Varying dimensional Bayesian acoustic waveform inversion for 1D semi-infinite heterogeneous media

S. Esmailzadeh [a], Z. Medina-Cetina [a,*], J.W. Kang [b], L.F. Kallivokas [c]

[a] Zachry Department of Civil Engineering, Texas A&M University, College Station, TX 77843, USA
[b] Department of Civil Engineering, Hongik University, Seoul, South Korea
[c] Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, Austin, TX 78712, USA

ARTICLE INFO

ABSTRACT

This paper introduces a methodology to infer the spatial variation of the acoustic characteristics of a 1D vertical elastic heterogeneous earth model via a Bayesian calibration approach, given a prescribed sequence of loading and the corresponding time history response registered at the ground level. This involves solving an inverse problem that maps the ground seismic response onto a random profile of the ground stratigraphy (i.e. a 1D continuous spatial random field). From a Bayesian point of view, the solution to an inverse problem is fully characterized by a posterior density function of the forward model random parameters, which explicitly overcomes the solution's non-uniqueness. This subsurface earth model is parameterized using a Bayesian partition model, where the number of soil layers, the location of the layers' interfaces, and their corresponding mechanical characteristics are defined as random variables. The partition model approach to an inverse problem is closely related to a Bayesian model selection problem, where the likely dimensionality of the inverse problem (number of unknowns) is inferred conditioned on the experimental observations. The main benefit of the proposed approach is that the explicit regularization of the inverted profile by global damping procedures is not required. A Reversible Jump Markov Chain Monte Carlo (RJMCMC) algorithm is used to sample the target posterior of varying dimension, dependent on the number of layers. A synthetic case study is provided to indicate the applicability of the proposed methodology.

1. Introduction

A subsurface earth model is composed of complex geophysical formations, which embodies a wide range of physical and mechanical heterogeneities. The aim of probabilistic inverse modeling is to reconstruct the random field structure of these subsurface properties, while accounting for various sources of uncertainty stemming from ground seismic geologic measurement errors, aleatory formations, and limited theoretical understanding about underground wave propagation.

In practice, one of the main goals of geophysical investigations is to identify the main geomorphological features of an unknown medium, meaning the spatial location and concentration of geological features such as the transition between materials, discontinuities and material concentrations [34]. In the case of a vertical 1D profile, this requires the definition of the location of the sharp transitions between layer interfaces (i.e. material properties), and the characterization of its corresponding mechanical properties.

In a horizontally stratified earth model, prior to making an inference about the likely variation of the elastic parameters within the geological layers, an assumption must be made concerning the number of layers in a certain depth range of interest. This assumption defines the dimensionality (i.e., the number of unknowns) of the inverse problem. In reality, however, such information is rarely available for the dimension and definition of the parameter space to be fixed. Consequently, fixing the number of layers based on an incorrect assumption results in an erroneous subsurface characterization.

To relax the hypothesis about the subsurface structure or spatial layering of the media's mechanical parameters, before the forward model is calibrated, it is proposed to define the number of layers, their locations, and their corresponding mechanical parameters, as random variables. From a Bayesian perspective, this set up is closely associated with probabilistic model selection, where a collection of models with varying number of parameters are

* Corresponding author.
E-mail addresses: saba61@tamu.edu (S. Esmailzadeh),
zenon@tamu.edu (Z. Medina-Cetina).

presented for inversion, and the task is to select the models that most likely describe the experimental observations.

To illustrate the applicability of the proposed probabilistic calibration method, a one dimensional horizontally stratified medium is presented in terms of a Bayesian partition model [8]. The partition model divides the unknown material field into a number of non-overlapping regions, where each region represents a soil layer. Formulating the inverse medium problem in terms of a partition model may help reduce the dimensionality of the parameter space. Hence, regularizing the solution through specific prior distributions, which bears smoothness constraints (in a Bayesian inversion framework [45,12,20]), or regularization terms (in deterministic optimization problems [44,36,15]), is precluded.

A generalization of the simulation-based Markov Chain Monte Carlo methods, the so-called reversible jump [18], is used to sample the posterior distribution of varying dimensionality. In this setting, the Markov Chain is capable of undergoing dimension changes while moving among a number of candidate models. The key aspect of the reversible jump algorithm is the introduction of some auxiliary random variables to equalize the dimensionality of the parameter space across models. A series of one-to-one deterministic functions are defined to perform dimension matching such that the detailed balance condition is satisfied. Balance condition is the necessary condition for a Markov Chain to converge to the target density [7].

Since the introduction of Bayesian inference methods to the geophysical community, this has received a great deal of attention in a variety of applications [14,16,17,45,41,43]. However, a limited number of studies have addressed the subsurface parameter estimation as a model selection problem, many of which resort to approximate methods to fulfill the model determination [10,13]. The varying dimensional formulation was first introduced to the geophysics literature by Malinverno [31] in a 1D-DC resistivity sounding inversion, and later implemented in a number of geophysical probing inverse problems [39,11,1,35].

The major impact of utilizing probabilistic calibration via a Bayesian approach is the systematic exploration of all combinations of the model parameters through a transparent definition of the impact of the participating uncertainty sources. During such exhaustive parameter exploration, a probability metric is defined to assess the likelihood of selecting sets of parameters that serve to approximate the experimental observations (likelihood); but also a probability density is defined to reflect the degree of knowledge on the model parameters (prior) before the model inversion. The combination of these two states of knowledge about the model of interest yields the following benefits: a transition from deterministic to probabilistic model parameters, assessment of the type and degree of correlation between the model parameters (e.g. linear or non-linear), measurement of the impact of the varying experimental observations (e.g. the effect of the number of observations on the prediction of confidence levels), assessment of the model performance, and most importantly, that among a number of competing models to choose from, selecting the best model which can describe the process that generated the observations. Key applications of the probabilistic subsurface imaging include integrated site investigation, since the recovery of geophysical mechanical parameters allows enhanced geomechanical characterization. [33,34]. The varying parameter dimensionality is formulated through a Bayesian inversion, to populate likely configurations of a heterogeneous elastic medium occupying a semi-infinite domain.

A key defining characteristic of full waveform inversion is the numerical solution of the equations of motion. The governing forward physics consist of a 1D transient scalar acoustic wave propagation, where in order to model the semi-infinite extent of the physical domain, a perfectly matched layer (PML) is introduced at the truncation boundary to emulate the infiniteness of the earth structure [22]. A displacement–stress mixed finite element scheme is used for the numerical solution of the PML-augmented wave PDE.

2. Bayesian approach to inverse problems

An inverse problem is described as the process of estimating some characteristics or parameters of a physical system from a set of directly measurable responses of the system (observations). The vector of model parameters θ, and the vector of observable quantities $\mathbf{d}_{obs}$ are mapped through a forward model. The forward model operator G relies on a physical theory to predict the outcome of a possible experiment, or in other words to approximate the reality: $\mathbf{d}_{obs} \approx G(\theta)$, or

$$\mathbf{d}_{obs} = G(\theta) + \epsilon \qquad (1)$$

where ϵ is the tradeoff component which quantifies the deviation between model predictions and experimental observations. This random term contains both theoretical and measurement errors (assuming that the forward model is an unbiased estimate to the true physical process). Explicit distinction, however, could be made between model and observational errors in a full uncertainty quantification framework (UQ) [32].

