24th INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
23–27 July 2017, London

ICSV24
23-27JULY 2017
LONDON CALLING

NUMERICAL APPROACHES IN GEOACOUSTIC INVERSION

Stan E. Dosso
University of Victoria, School of Earth & Ocean Sciences, Victoria, British Columbia, Canada
email: sdosso@uvic.ca

Hefeng Dong
Norwegian University of Science & Technology, Dept. of Electronic Systems, Trondheim, Norway

Kenneth Duffaut
Norwegian University of Science & Technology, Dept. of Geoscience & Petroleum, Trondheim, Norway

This paper considers numerical approaches to solve geoacoustic inverse problems to infer seabed geophysical profiles from ocean acoustic measurements. Within a Bayesian formulation, the solution is given by the posterior probability density over the seabed parameters of interest, which combines data and prior information. For nonlinear geoacoustic inversion, the posterior probability density can be sampled numerically using Markov-chain Monte Carlo methods. These methods can be computationally intensive, and a number of numerical approaches can be applied to provide efficient sampling while ensuring sufficiently wide exploration of the parameter space, including principal-component sampling and parallel tempering. A key component for efficient/effective inversion is that of choosing an appropriate parameterization for the seabed, i.e., model selection. Recent trans-dimensional inversion methods model the seabed as sequence of discontinuous, uniform layers and sample probabilistically over the number of layers. However, in some applications it is preferable to consider smooth, continuous gradients in geoacoustic properties, which can be accomplished by parameterizing geoacoustic profiles in terms of Bernstein-polynomial basis functions. These approaches are compared for the problem of estimating seabed geoacoustic profiles from the dispersion of seabed seismo-acoustic interface waves.
Keywords: Geoacoustic inversion, Bayesian inference, model selection, numerical sampling, interface-wave dispersion.

1. Introduction

The remote sensing of seabed geoacoustic properties using ocean acoustic and seismo-acoustic observations (data) is an important problem with applications in sonar, geology/geophysics, and marine geotechnical studies. The specific problem considered here involves estimating the depth-dependent shear-wave velocity profile, $v_s(z)$, from measurements of the dispersion (phase velocity as a function of frequency) of seismo-acoustic interface or Scholte waves [1]–[6], which requires solving a nonlinear inverse problem. Although geoacoustic inverse problems problems such as this were traditionally solved via linearization, in recent years fully-nonlinear (numerical) Bayesian inversion methods based on Markov-chain Monte Carlo (MCMC) sampling [7, 8, 9] have been applied. Important components of a complete solution to an inverse problem include an objective approach to model selection (e.g., choosing an appropriate seabed parameterization) and parameter inference (estimating uncertainties of the model parameters). Approaches to model selection considered here include the Bayesian information criterion (BIC), and trans-dimensional (trans-D) sampling over a set of possible models

(choices of parameterization) [10, 11]. Efficient and effective sampling is of key importance in numerical inversion, and methods such as principal-component sampling [11] and parallel tempering [12, 13] are described.

Regarding model parameterizations, it is well known from both theoretical considerations and observations that the shear-wave velocity profile in unconsolidated marine sediments of a uniform composition generally varies in terms of a smooth, continuous gradient, rather than in discontinuous layers, and often approximates a power law [3, 4, 5], such that inversions are sometimes carried out directly for the parameters of a power-law relationship. However, parameterizing the model in terms of a power law represents strong prior information which may not always be justified. This approach does not provide independent verification that the sediment profile actually follows a power law, and can lead to poor results if it does not. The goal of this paper is to compare the results of nonlinear inversion based on a power-law profile to more general approaches to model selection for parameterization. These include trans-D inversion based on an unknown number of uniform layers [10, 11],and a new parameterization in terms of Bernstein-polynomial (BP) basis functions [14] which provides general gradient models.

2. Inverse theory and algorithms

This section provides a brief overview of a nonlinear (numerical) approach to Bayesian inference and model selection which is applicable to geoacoustic inversion; more complete treatments are given in [7]–[16].

