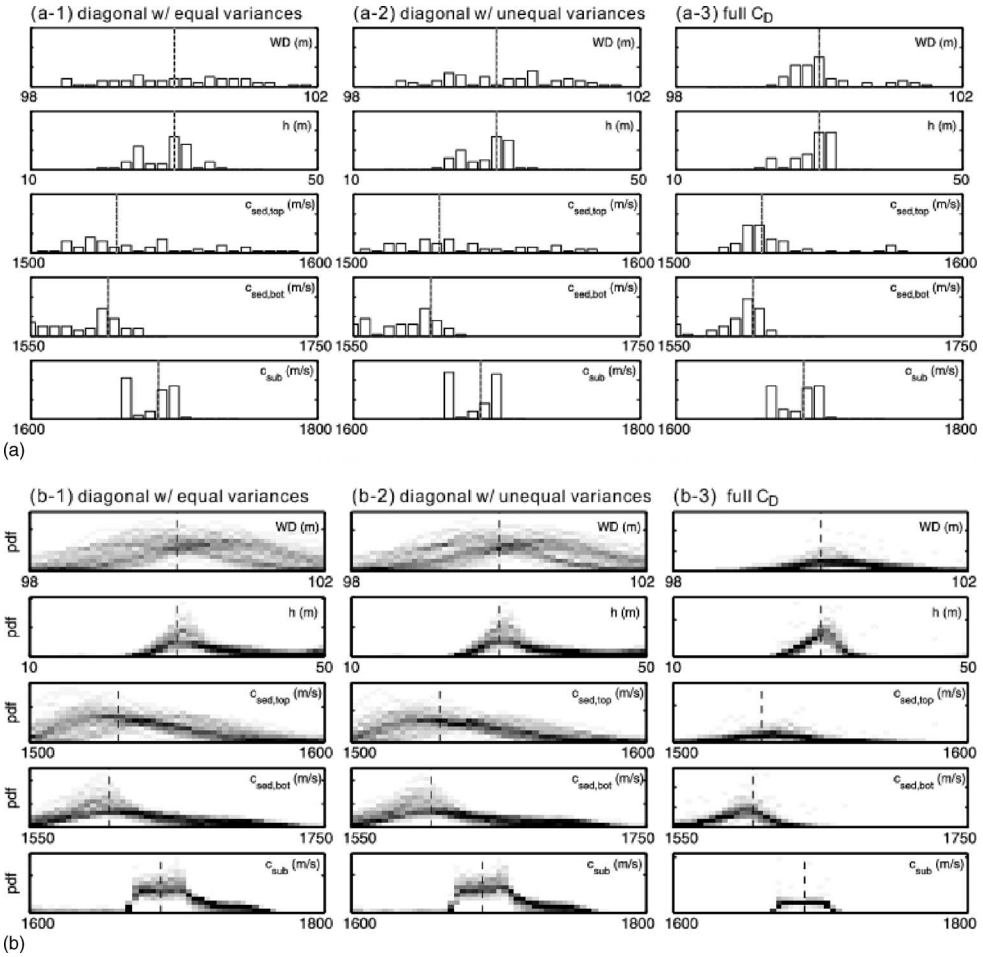

Fig. 3. (a) Histograms of the maximum *a posteriori* (MAP) estimates and (b) uncertainty of the marginal 1D posterior probability densities (PPDs) for the model parameters computed from 60 independent inversions. Vertical dotted line indicates the true value for the model parameter. The columns represent from left to right (1) diagonal $\mathbf{C}_D$ with equal variances; (2) diagonal $\mathbf{C}_D$ with unequal variances; and (3) full $\mathbf{C}_D$, respectively.

To compare these three approaches, a total of 60 independent inversions is carried out for each approach. The solution of an inversion is represented by both the maximum *a posteriori* (MAP) estimate and 1D marginal PPD for each model parameter. For each inversion, the observed data vector is generated by computing the pressure field from the baseline environment and adding a realization of the correlated error vector from the full $\mathbf{C}_D$ in Fig. 2(a), e.g., the three realizations shown in the top panel of Fig. 2(b).