The basis of the proposed UQ framework is found on the definition of a "true process" vector $\mathbf{d}$, which in general represents values of observable variables (in this case displacement time history response of earth at the surface level). Notice that in a typical geomechanical processes, $\mathbf{d}$ is not known. However, if the true process is assumed to be random, $\mathbf{d}$ can be defined as a vector of random variables. On the other hand, what the modeler can determine are the following: (1) a vector of physical observations $\mathbf{d}_{obs}$, and (2) a vector of model predictions $\mathbf{d}_{pred}$ (prescribed at the same control points in space and time). $\mathbf{d}_{pred}$ represents a vector of predictions stemming from the forward model G, conditioned on a vector of control parameters θ. $\mathbf{d}_{pred}$ could deviate from the true process ($\mathbf{d}$) as a result of the model not fully capturing the underlying physics, due, for example to the fact that either the governing PDE is an inadequate idealization of the true process, initial/boundary conditions are insufficiently modeled, or due to the deficiency of the computational scheme or lack of resolution of the numerical solver.

Having denoted physical random deviations between $\mathbf{d}$ and $\mathbf{d}_{obs}$ (observation error), and between $\mathbf{d}$ and $\mathbf{d}_{pred}$ (model error) by $\delta\mathbf{d}_{obs}$ and $\delta\mathbf{d}_{pred}$ respectively ($\delta\mathbf{d}_{obs} = \mathbf{d} - \mathbf{d}_{obs}$ and $\delta\mathbf{d}_{pred} = \mathbf{d} - \mathbf{d}_{pred} = \mathbf{d} - G(\theta)$), the observed data is defined as $\mathbf{d}_{obs} = \mathbf{d}_{pred} + \delta\mathbf{d}_{obs} + \delta\mathbf{d}_{pred} = \mathbf{d}_{pred} + \epsilon$.

In general, the error components $\delta\mathbf{d}_{obs}$ and $\delta\mathbf{d}_{pred}$ are not identifiable, meaning that several different combinations of values could be equally consistent with the observed data. However, this does not mean that all the possible values are equally likely [28]. For example, error trends that significantly deviate from zero most likely imply either a bias in the model or a mis-calibration of the data acquisition instrument. The Bayesian method provides a basis for quantifying a priori and a posteriori measures of plausibility of each type of error [27]. In this study, the model discrepancy term $\delta\mathbf{d}_{pred}$ vanishes, since the data is synthesized by perturbing the model output. Therefore, the error component can be defined by a single uncertainty metric as shown in Eq. (1). Notice that this latter formulation is valid also when the model predictions are unbiased along the domain of interest (where $\mathbf{d}$ is defined). That is, when the probabilistic expectation $\mathbb{E}[\delta\mathbf{d}_{obs} - \delta\mathbf{d}_{pred}] = 0$ [32].

In a Bayesian approach to inverse problems, a prior distribution $p(\theta)$ is incorporated in estimating each model unknown, which

quantifies the initial uncertainty about the parameter. Ideally, this density limits the space of plausible parameters by giving higher probability to those which can help us to describe the system's response more accurately. The objective of the inversion is to find the posterior distribution $p(\theta|\mathbf{d}_{obs})$, built to fully describe the model parameters in terms of a density function, given the data $\mathbf{d}_{obs}$ is observed. The Bayes theorem in this context is defined as

$$p(\theta|\mathbf{d}_{obs}) = \frac{p(\mathbf{d}_{obs}|\theta)p(\theta)}{\int_{\Theta} p(\mathbf{d}_{obs}|\theta)p(\theta)\, d\theta} \qquad (2)$$

where the likelihood function $p(\mathbf{d}_{obs}|\theta)$ assess the probability that the observed realization $\mathbf{d}_{obs}$ is produced by the model parameters θ. Under the customary assumption that the random error components $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^T$ are such that $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, with $\mathbf{I}_n$ being an $n \times n$ identity matrix (i.e., uncertainty associated with the data is normal with mean zero and standard deviation σ and data points are independent of each other), the likelihood function is found with reference to a multivariate normal density:

$$p(\mathbf{d}_{obs}|\theta) = \frac{1}{[(2\pi)^n|\mathbf{C}_d|]^{1/2}} \times \exp\left[-\frac{1}{2}(G(\theta) - \mathbf{d}_{obs})^T \mathbf{C}_d^{-1}(G(\theta) - \mathbf{d}_{obs}) \right] \qquad (3)$$

where n is the number of observations and $\mathbf{C}_d = \sigma^2 \mathbf{I}_n$ is the covariance of the error term. The quantity in the denominator of Eq. (2) (the probability of observing the data $\mathbf{d}_{obs}$) is a normalizing constant, such that the posterior is integrated to one.

3. Forward model

This section introduces the forward model used in the model inversion. The forward physics describes seismic vertical propagation of compressional waves within a horizontally stratified semi-infinite elastic earth, when the media is subjected to a uniform excitation $p(t)$ over the surface. This problem can be treated as a one dimensional problem along the depth direction. In a computational setting, a major issue associated with this geo-acoustic inverse problem is to model the semi-infinite physical domain. In order to arrive at a computationally finite region the medium must be truncated at some depth. If the truncated boundary is fixed or inadequately modeled, the propagating waves are (partially) reflected in the domain, and distort the inverted profile [23].

To address the issue, a perfectly matched layer (PML) approach is adopted, and a PML buffer zone is introduced at the truncation interface [22]. The PML enforces the rapid decay of the wave motion within the buffer zone, with ideally no reflection back into the domain. Fig. 1 illustrates the schematic representation of the
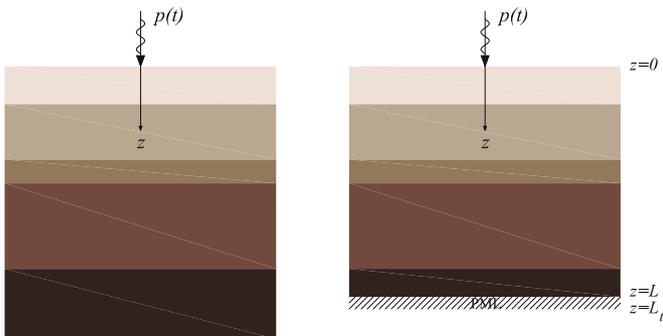
Fig. 1. Schematic presentation of the 1D problem. (a) Original semi-infinite soil media. (b) PML truncated domain.

problem. We refer to the original work for the extensive derivations of the model [22], however, for the sake of completeness we only include the governing wave equation: find $\nu(z, t)$ and $\sum(z, t)$ such that

$$\frac{\partial^2 \nu(z, t)}{\partial t^2} + c(z)g(z)\frac{\partial \nu(z, t)}{\partial t} - \frac{\partial \sum(z, t)}{\partial z} = 0,$$
$$\text{for } z \in (0, L_t), \ t \in (0, T],$$
$$\frac{\partial \sum(z, t)}{\partial t} + c(z)g(z)\sum(z, t) - c^2(z)\frac{\partial^2 \nu(z, t)}{\partial z \partial t} = 0,$$
$$\text{for } z \in (0, L_t), \ t \in (0, T], \qquad (4)$$

subject to

$$\nu(L_t, t) = 0$$
$$\sum(0, t) = p(t)$$
$$\nu(z, 0) = 0$$
$$\frac{\partial \nu}{\partial t}(z, 0) = 0$$
$$\sum(z, 0) = 0 \qquad (5)$$

where ν is the normalized displacement with respect to the soil's density ρ (i.e., $\nu = \rho u$) and $\sum$ denotes the stress. $g(z)$ is an attenuation function which accounts for the artificial dissipation of the wave motion within the buffer zone, and $c(z)$ is the 1D soil compressional wave velocity random field which is the inverse problem parameter. Eq. (4) presents the displacement (ν)–stress ($\sum$) mixed equations governing wave propagation in a PML truncated one dimensional domain. The reader is cautioned not to confuse the standard deviation $\sum$ appeared in Section 2 with stress denoted by $\sum$ (i.e. upright Greek letter).