2.1 Nonlinear Bayesian inference

Let $\mathcal{M}$ represent the model of a system of interest (e.g., the seabed), including the parameterization in terms M unknown parameters, and let $\mathbf{m}$ be a vector of a parameter values, assumed to be random variables, constrained by a vector of N data $\mathbf{d}$ and prior information $P(\mathbf{m}|\mathcal{M})$. Bayes' theorem states

$$P(\mathbf{m}|\mathbf{d},\mathcal{M}) \;=\; \frac{P(\mathbf{m}|\mathcal{M})\,P(\mathbf{d}|\mathbf{m},\mathcal{M})}{P(\mathbf{d}|\mathcal{M})}. \tag{1}$$

In Eq. (1), $P(\mathbf{d}|\mathbf{m},\mathcal{M})$ is the conditional probability of $\mathbf{d}$ given model $\mathcal{M}$ and parameters $\mathbf{m}$, and represents the data information. Interpreted as a function of $\mathbf{d}$, this term represents the residual error density. However, when $\mathbf{d}$ represents the (fixed) observed data, the term is interpreted as the likelihood of $\mathbf{m}$. For example, assuming Gaussian-distributed errors of covariance matrix $\mathbf{C}_d$, the likelihood is

$$\mathcal{L}(\mathbf{m}) = \frac{1}{(2\pi)^{N/2}|\mathbf{C}_d|^{1/2}}\,\exp\left[-\left(\mathbf{d}-\mathbf{d}(\mathbf{m})\right)\,\mathbf{C}_d^{-1}\left(\mathbf{d}-\mathbf{d}(\mathbf{m})\right)/2\right], \tag{2}$$

where $\mathbf{d}(\mathbf{m})$ represents data predicted for model parameters $\mathbf{m}$. On the left side of Eq. (1), $P(\mathbf{m}|\mathbf{d})$ is the posterior probability density (PPD), representing the total information for the model parameters given the data, prior, and choice of model. The normalization term $P(\mathbf{d}|\mathcal{M})$ represents the probability of the data given the choice of model (independent of $\mathbf{m}$), referred to as the Bayesian evidence for $\mathcal{M}$ (the evidence can be considered the likelihood of the $\mathcal{M}$ [15]).

For nonlinear inverse problems the PPD can be estimated numerically via MCMC sampling methods. Metropolis-Hastings sampling consists of generating new model parameters $\mathbf{m}'$ via a proposal density based only on the current model, $Q(\mathbf{m}'|\mathbf{m})$, and accepting the proposed model as the next step in the Markov chain with acceptance probability [8]

$$A(\mathbf{m}'|\mathbf{m}) = \min\left[1, \frac{Q(\mathbf{m}|\mathbf{m}')}{Q(\mathbf{m}'|\mathbf{m})}\,\frac{P(\mathbf{m}')}{P(\mathbf{m})}\,\frac{\mathcal{L}(\mathbf{m}')}{\mathcal{L}(\mathbf{m})}\right]. \tag{3}$$

The Metropolis-Hastings criterion is applied by drawing a random number ξ from a uniform distribution on $[0, 1]$ and accepting the new model $\mathbf{m}'$ if $\xi < A(\mathbf{m}'|\mathbf{m})$. If $\mathbf{m}'$ is not accepted, another copy of the current model $\mathbf{m}$ is included as the next step in the chain.

While MCMC provides PPD sampling for nonlinear inversion, it can be numerically intensive, and it is important to develop efficient algorithms. The choice of proposal density is of key importance in an efficient algorithm. The optimal proposal density is given by the PPD itself; in this case all terms cancel in the acceptance criterion such that all proposals accepted. Although the PPD is not available as a proposal density in practical problems, a local linearized approximation can be used to define an efficient proposal scheme. According to standard linearized inverse theory [16], the PPD can be approximated by an M-dimensional Gaussian distribution with posterior model covariance matrix

$$\mathbf{C}_m = \left[\mathbf{J}^\top \mathbf{C}_d^{-1}\, \mathbf{J} + \mathbf{C}_p^{-1}\right]^{-1}, \tag{4}$$