The estimated parameters are water depth WD, sediment thickness h, top and bottom sediment sound speeds $c_{\text{sed,top}}$ and $c_{\text{sed,bot}}$, and subbottom sound speed c_{sub}, with the search bounds being the same as the horizontal axis for each parameter in Fig. 3. Each inversion was carried out using Markov chain Monte Carlo sampling[8] and ORCA (Ref. 9) was used for forward-field calculations.

Figure 3(a) shows histograms of the MAP estimates from all 60 inversions for each model parameter. Each column represents one of the above approaches for $\mathbf{C}_D$. Using the full $\mathbf{C}_D$ [Fig. 3(a-3)], the MAP estimates have less spread. Without including the correlation structure in the objective function [Figs. 3(a-1 and 2)], the smooth pattern of the error vector can be mistaken as a component of the signal. The inversion process treats the error as signal, resulting

in a biased MAP estimate. With full $\mathbf{C}_D$ [Fig. 3(a-3)], the bias induced by the correlated error is eliminated; the MAP estimates are more centered around the true solution.

Figure 3(b) shows the superimposed 1D marginal PPDs for each of the model parameters from the 60 independent inversions. A gray scale is used to indicate the distribution of the PPD curves. The darker color represents more curves falling into a particular bin of the grid. It displays uncertainty of the marginal 1D PPDs for each model parameter. When the full error covariance matrix is properly accounted for in the likelihood function, the parameter uncertainty derived from an observed data set is more reliable.

5. Conclusions

We have described an approach for modeling data error correlations on matched-field geoacoustic inversion for vertical array data. The data error correlations considered here originate from inhomogeneities that are not included in the environmental model. These inhomogeneities could be due to a random environment or result from the chosen parametrization of the environment. We have only simulated modeling errors resulting from our inability to model the inhomogeneities in the ocean sound-speed structure.

In the presence of error correlation, the smooth pattern of the error vector can be mistaken as a component of the signal. This influences the posterior probabilities if not taken into consideration. By including the data error covariance matrix $\mathbf{C}_D$ in the likelihood function, the impact of the smooth pattern induced by the data errors may be reduced. Consistent 1D PPDs are obtained and MAP estimates are more centered around the true solution.

Multifrequency inversions might have a mitigating effect on the geoacoustic inversion results presented here. However, the environmental perturbations might result in cross frequency correlation.

References and links

[1]P. Gerstoft and C. F. Mecklenbräuker, "Ocean acoustic inversion with estimation of *a posteriori* probability distributions," J. Acoust. Soc. Am. **104**, 808–819 (1998).
[2]C. F. Mecklenbräuker and P. Gerstoft, "Objective functions for ocean acoustic inversion derived by likelihood methods," J. Comp. Acoust. **8**, 259–270 (2000).
[3]C.-F. Huang, P. Gerstoft, and W. S. Hodgkiss, "Uncertainty analysis in matched-field geoacoustic inversions," J. Acoust. Soc. Am. **119**, 197–207 (2006).
[4]S. E. Dosso, P. L. Nielsen, and M. J. Wilmut, "Data error covariance in matched-field geoacoustic inversion," J. Acoust. Soc. Am. **119**, 208–219 (2006).
[5]A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation* (SIAM, Philadelphia, 2005).
[6]W. P. Gouveia and J. A. Scales, "Bayesian seismic waveform inversion: Parameter estimation and uncertainty analysis," J. Geophys. Res. **103**, 2759–2779 (1998).
[7]M. D. Collins and E. K. Westwood, "A higher-order energy-conserving parabolic equation for range-dependent ocean depth, sound speed, and density," J. Acoust. Soc. Am. **89**, 1068–1075 (1991).
[8]P. Gerstoft, *SAGA Users guide 5.0, an inversion software package*, An updated version of "SAGA 2.0," SACLANT Undersea Research Centre, SM-333, La Spezia, Italy (1997).
[9]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).
[10]A. Tolstoy, N. R. Chapman, and G. Brooke, "Workshop'97: Benchmarking for geoacoustic inversion in shallow water," J. Comp. Acoust. **6**, 1–28 (1998).