4. Bayesian partition models

As described in the preceding section, our geo-acoustic inverse problem identifies the spatially dependent coefficient of a PML augmented wave equation $c(z)$, given the probed medium's response to a known excitation. This describes a functional inverse problem where the unknown quantity is a function of the spatial coordinate. Hence, in our Bayesian probabilistic setup, the inverse problem parameter comprises a real-valued random field $c(z)$ (of infinite dimensionality), which assigns a probability density function to the subsurface property of interest at each point in the spatial domain. In order to arrive at a computationally feasible problem, this random field (and the forward model) must be approximated by its discretized version. Hence the velocity field is approximated with an N-dimensional joint probability density $p(c_1(z_1), \ldots, c_N(z_N)|\mathbf{d}_{obs})$, with N being the number of grid blocks in the domain.

One way of treating the problem is to assign a prior to each random variable $\mathbf{c} = (c_1, \ldots, c_N)^T$, and directly apply the Bayesian formulation to form the posterior density of $\mathbf{c}|\mathbf{d}_{obs}$, and implement MCMC methods to explore the resulting very high-dimensional posterior density. Although MCMC methods converge to the posterior by definition as the number of samples grows, in such high dimensional, highly correlated target density configurations, slow chain mixing and serious lack of convergence may arise, which render the whole inversion procedure almost computationally impractical.

Instead of exploring the value of $\mathbf{c}(z)$ at each of the N grid blocks, we introduce a varying dimensional Bayesian model to parameterize the velocity random field. We opt for a Bayesian partition model [9,19] representation of the material field, assuming that $c(z)$ takes the form of a step function (Fig. 2). This
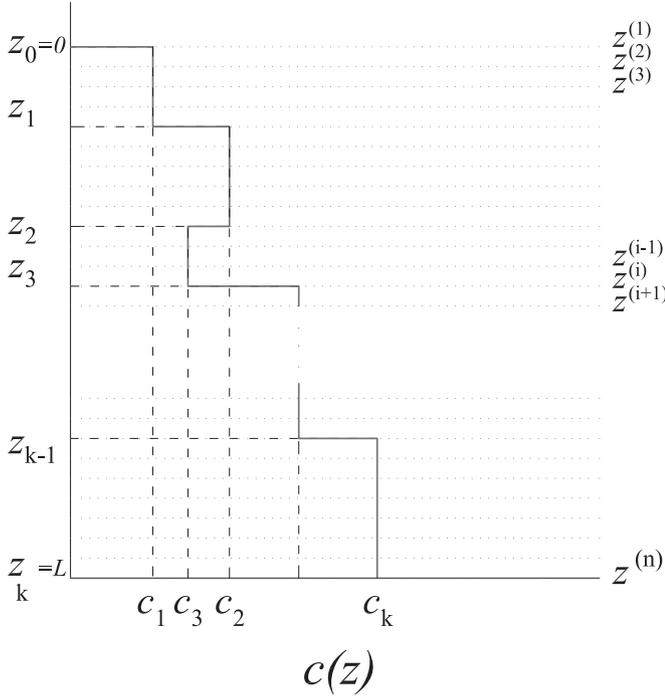
Fig. 2. Partition model presentation of the 1D velocity random field.

setting is well suited for parameterizing a stratified earth model with sharp transition between materials. The Bayesian partition model can be defined by

$$c(z) = \mathbf{Z}\mathbf{c} \tag{6}$$

$$\mathbf{Z} = \begin{pmatrix} I(z^{(1)} \leq z_1) & I(z_1 < z^{(1)} \leq z_2) & \cdots & I(z^{(1)} > z_{k-1}) \\ I(z^{(2)} \leq z_1) & I(z_1 < z^{(2)} \leq z_2) & \cdots & I(z^{(2)} > z_{k-1}) \\ \vdots & \vdots & \ddots & \vdots \\ I(z^{(N)} \leq z_1) & I(z_1 < z^{(N)} \leq z_2) & \cdots & I(z^{(N)} > z_{k-1}) \end{pmatrix} \tag{7}$$

where $\mathbf{Z}$ is called the basis matrix, each column of which forms a basis function. This formulation states that the true k-layer soil is made up of a linear combination of these basis functions and the corresponding coefficients ($\mathbf{c}$). $\mathbf{c} = (c_1, ..., c_k)^T$ hold the value of partition weights (i.e. wave velocity at each layer), and $I(.)$ is the indicator function which assumes the value one, if its argument is true, and zero otherwise. The vector $(z_1, ..., z_{k-1})$ denotes the $k - 1$ change point locations (position of the layer interfaces), where k is unknown number of partitions (layers). $z^{(1)}, ..., z^{(N)}$ are the co-ordinates of N prespecified grid points, which conveniently coincide with the forward model discretization mesh. Fig. 2 shows the partition model presentation of a 1D velocity random field. By partitioning the velocity field the reduced dimension parameter vector will be $(k, \{c_i\}_{i=1}^k, \{z_i\}_{i=1}^{k-1})$.

5. Bayesian model selection

The proposed Bayesian partition model is categorized within a special class of models namely variable dimension models. A variable dimension model is defined as a model whose number of unknowns is an unknown itself [9]. A Bayesian variable dimension model is composed of a set of plausible models $\{\mathcal{M}_k\}_{k=1}^K$, each reflecting a hypothesis about the data. Having K such competing models, it is desired to find the model that best describes the observations. Here, $\mathcal{M}_k$ represents a k-layer subsurface structure.

This definition by nature pertains to the spatial case of a Bayesian model selection problems, where the competing models belong to the same family, with differing number of parameters, namely nested models [38].

A variable dimension model can be formulated as an extension to the standard Bayesian inference (see Section 2), where a prior distribution is assigned on the model indicator $\mathcal{M}_k$, which implies extending the prior modeling from parameters to models. $\boldsymbol{\Theta} = \bigcup_{k \in \mathcal{K}} \{k\} \times \boldsymbol{\Theta}_k$ is the parameter space associated with the set of models $\mathcal{M}_k$, where $\mathcal{K}$ denotes the space of model indicators. Having defined priors π_k on the indicator parameter $\mathcal{M}_k$ (being considered now part of the parameters), and the parameter sub-space $\boldsymbol{\Theta}_k$, by virtue of Bayes's theorem

$$p(\mathcal{M}_k, \boldsymbol{\theta}_k | \mathbf{d}_{obs}) = \frac{p(\mathbf{d}_{obs} | \mathcal{M}_k, \boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k | \mathcal{M}_k) \pi_k}{\sum_{k \in \mathcal{K}} \pi_k \int_{\Theta_k} p(\mathbf{d}_{obs} | \boldsymbol{\theta}_k, \mathcal{M}_k) p(\boldsymbol{\theta}_k | \mathcal{M}_k) \, d\boldsymbol{\theta}_k} \tag{8}$$

where $\boldsymbol{\theta}_k$ is a set of parameters specific to model $\mathcal{M}_k$, and $p(\mathbf{d}_{obs} | \mathcal{M}_k, \boldsymbol{\theta}_k)$ and $p(\boldsymbol{\theta}_k | \mathcal{M}_k)$ are the likelihood function and the prior density of the model specific parameters $\boldsymbol{\theta}_k$, respectively, given $\mathcal{M}_k$ is the true model. Bayes factors in the Bayesian model selection context offer a thorough criterion to pairwise comparison of members in $\{\mathcal{M}_k\}$. The relative plausibility of model i versus model j, having observed $\mathbf{d}_{obs}$, is determined by the Bayes factor given by

$$BF[\mathcal{M}_i : \mathcal{M}_j] = \frac{p(\mathbf{d}_{obs} | \mathcal{M}_i)}{p(\mathbf{d}_{obs} | \mathcal{M}_j)} \tag{9}$$

where $p(\mathbf{d}_{obs} | \mathcal{M}_i)$ is the marginal likelihood of data under model $\mathcal{M}_i$, which is the normalizing constant of the posterior density (see the denominator in Eq. (8)), and is given by

$$p(\mathbf{d}_{obs} | \mathcal{M}_i) = \int_{\Theta_i} p(\mathbf{d}_{obs} | \mathcal{M}_i, \boldsymbol{\theta}_i) p(\boldsymbol{\theta}_i | \mathcal{M}_i) \, d\boldsymbol{\theta}_i \tag{10}$$