where $\mathbf{J}$ is the Jacobian matrix of partial derivatives, $J_{ij} = \partial d_i(\hat{\mathbf{m}})/\partial m_j$, $\mathbf{C}_d$ is the data covariance matrix, and $\mathbf{C}_p$ is the prior model covariance matrix of an assumed Gaussian prior density. To draw individual parameters but take advantage of the full M-dimensional form of the Gaussian proposal, perturbations are applied in a principal-component (PC) parameter space where the axes align with the dominant correlation directions (i.e., PC parameters are uncorrelated). The orthogonal transformation (rotation) between physical parameters $\mathbf{m}$ and PC parameters $\tilde{\mathbf{m}}$ is

$$\tilde{\mathbf{m}} = \mathbf{U}^\top \mathbf{m}, \qquad \mathbf{m} = \mathbf{U}\, \tilde{\mathbf{m}}, \tag{5}$$

where $\mathbf{U}$ is the column-eigenvector matrix of the model covariance matrix,

$$\mathbf{C}_m = \mathbf{U}\, \mathbf{W}\, \mathbf{U}^\top, \tag{6}$$

and $\mathbf{W} = \mathrm{diag}[w_i]$ is the eigenvalue matrix, with w_i representing the PC parameter variances. Thus, the PC decomposition provides both directions and length scales for effective (Gaussian-distributed) parameter proposals [11]. The procedure used here is initiated using the linearized model covariance estimate given by Eq. (4), with partial derivatives computed numerically. Uniform bounded priors of width Δm_i are approximated by taking $\mathbf{C}_p$ to be a diagonal matrix with variances equal to those of the uniform distributions, i.e., $(\Delta m_i)^2/12$. Following this initialization, nonlinear estimation of the model covariance matrix is carried out numerically based on the ongoing MCMC sampling, with $\mathbf{C}_m$ updated periodically (a diminishing adaptation).

The method of parallel tempering [12, 13] is a powerful approach for efficient/effective sampling based on running a series of parallel, interacting Markov chains for which the acceptance criterion is relaxed by raising the likelihood to powers $1/T$ (where $T \geq 1$ is referred to as temperature). High-T chains have an increased probability of accepting low-likelihood models, and hence provide a wider sampling of the parameter space with increased probability of accepting large moves in the parameter space, potentially bridging isolated modes. Conversely, low-T chains provide concentrated sampling but are prone to become trapped in localized regions of the space. Parallel tempering improves sampling by providing probabilistic interchange between chains with different temperatures, ensuring that low-T chains can access all regions of the space, providing a robust ensemble sampler. For randomly-chosen chains and uniform prior densities the acceptance probability of interchange between chains i and j is

$$A((\mathbf{m}_j, T_i), (\mathbf{m}_i, Tj)|(\mathbf{m}_i, T_i), (\mathbf{m}_j, T_j)) = \min\left[1, \left\{\frac{\mathcal{L}(\mathbf{m}_i)}{\mathcal{L}(\mathbf{m}_j)}\right\}^{(1/T_j - 1/T_i)}\right]. \tag{7}$$

Since chains at $T > 1$ provide biased sampling of the PPD, only the samples collected at $T = 1$ are retained. Combining PC sampling and parallel tempering, a separate PC decomposition/rotation is carried out based on the sampling for each chain initiated from an estimate

$$\mathbf{C}_m(T) = \left[\mathbf{J}^\top \left(T\, \mathbf{C}_d\right)^{-1} \mathbf{J} + \mathbf{C}_p^{-1}\right]^{-1}. \tag{8}$$

2.2 Model selection

Objectively determining the best choice of model (e.g., parameterization of the seabed) is a problem of fundamental importance. One approach to model selection is to adopt the model $\mathcal{M}$ which maximizes the Bayesian evidence. However, evidence computation require numerical estimation of a particularly challenging integral [15],

$$P(\mathbf{d}|\mathcal{M}) = \int P(\mathbf{m}'|\mathcal{M})\, P(\mathbf{d}|\mathbf{m}', \mathcal{M})\, d\mathbf{m}'. \tag{9}$$

Alternatively, the BIC, which represents an asymptotic point estimate of evidence, can be applied:

$$-2 \log_e P(\mathbf{d}|\mathcal{M}) \approx \mathrm{BIC} = -2\, \log_e L(\hat{\mathbf{m}}) + M \log_e N, \tag{10}$$

where $\hat{\mathbf{m}}$ is the maximum-likelihood set of parameter values for a particular choice of $\mathcal{M}$. Since the BIC is defined in terms of the negative logarithm of evidence, the goal is to determine the model that minimizes the BIC. The first term on the right of Eq. (10) favors models with low misfits; however, this is balanced by the second term which applies a penalty for additional parameters. Minimizing the BIC provides the model with the smallest number of parameters required to fit the data, prevents over-parameterizing the model, and provides the preferred solution according to Occam's razor.