The above quantity (equation (10)) is the basis for the Bayesian method's natural penalty against complex models, also known as Occam's razor [8]. The Bayesian embodiment of Occam's razor is briefly explained in Appendix A. Non-Bayesian model selection procedures [2,42,25] rely on comparing (penalized) maximum likelihoods of the competing models [46]. Such criterion breaks down when a set of nested models is to be compared, and over-fitting becomes a serious problem. Since a more flexible model is able to describe the data better, it gives rise to a higher likelihood measure, and therefore is favored in a likelihood-based test, while it performs poorly in terms of prediction. The Bayesian method's built-in penalty against overly complex models provides a robust tool to treat varying dimensional problems.

6. Prior elicitation

The first step in a Bayesian inference analysis setting is to specify prior densities to the model parameters $\boldsymbol{\theta}$ (given the model representation ($\mathcal{M}$) is chosen). The prior distribution $p(\boldsymbol{\theta})$ is basically a tool to summarize the initially available information on the process, and to quantify the uncertainty associated with this information. Selecting standard vague or non-informative priors is favored in this case in order to base the inference only on the experimental observations.

A number of techniques are currently available for constructing such standard priors [26]. The use of non-informative priors, however, is rather delicate for varying dimensional model settings, since the majority of non-informative priors are improper. Improper priors are only defined up to a proportionality constant (i.e., not integrable) [26]. In general, improper priors cannot be assigned to model specific parameters in Bayesian model selection

problems, as the choice of the arbitrary normalizing constant will influence the Bayes factor (Eq. (9)). Notice that the Bayes factor is a multiple of this normalizing constant (Eqs. (9) and (10)). Proper vague priors (proper priors with large dispersion parameter) also do not address the difficulty, for they give rise to the so-called Jeffreys–Lindley paradox [29,24]. The Jeffreys–Lindley paradox is a problem related to the stability of the Bayes factor, which causes the simplest model (which might be a very poor reflection of the data) to always be favored by the Bayes factor.

We address these concerns in our choice of priors. In particular, we propose the use of a hierarchical Bayes approach to model the lack of information on the parameters of the prior distribution, by introducing a second level of prior distributions on these parameters. Hence we refrain from using improper priors, yet avoiding any subjective input to the inference by introducing unground informative priors. The posterior kernel (of varying dimension) according to the Bayes rule is

$$p(\boldsymbol{\theta}_k, k|\mathbf{d}_{obs}) \propto p(\mathbf{d}_{obs}|\boldsymbol{\theta}_k, k)p(\boldsymbol{\theta}_k, k) \tag{11}$$

where $\boldsymbol{\theta}_k$ is the parameter vector associated to the k soil layer model. $\mathbf{d}_{obs}$ denotes the $n \times 1$ vector of normalized displacement response, recorded at the soil surface. Introducing the second layer of hierarchy will lead to

$$p(\boldsymbol{\theta}_k, k|\mathbf{d}_{obs}) \propto p(\mathbf{d}_{obs}|\boldsymbol{\theta}_k, k)p(\boldsymbol{\theta}_k|k)p(k)$$
$$\propto p(\mathbf{d}_{obs}|\boldsymbol{\theta}_k, k)p(\mathbf{c}|\sigma^2, v, k)p(\mathbf{z}|k)p(\sigma^2)p(v)p(k|\lambda)p(\lambda) \tag{12}$$

For the ease of notation, we define vectors $\boldsymbol{\theta}_P$ and $\boldsymbol{\theta}_H$ which contain the model specific parameters and the global hyperparameters, respectively. The global parameters are unknowns which bear on parameters common to all models. Thus $\boldsymbol{\theta}_P = (\mathbf{c}_{k \times 1}^T, \mathbf{z}_{(k-1) \times 1}^T)^T$, $\boldsymbol{\theta}_H = (\sigma^2, v, \lambda)^T$, $\boldsymbol{\theta}_k = (\boldsymbol{\theta}_P^T, \boldsymbol{\theta}_H^T)^T$, and $\boldsymbol{\theta} = (\boldsymbol{\theta}_k^T, k)^T$. Superscript T denotes transposition. The definitions of the priors are

$$\log \mathbf{c}|\sigma^2, v, k \sim \mathcal{N}(\mathbf{c}_0, \sigma^2 v \mathbf{I}_{k \times k}) \tag{13a}$$

$$\sigma^2 \sim \mathcal{IG}(\alpha_0, \delta_0) \tag{13b}$$

$$v \sim \mathcal{G}a(\zeta_0, \eta_0) = \frac{\eta_0^{\zeta_0}}{\Gamma(\zeta_0)} v^{(\zeta_0 - 1)} e^{(-\eta_0 v)} \tag{13c}$$

$$p(\mathbf{z}|k) \propto \binom{T}{k-1}^{-1} \tag{13d}$$

$$k|\lambda \sim \frac{1}{\sum_{i=1}^{K}(\lambda^i/i!)} \frac{\lambda^{k-1}}{(k-1)!}, \quad k = 1, \ldots, K \tag{13e}$$

$$\lambda \sim \mathcal{G}a(\iota_0, \kappa_0) \tag{13f}$$

We, a priori, assume that the velocity field within each layer is populated from a log-normal distribution. Hence, the log-velocity field has a multi-variate Gaussian prior density (Eq. (13a)). This assumption ensures that velocity is a positive-valued random filed. We further suppose that $c_1, \ldots, c_k$ are a priori independent. The correlation structure of the layered elastic properties will be reconstructed a posteriori (if there exists any). $\mathbf{c}_0$ is set to 200 m/s, meaning that before the inversion, the media is assumed to be homogeneous.

The hyper parameters σ^2 and v are the noise variance and the precision parameter respectively. We opt for a broadly non-informative priors for these parameters ($\alpha_0 = \delta_0 = 0.01$ and $\zeta_0 = \eta_0 = 0.01$), (Eqs. (13b) and (13c)). Since the inference is highly sensitive to the choice of v, it is important to avoid fixing this parameter [8]. By considering it as a random variable we elevate

the robustness of the method against poor choices of v. There is no restriction in using improper priors for the global parameters (which are common among all the models) [3], since in Bayes factor calculation (Eq. (9)) the arbitrary constant of proportionality cancels out. Notice that the Bayes factor is a multiple of the prior normalizing constant. Hence, the problem with the arbitrary proportionality constant, which brings about Lindley's paradox, is removed.

$\mathbf{z}$ is the position vector of the $k-1$ layer interfaces. $p(\mathbf{z}|k)$ (Eq. (13d)) reflects the prior assumptions about the position of the material interfaces. We define an underlying grid of T points (which coincides with the finite element discretization of the physical domain). This prior suggests that given a k layer model is the true process, and there are T candidate nodes to locate $k-1$ interfaces, any combination of $(z_1, \ldots, z_{k-1})$ is equally likely. We assign a hierarchical truncated Poisson prior on $k|\lambda$, with K being the maximum number of layers in partitioning (Eq. (13e)). This setting controls the prior weights given to over-parameterized models. λ is a hyperparameter to be elicited from the data. A natural choice of prior on this parameter is a flat Gamma distribution ($\iota_0 = \kappa_0 = 0.01$). The defined prior $p(\boldsymbol{\theta}_k, k)$ does not place an explicit penalty on the model complexity. However, as stated earlier, the marginal likelihood contains a built-in penalty on the model dimension, which strongly depends on the prior variance v of the coefficients $\mathbf{c}$ [8].

7. Posterior computation

A customary burden of using Bayes factors (Eq. (9)) is the computation of, oftentimes, high dimensional marginal likelihood integrals (Eq. (10)). To circumvent this difficulty, one may resort to alternative solutions such as Monte Carlo simulation based methods (e.g., pseudo-priors [4,5]), or asymptotic approximation to Bayes factors (e.g., Schwartz's criteria also known as BIC) [42]. The latter is widely used in a variety of applications including geophysical modeling (e.g., see [13,10,45]) due to the ease of its implementation. BIC provides a first-order approximation to the logarithm of the Bayes factor as the sample size grows. The applicability of the approximation, however, is restricted to models with regular likelihoods, and i.i.d. data structures. Also the method calls for the derivation of maximum likelihood estimates for the parameters of all models, which is an unfavorable fact when K is large. Monte Carlo methods provide essentially the only accurate mean of inferring the posterior which does not depend on the knowledge of the proportionality. Markov Chain Monte Carlo (MCMC) is an iterative stochastic method designed to generate random samples from the posterior kernel. A generalization of MCMC, the so-called Reversible Jump MCMC (RJMCMC), was introduced by Green [18] to generate random samples which are distributed according to a varying dimensional posterior. The key aspect of the reversible jump algorithm is the introduction of some auxiliary random variables to equalize the dimensionality of the parameter space across models in order for the Markov chain to be able to move among different dimensions. RJMCMC has recently become increasingly popular in geophysical inversion as a robust tool for subsurface modeling. A detailed introduction to geophysical transdimensional Bayesian inversion can be found in Sambridge et al. [40].

7.1. Reversible jump MCMC implementation as birth–death process

In this section we extract the details involved in the RJMCMC algorithm designed for our specific inversion setup. Suppose we want to draw random samples from varying dimensional target

distribution $p(\theta_k, k|\mathbf{d}_{obs})$ (Eq. (11)), where the sequence of random samples constructs a Markov chain.

In order to traverse the varying dimensional posterior surface, we iteratively perform four types of transitions: Birth (B), Death (D), Move (M), and Perturb (P). Different search strategies have been designed depending on the application (e.g., see the original work by Green [18], and Denison et al. [6]). As long as the algorithm satisfies the balance condition, and the acceptance ratio remains computationally efficient, we assume that our approach offers a flexible design.

Let us suppose that at the (s) th step the chain is at $k^{(s)}, \theta_P^{(s)}, \theta_H^{(s)}$ (denoting number of layers, model specific parameters $\mathbf{c}^{(s)} = (c_1^{(s)}, ..., c_k^{(s)})^T$, $\mathbf{z}^{(s)} = (z_1^{(s)}, ..., z_{k-1}^{(s)})^T$, and hyper-parameters $\sigma^{2(s)}, v^{(s)}, \lambda^{(s)}$ respectively). The possible transitions are the following: (B) add an intersection at a random location with probability $p_{k^{(s)}}^{(B)}$. (D) Delete a randomly chosen intersection with probability $p_{k^{(s)}}^{(D)}$. (M) Swap a randomly chosen intersection for a randomly chosen available node in $\mathcal{T}$ with probability $p_{k^{(s)}}^{(M)}$, where $\mathcal{T}$ is the set of candidate node locations, and T is the size of the set $\mathcal{T}$ ($|\mathcal{T}| = T$). (P) Perturb velocity of a randomly chosen layer with probability $p_{k^{(s)}}^{(P)}$. Where $p_{k^{(s)}}^{(B)} + p_{k^{(s)}}^{(D)} + p_{k^{(s)}}^{(M)} + p_{k^{(s)}}^{(P)} = 1$, $\forall k^{(s)}$. Notice that (B) and (D) involve dimension changes in $\theta_P^{(s)}$, while (M), and (P), proposes moves within the current dimension, hence the latter proceeds similar to regular Metropolis–Hastings algorithm [37]. Below is the definition of each transition:

- *Birth*: $k^* = k^{(s)} + 1$.
  With probability $p_{k^{(s)}}^{(B)} = q_{k^{(s)},k^*}$, a Birth move is proposed, and a layer interface i is added at an available random grid location (Fig. 3a). This random location is proposed from the discrete probability $q_z(\mathbf{z}^*|\mathbf{z}^{(s)}, k^{(s)})$ (Eq. (16b)). The birth increases the dimensionality of the parameter space by two, the difference being accounted for one discrete variable, the new interface position z_i^*, and one continuous variable, the new velocity c_i^* (i.e., $\dim(\boldsymbol{\Theta}_{k^*}) - \dim(\boldsymbol{\Theta}_{k^{(s)}}) = 2$, where $\boldsymbol{\Theta}_{k^{(s)}}$ denotes the space of current state and $\boldsymbol{\Theta}_{k^*}$ is the candidate parameter space). In order for the Markov chain to jump between subspaces $\boldsymbol{\Theta}_{k^{(s)}}$ and $\boldsymbol{\Theta}_{k^*}$, $\theta^{(s)}$ is augmented with a *two* dimensional auxiliary variable $\mathbf{u}$ drawn from a proposal distribution $\psi(\mathbf{u})$. The new state of the chain θ^* is found from the transformation $\mathcal{T}$ such that $\theta^* = \mathcal{T}_{k^{(s)},k^*}(\theta^{(s)}, \mathbf{u})$. $\mathcal{T}_{k^{(s)},k^*}$ is a deterministic mapping, the so-called dimension matching transformation. The proposed layer

velocities $\mathbf{c}^*$ from the transformation $\mathcal{T}_{k^{(s)},k^*}(\mathbf{c}^{(s)}, u)$ is given by

$$\mathcal{T}_{k^{(s)},k^*}(\mathbf{c}^{(s)}, u) = \begin{cases} c_1^* = c_1^{(s)} \\ \quad\vdots \\ c_{i-1}^* = c_{i-1}^{(s)} \\ c_i^* = c_i^{(s)} - \varsigma_c u \\ c_{i+1}^* = c_i^{(s)} + \varsigma_c u \\ \quad\vdots \\ c_{k^*}^* = c_{k^{(s)}}^{(s)} \end{cases} \tag{14}$$