Another approach to model selection is trans-D inversion [7, 10, 11], which provides an ensemble solution over a set of possible model parameterizations, sampled according to their probability. Let $\mathcal{K}$ be a set, indexed by k, specifying parameterization choices with M_k parameters denoted $\mathbf{m}_k$. Bayes' theorem for a hierarchical model including hyper-parameter k can be written

$$P(k, \mathbf{m}_k|\mathbf{d}) = \frac{P(k)\, P(\mathbf{m}_k|k)\, P(\mathbf{d}|k, \mathbf{m}_k)}{\sum_{k' \in \mathcal{K}} \int P(k')\, P(\mathbf{m}'_{k'}|k')\, P(\mathbf{d}|k', \mathbf{m}'_{k'})\, d\mathbf{m}'_{k'}}. \tag{11}$$

In (11),$P(k)P(\mathbf{m}_k|k)$ is the prior probability of the state $(k, \mathbf{m}_k)$, and $P(\mathbf{d}|k, \mathbf{m}_k)$ is the conditional probability of $\mathbf{d}$ given $(k, \mathbf{m}_k)$, which is interpreted as the likelihood $\mathcal{L}(k, \mathbf{m}_k)$. The PPD $P(k, \mathbf{m}_k|\mathbf{d})$ is defined over the trans-D parameter space spanning all choices of parameterization.

To sample the trans-D parameter space in (11), a Markov chain must transition between models with differing numbers of parameters. For such steps, acceptance is generalized to the Metropolis-Hastings-Green criterion [7]

$$A(k', \mathbf{m}'_{k'}|k, \mathbf{m}_k) = \min \left[1, \frac{Q(k, \mathbf{m}_k|k', \mathbf{m}'_{k'})}{Q(k', \mathbf{m}'_{k'}|k, \mathbf{m}_k)} \frac{P(k')P(\mathbf{m}'_{k'}|k')}{P(k)\, P(\mathbf{m}_k|k)} \frac{\mathcal{L}(k', \mathbf{m}'_{k'})}{\mathcal{L}(k, \mathbf{m}_k)} |\mathbf{J}| \right], \tag{12}$$

where $|\mathbf{J}|$ is the determinant of the Jacobian matrix for the transition from $(k, \mathbf{m}_k)$ to $(k', \mathbf{m}'_k)$. Trans-D sampling according to (12) is referred to as reversible jump Markov-chain Monte Carlo (rjMCMC). In geoacoustic inversion, rjMCMC adds and deletes layer interfaces (referred to as birth and death steps, respectively) in such a manner that $|\mathbf{J}| = 1$. The requirements are that the number of interfaces in the model changes by one at a step; that interface depths are independent from layer parameters; and that only the parameters added/deleted in birth/death steps are changed, with proposed parameters in a birth step depending only on the parameters at that depth in the current state. PC proposal densities and parallel tempering have been shown to substantially improve efficiency in trans-D inversion [11].