This implies that the velocity of the chosen layer is perturbed from a Gaussian proposal to attain the velocity of the two emerged layers. ς_c is a variance measure, defining size of the search step. Notice that the hyperparameters remain unchanged in the Birth (also in Death) move. The candidate state is accepted with probability

$$r_{k^{(s)},k^*}(\theta_P^{(s)}, \theta_{\tilde{P}}^*)$$

$$= \min\left\{ 1, \underbrace{\frac{p(\theta_{\tilde{P}}^*, k^*)}{p(\theta_P^{(s)}, k^{(s)})}}_{\text{prior ratio}} \underbrace{\frac{p(\mathbf{d}_{obs}|\theta_{\tilde{P}}^*, \theta_H^{(s)}, k^*)}{p(\mathbf{d}_{obs}|\theta_P^{(s)}, \theta_H^{(s)}, k^{(s)})}}_{\text{likelihood ratio}} \underbrace{\frac{q_{k^*,k^{(s)}}\psi_{k^*,k^{(s)}}(u'|\mathbf{c}^*)q_z(\mathbf{z}^{(s)}|\mathbf{z}^*, k^*)}{q_{k^{(s)},k^*}\psi_{k^{(s)},k^*}(u|\mathbf{c}^{(s)})q_z(\mathbf{z}^*|\mathbf{z}^{(s)}, k^{(s)})}}_{\text{proposal ratio}} \right.$$

$$\times \underbrace{\left| \frac{\partial \mathcal{T}_{k^{(s)},k^*}(\mathbf{c}^{(s)}, u)}{\partial \mathbf{c}^{(s)} \partial u} \right|}_{\text{Jacobian}} \right\} \tag{15}$$

where the prior ratio is

$$\frac{p(\theta_{\tilde{P}}^*, k^*)}{p(\theta_P^{(s)}, k^{(s)})} = \frac{p(\mathbf{c}^*|\sigma^{2(s)}, v^{(s)}, k^*)p(\mathbf{z}^*|k^*)p(k^*|\lambda^{(s)})}{p(\mathbf{c}^{(s)}|\sigma^{2(s)}, v^{(s)}, k^{(s)})p(\mathbf{z}^{(s)}|k^{(s)})p(k^{(s)}|\lambda^{(s)})}$$

and

$$q_z(\mathbf{z}^{(s)}|\mathbf{z}^*, k^*) \propto \frac{1}{k^* - 1} \tag{16a}$$

$$q_z(\mathbf{z}^*|\mathbf{z}^{(s)}, k^{(s)}) \propto \frac{1}{T - (k^{(s)} - 1)} \tag{16b}$$
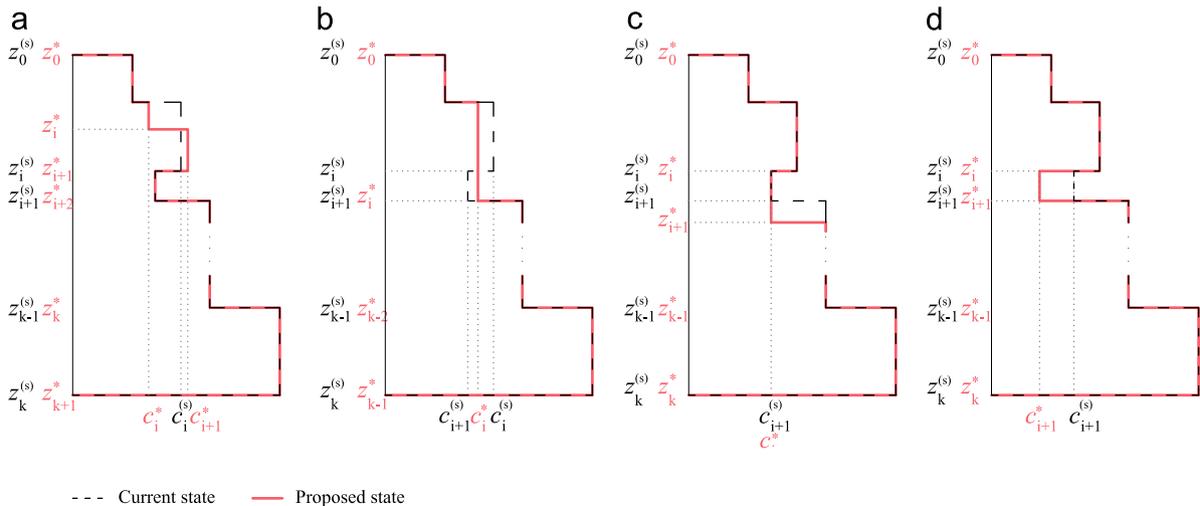
Fig. 3. The RJMCMC transitions: Birth, Death, Move and Perturb. By random combination of these four types of transitions, MCMC searches the parameter space for the solution.

$$\psi_{k*,k^{(s)}}(u'|\mathbf{c}^*) \propto 1 \tag{16c}$$

$$\psi_{k^{(s)},k*}(u|\mathbf{c}^{(s)}) \sim \frac{1}{\varsigma_c}\mathcal{N}\left(0,\ 1|u\right) \tag{16d}$$

$$q_{k^{(s)},k*} = \begin{cases} 1/2, & k^{(s)} = 1 \\ 1/4 & \text{otherwise} \end{cases} \tag{16e}$$

$$q_{k*,k^{(s)}} = \begin{cases} 1/3, & k^{(s)} = K \\ 1/4 & \text{otherwise} \end{cases} \tag{16f}$$

$$\left|\frac{\partial \mathcal{T}_{k^{(s)},k*}(\mathbf{c}^{(s)},\ u)}{\partial \mathbf{c}^{(s)}\partial u}\right| = 2\varsigma_c \tag{16g}$$

$p(\mathbf{d}_{obs}|\boldsymbol{\theta}_P, \boldsymbol{\theta}_H, k)$ is the likelihood function, which is constructed according to Eq. (3).

Notice that the parameter space for the interface positions is discrete (whereas it is continuous for the velocity space), and a discrete uniform random variable was drawn to propose the position of the new interface (16b). Denison et al. [9] show that the Jacobian term is always unity for discrete transformations. Therefore only where continuous model spaces change in dimensions, determining the Jacobian is required, and the dimension matching transformation $\mathcal{T}$ maps the continuous

- subspace: $(\{c^{(s)}\}_{i=1}^{k^{(s)}},\ u) \underset{\mathcal{T}^{-1}}{\overset{\mathcal{T}}{\rightleftharpoons}} \mathcal{T}\left(\{c^*\}_{i=1}^{k*}\right)$.
  Death: $k* = k^{(s)} - 1$.
  With probability $p_{k^{(s)}}^{(D)} = q_{k^{(s)},k*}$, a Death move is proposed. A current interface i is randomly chosen from the probability $q_z(\mathbf{z}*|\mathbf{z}^{(s)}, k^{(s)})$ and removed (Fig. 3b). The proposed velocity profile $\mathbf{c}*$ is determined from the deterministic Death transformation (which is the reverse Birth transformation)

$$\mathcal{T}_{k^{(s)},k*}(\mathbf{c}^{(s)},\ u) = \begin{cases} c_1^* = c_1^{(s)} \\ \quad\vdots \\ c_{i-1}^* = c_{i-1}^{(s)} \\ c_i^* = \frac{1}{2}\left(c_i^{(s)} + c_{i+1}^{(s)}\right) \\ c_{i+1}^* = c_{i+2}^{(s)} \\ \quad\vdots \\ c_{k*}^* = c_{k^{(s)}}^{(s)} \end{cases} \tag{17}$$