Trans-D inversion is an automated scheme which has the advantage that the uncertainty in the parameterization is inherently included in the parameter uncertainty estimates. However, as trans-D inversion is typically parameterized in terms of an unknown number of discontinuous, uniform layers, it may not be appropriate in applications where the model (seabed) is known to involve smooth, continuous gradients. A recent advance in Bayesian geoacoustic inversion for gradient-based models is

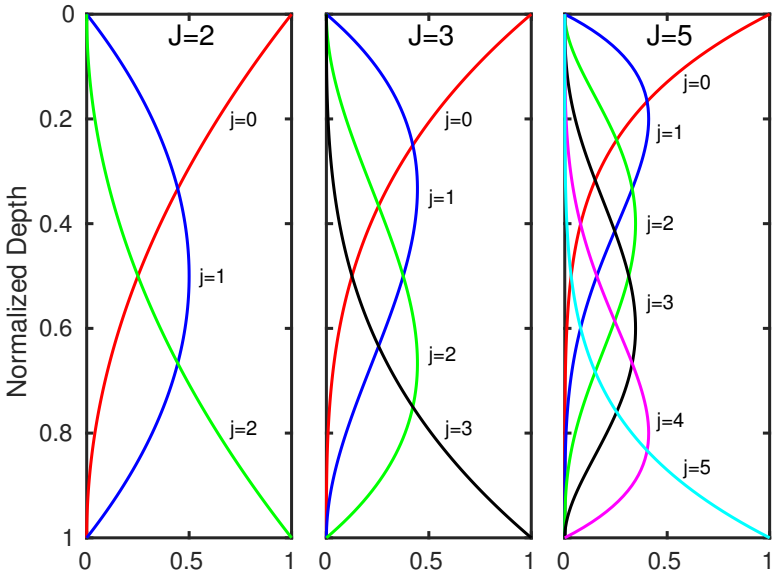

Figure 1: Bernstein-polynomial basis functions versus normalized depth $\tilde{z}$ for orders $J = 2,\ 3,\ 5$.

the parameterization in terms of BP basis functions [14]. In this approach, the profile of a geoacoustic parameter u over some depth range $z \in [0, z_{\max}]$ is represented as a BP

$$u(\tilde{z}) = \sum_{j=0}^{J} g_j\, b_j(\tilde{z}, J), \tag{13}$$

where $\tilde{z} = z/z_{\max}$ is normalized depth, $\{g_j, j = 0, \ldots, J\}$ are a set of $J+1$ coefficients (unknown parameters to be determined in the inversion), and

$$b_j(\tilde{z}, J) = \binom{J}{j} (1 - \tilde{z})^{J-j}\, \tilde{z}^j \tag{14}$$

are the corresponding set of BP basis functions, illustrated in Fig. 1 for orders $J = 2$, 3, and 5. The BP basis functions vary smoothly, are localized with peaks at successively greater (normalized) depths, and sum to unity at all depths. Further, BPs are stable in the sense that perturbing a coefficient (model parameter) only effects the profile over a limited depth range, a desirable property for non-linear inversions based on perturbing/accepting models. If BPs are chosen to represent geoacoustic parameter gradients, choosing the order (J) represents a secondary model-selection decision which can be carried out objectively using the BIC.

3. Interface-wave dispersion inversion

This section illustrates Bayesian inference and model selection for the problem of estimating the seabed shear-wave velocity profile, $v_s(z)$, from inversion of seismo-acoustic interface-wave dispersion data [1]–[6]. This geoacoustic inverse problem has been considered previously using approaches such as linearized inversion (with uncertainty estimation degraded by linearization errors and subjective regularization), and nonlinear inversions including inversions for specific profile shapes (e.g., power-law profiles), layered models, and trans-D inversion.

Fig. 2 shows noisy synthetic dispersion data generated for two shear-wave velocity profiles: Fig. 2(a) for a power-law profile, and Figure 2(b) for a linear gradient. In each case $v_s(z)$ increases smoothly from 0 to 20 m depth, with a discontinuous jump to a higher velocity in the basement half-space (true models are shown in Figs. 3 and 4, respectively). The data include include independent Gaussian errors of standard deviation 1 m/s, which are representative of measured dispersion errors.

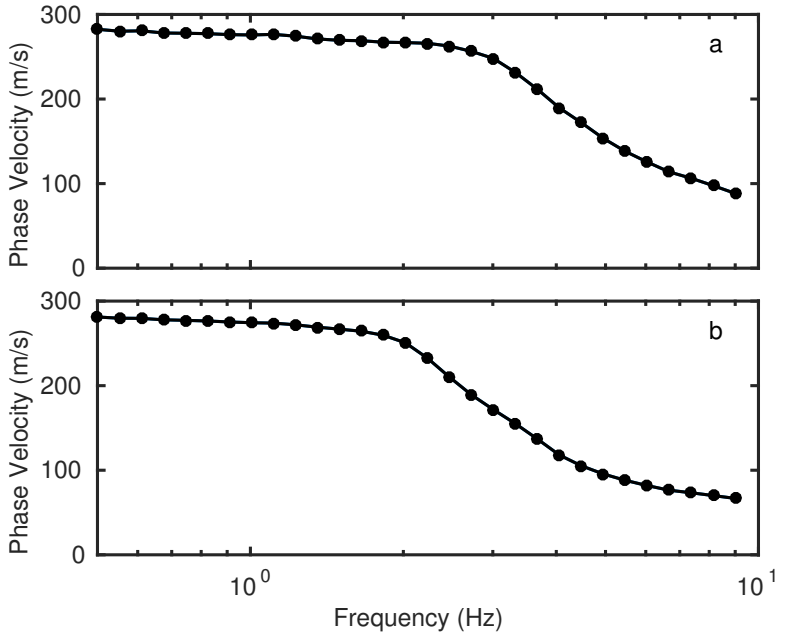

Figure 2: Noisy synthetic interface-wave dispersion data sets considered in this paper: (a) corresponds to a power-law shear-wave velocity profile; (b) to a linear profile.

Fig. 3 show inversion results for the first (power-law) data set considering three different model parameterizations, including inverting directly for the parameters defining a power law, trans-D inversion, and BP inversion. The power-law inversion provides excellent results in terms of both the depth to basement and the ability to resolve the profile shape. Uncertainties on shear-wave velocity generally increase with depth over the 20 m gradient, and the basement depth and velocity are well determined. In this case, including knowledge that the profile corresponds to a power law in the model selection provides strong prior information. However, in many practical situations definitive prior knowledge of the profile shape is not available; the other two inversion approaches in Fig. 3 do not re-

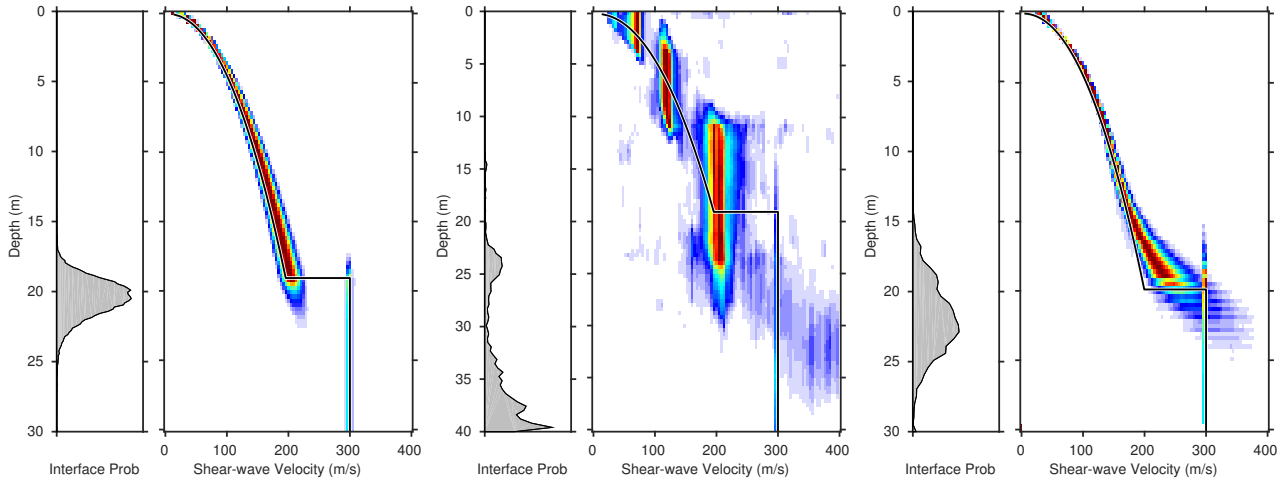

Figure 3: Marginal posterior probability profiles from noisy synthetic data generated for a power-law shear-wave v_s profile (solid line) based on: inversion for a power-law model (left two panels), trans-D inversion (centre two panels), and BP inversion (right two panels). For each set of two panels, the left panel shows the marginal probability density for the basement interface depth, and the right panel shows the marginal probability density of $v_s(z)$ (hot colours are high probabilities, and each depth is normalized independently for display purposes).