The acceptance probability is the same as Eq. (15), with the following modifications:

$$q_z(\mathbf{z}^{(s)}|\mathbf{z}*, k*) \propto \frac{1}{T - (k^{(s)} - 1)} \tag{18a}$$

$$q_z(\mathbf{z}*|\mathbf{z}^{(s)}, k^{(s)}) \propto \frac{1}{k* - 1} \tag{18b}$$

$$\psi_{k*,k^{(s)}}(u'|\mathbf{c}*) \sim \frac{1}{\varsigma_c}\mathcal{N}(0,\ 1|u) \tag{18c}$$

$$\psi_{k^{(s)},k*}(u|\mathbf{c}^{(s)}) \propto 1 \tag{18d}$$

$$q_{k^{(s)},k*} = \begin{cases} 1/3, & k^{(s)} = 1 \\ 1/4 & \text{otherwise} \end{cases} \tag{18e}$$

$$q_{k*,k^{(s)}} = \begin{cases} 1/2, & k^{(s)} = K \\ 1/4 & \text{otherwise} \end{cases} \tag{18f}$$

$$\left|\frac{\partial \mathcal{T}_{k^{(s)},k*}(\mathbf{c}^{(s)},\ u)}{\partial \mathbf{c}^{(s)}\partial u}\right| = \frac{1}{2\varsigma_c} \tag{18g}$$

- *Move*: $k* = k^{(s)}$.
  With probability $p_{k^{(s)}}^{(M)} = q_{k^{(s)},k*}$, a "Move" move is proposed. A layer interface is randomly chosen from a uniform probability, and moved to an available knot location (Fig. 3c). A new set of hyper parameters $\boldsymbol{\theta}_H^*$ is drawn from the probability $q(\boldsymbol{\theta}_H^*|\boldsymbol{\theta}_H^{(s)})$. Log-normal proposals are used to update all the hyperparameters.

  In a Move step, as the number of material layers is fixed, the algorithm reduces to the regular Metropolis–Hastings MCMC with the acceptance probability of the following form: (Notice that the hyperparameters of the model are also updated in Move and Perturb.)

$$r_{k^{(s)},k*}(\boldsymbol{\theta}^{(s)}, \boldsymbol{\theta}*)$$

$$= \min\left\{1,\ \underbrace{\frac{p(\boldsymbol{\theta}*)}{p(\boldsymbol{\theta}^{(s)})}}_{\text{prior ratio}}\ \underbrace{\frac{p(\mathbf{d}_{obs}|\boldsymbol{\theta}_P^*, \boldsymbol{\theta}_H^*, k^{(s)})}{p(\mathbf{d}_{obs}|\boldsymbol{\theta}_P^{(s)}, \boldsymbol{\theta}_H^{(s)}, k^{(s)})}}_{\text{likelihood ratio}}\ \underbrace{\frac{q(\boldsymbol{\theta}_H^{(s)}|\boldsymbol{\theta}_H^*)}{q(\boldsymbol{\theta}_H^*|\boldsymbol{\theta}_H^{(s)})}}_{\text{proposal ratio}}\right\} \tag{19}$$

- *Perturb*: $k* = k^{(s)}$.
  With probability $p_{k^{(s)}}^{(P)} = q_{k^{(s)},k*}$, a Perturb move is proposed. A layer is randomly picked from a uniform density, and its material property is perturbed with a Gaussian proposal (Fig. 3d). It is also attempted to update the model hyperparameters from log-normal proposal densities (similar to the M move). The probability of accepting the candidate state is found from Eq. (19). Notice that the uniform and Gaussian proposals to update $\boldsymbol{\theta}_P$ do not appear in this ratio (also in the M step), for reasons of symmetry.

8. Application to a synthetic case

The inversion scheme outlined in the preceding sections is applied to a synthetic data set to deduce the subsurface elastic properties of a soil model. We consider the horizontally stratified semi-infinite soil medium depicted in Fig. 4. The medium is modeled as a one-dimensional PML-truncated domain, with the regular domain extending to $z = 100$ m, and the PML buffer zone thickness being 10 m. Fig. 4 illustrates the target wave velocity profile, which reflects sharp transitions between different materials in depth. The medium is probed with a Gaussian pulse-type excitation $p(t)$ applied at the soil surface (Fig. 5a). The maximum frequency of the excitation is 40 Hz and the peak amplitude is 10 kPa. Fig. 5b depicts the frequency spectrum of the excitation.

Fig. 6a shows the displacement time history response of the soil model depicted in Fig. 4 to the prescribed load. This is obtained by solving the forward problem (4) and (5) using a mixed finite element method. 220 elements of length 0.5 m are used in the analysis. The readings are recorded every 0.001 s for a total duration of 2 s. Displacement response, as a measurable characteristic of the wave field, will serve as the input to our inversion scheme. We generate the synthetic data by perturbing the
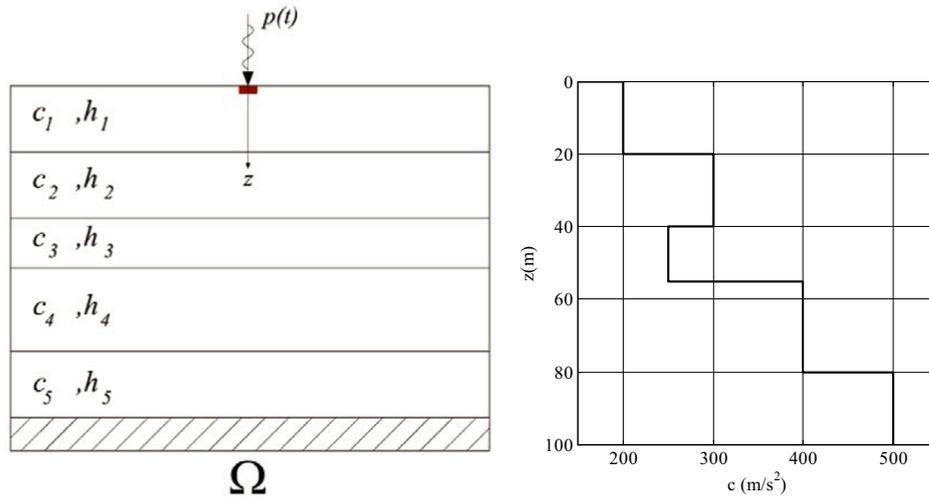
Fig. 4. Benchmark soil velocity profile.

displacement response of the soil model computed at the ground level $v(0, t)$ with 20% Gaussian noise. Fig. 6b illustrates this data set. The attenuation effect is disregarded in this study, and the soil density is assumed to be known a priori ($\rho = 2000$ kg/m³).

9. Results

In this section, we illustrate the applicability of the Bayesian varying dimensional inversion, and model determination using the methodology introduced in the preceding sections. The inversion is allowed for a maximum of 40 soil layers (up to the truncation interface), which indicates maximum number of 83 model unknowns. This maximum resolution is attributed to the frequency of the exerted load (maximum frequency 40 Hz). The simplest earth model is $k=1$, which corresponds to the state of a homogeneous medium. No additional assumption is made concerning the regularization of the deduced velocity profile.

We started the inversion with homogeneous initial guess ($k = 1$, $c = 200$ m/s). The RJMCMC sampler was run, and a total of 100K iterations were stored as the generated samples. The first 20K samples were discarded as burn-in iterations. Every fifth visited sample was kept in the chain as high dependency is expected, especially between successive values of k, since the

difference between the current and the proposed k values could be at most one. Fig. 7 illustrates the first 300 RJMCMC sampling sequence for the model index (number of layers), starting from $k=1$. This figure shows that k increases rapidly up to $k=10$ and in about 200 iterations, then it settles down to the five layer target model. This figure also implies that even though our sampling strategy does not force the model to undergo dimension changes at every iteration (we are pointing to M and P moves) the waiting time at a single model is not long. Hence the sampler promptly explores the space of plausible models until it converges to the target model $k=5$. The rest of the simulation effort is committed to arriving at the stationary condition in sampling the soil parameters of the few favored models. This observation confirms the efficiency of the algorithm design and of the proposal density formulations.