quire such information and represent more general approaches to model selection and inversion. The trans-D inversion results in Fig. 3 (with 1-10 interfaces allowed) captures the correct trend of increasing shear-wave velocity with depth, but not the profile shape. Rather, the profile consists of a series of three approximately-uniform layers, separated by discontinuities, over the basement. Basement velocity, but not depth, is well determined. The trans-D inversion fit the data to within uncertainties, but the structural form does not correspond to prior knowledge of a smooth, gradient-based profile. BP inversion results (polynomial order $J = 3$ determined via the BIC) provide a good approximation to the power-law profile shape and to the basement depth and velocity. Compared to the power-law inversion, the BP profile slightly under-estimates the strong profile curvature at very shallow depths (which the data have little sensitivity to) and has higher uncertainties below 15 m depth where the data can be fit by either a stronger velocity gradient or a transition to the half-space. Overall, BP inversion provides excellent results, given that no prior knowledge of a power-law profile shape is applied (i.e., the inversion is general).

Fig. 4 shows inversion results for the linear-gradient data set (Fig. 2(b)). In this case, the inversion based on a power-law parameterization produces good results in terms of $v_s(z)$ from about 2–12 m depth, but diverges from the true model at shallower and deeper depths due to curvature of the profile. This also leads to a transition to the half-space at about 15 m depth which is erroneous but estimated with high probability due to model mismatch. The trans-D inversion result again provides a reasonable approximation to the true profile in terms of discontinuous layers, but is highly uncertain on basement depth. Finally, the BP inversion ($J = 3$ from BIC) provides an excellent approximation to the true linear gradient with uncertainty increasing with depth, particularly near the half-space transition depth (which is relatively uncertain but peaked near the true value).

4. Summary

This paper considered Bayesian inference and model selection for the geoacoustic inverse problem of estimating the seabed shear-wave velocity profile from the dispersion of seismo-acoustic interface waves. Marginal posterior probability profiles were determined by Markov-chain Monte Carlo numerical sampling methods. Three approaches to model selection were considered, including assuming the shear-wave velocity profile is a power law, trans-dimensional inversion, and inversion in terms of a Bernstein-polynomial expansion. The two (synthetic) examples showed that BP inversion is well suited to geoacoustic inversion where a smooth gradient in seabed structure is expected but the profile shape is not known *a priori*.

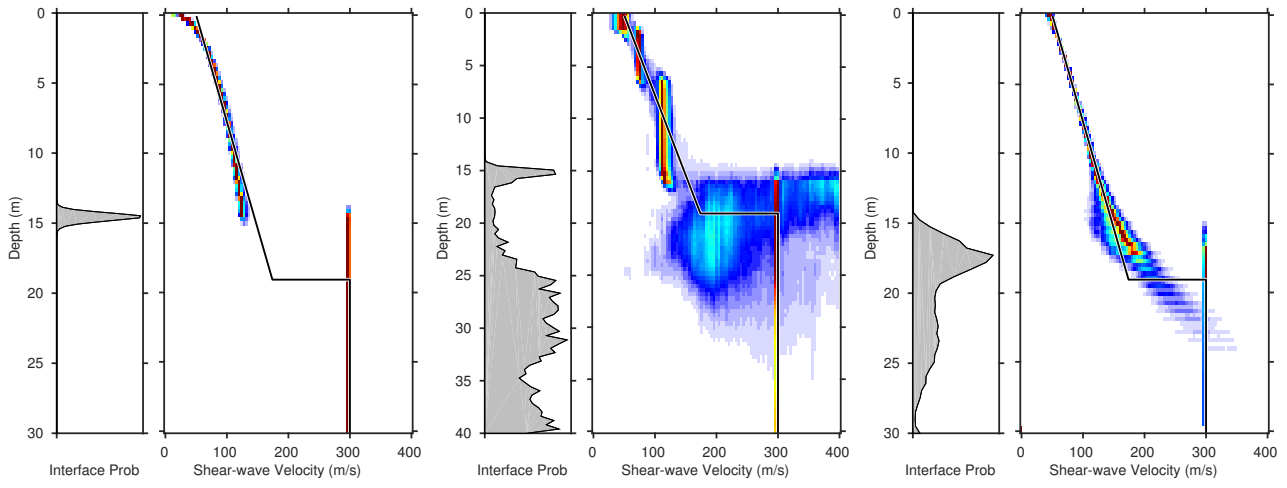

Figure 4: Marginal posterior probability profiles from noisy synthetic data generated for a linear-gradient shear-wave velocity profile (solid line) based on: inversion for a power-law model (left two panels), trans-D inversion (centre two panels), and BP inversion (right two panels).

REFERENCES

1. Jensen, F. B. and Schmidt, H. Shear properties of ocean sediments determined from numerical modeling of Scholte waves data, in *Ocean Seismo-Acoustics*, Ed: T. Akal and J. M. Berkson, Plenum, New York (1985).

2. Caiti, A., Akal, T. and Stoll, R. D. Estimation of shear wave velocity in shallow marine sediments, *IEEE J. Ocean. Eng.*, **19**, 58–72 (1994).

3. Godin, O. A. and Chapman, D. M. F. Dispersion of interface waves in sediments with power-law shear speed profiles I: Exact and approximate analytic results, *J. Acoust. Soc. Am.*, **110**,1890–1907 (2001).

4. Chapman, D. M. F. and Godin, O. A. Dispersion of interface waves in sediments with power-law shear speed profiles II: Experimental observations and seismo-acoustic inversions, *J. Acoust. Soc. Am.*, **110** 1908-1916 (2001).

5. Dong, H. and Dosso, S. E. Bayesian inversion of interface-wave dispersion for seabed shear-wave velocity profiles, *IEEE J. Oceanic Eng.*, **36**, 1–11 (2011).

6. Li, C. E., Dosso, S. E., Dong, H., Li L. and Yu D. Bayesian inversion of multi-mode interface-wave dispersion from ambient noise, *IEEE J. Oceanic Eng.*, **37**, 407–416 (2012).

7. Green, P. J. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, **82**, 711–732 (1995).

8. Gilks, W. R., Richardson, S. and Spiegelhalter, G. J. *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London (1996).

9. Sambridge, M. and Mosegard, K. Monte Carlo methods in geophysical inverse problems, *Reviews of Geophysics*, **40**, 3-1–3-29 (2002).

10. Dettmer, J., Dosso, S. E. and Holland, C. W. Trans-dimensional geoacoustic inversion, *J. Acoust. Soc. Am.*, **128**, 3393–3405 (2010).

11. Dosso, S. E., Dettmer, J., Steininger, G. and Holland, C. W. Efficient trans-dimensional Bayesian inversion for geoacoustic profile estimation. *Inverse Problems*, **30**, 114018, 29 pp. (2014).

12. Earl, D. J. and Deem, M. W. Parallel tempering: Theory, applications, and new perspectives, *Phys. Chem. Chem. Phys.*, **7**, 3910–3916 (2005).

13. Dosso, S. E., Holland, C. W. and Sambridge, M. Parallel tempering in strongly nonlinear geoacoustic inversion, *J. Acoust. Soc. Am.*, **132**, 3030–3040 (2012).

14. Quijano, J., Dosso, S. E., Dettmer, J. and Holland, C. W. Geoacoustic inversion for the shallow transition layer in mud sediments using a Bernstein-polynomial parameterization, *J. Acoust. Soc. Am.*, **140**, 4073–84 (2016).

15. Dettmer, J., Dosso, S. E. and Osler, J. C. Bayesian evidence computation for model selection in nonlinear geoacoustic inference problems, *J. Acoust. Soc. Am.*, **128**, 3406–3415 (2010).

16. Aster, R. C., Borchers, B. and Thurber, C. H. *Parameter Estimation and Inverse Problems*, International Geophysics Series, Elsevier, Amsterdam (2005).