Fig. 8a depicts the full sampling history for the same parameter, to further emphasize the stability of the RJMCMC chain. The marginal posterior probability mass function of k is shown in Fig. 8b, which quantifies the level of certainty in accepting each hypothesis. According to this figure, 6 layer profile is also a likely model to describe the observations with much less probability. The figure manifests the Bayesian inversion capability to deduce the true nature of the underlying process without imposing any regularization constraint to penalize overly complex models.
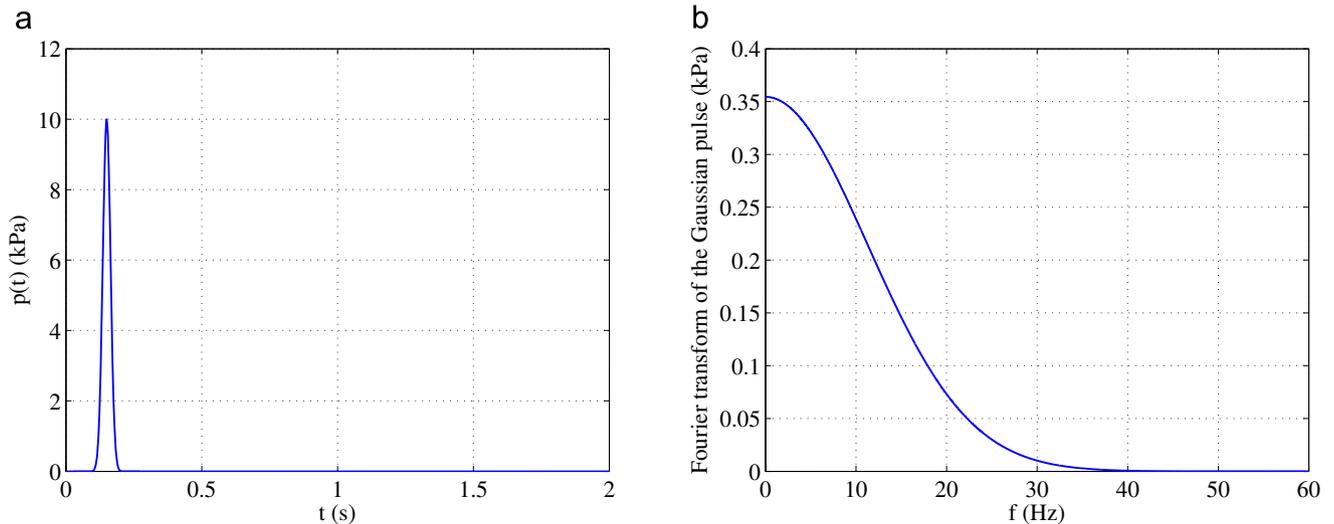
Fig. 5. (a) Time history of the applied stress $p(t)$. (b) Frequency spectrum of the applied stress $p(t)$.

a

b
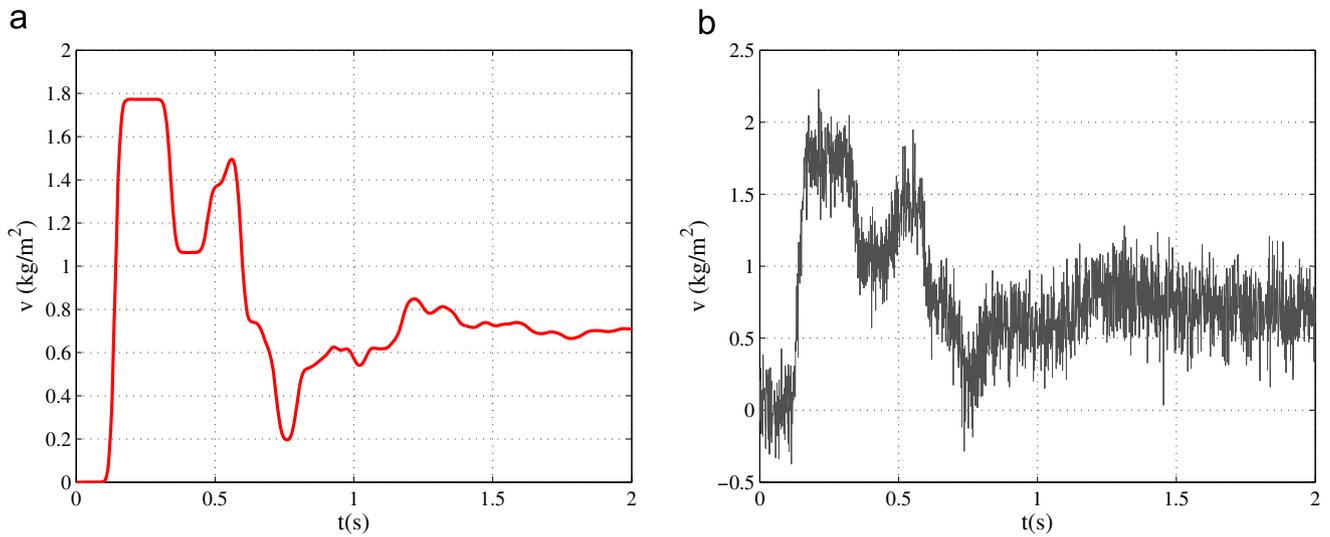
Fig. 6. (a) Computed displacement response at the surface (b) Synthetic data: computed displacement response at $z=0$ perturbed with 20% Gaussian noise.
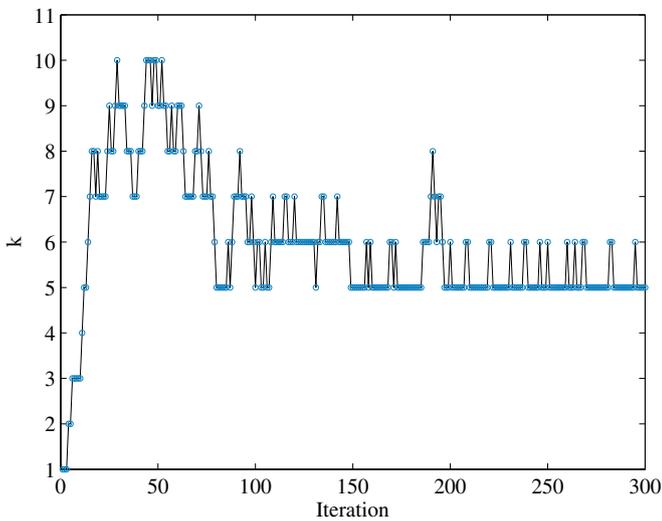
Fig. 7. Number of basis functions (layer interfaces) variation in first 300 RJ-MCMC iterations.

Figs. 9 and 10 illustrate the marginal posterior densities of the model specific parameters, given the true model $k=5$. Fig. 9 shows the posterior estimates for the layer thicknesses $p(\mathbf{z}|\mathbf{d}_{obs}, k = 5)$, and their associated uncertainties. The target values are also superimposed on each histogram (dashed lines as reference). The figure indicates the ability of the inversion scheme to deduce the target parameters (notice that the deviation of the posterior mean from the target values are about one to two element dimensions and that the scales of the figures are not the same). The thickness of the fifth layer is not included here, as it is considered to be semi-infinite.

Theoretically, the PML is assumed to be located at a depth beyond which homogeneity is ascertained. Fig. 10 shows the inverted acoustic soil velocities of the true model $p(\mathbf{c}|\mathbf{d}_{obs}, k = 5)$, together with the target values (notice the variation of scales on the figures as well).

Fig. 11 shows inference for model hyperparameters. Although these parameters might not be incorporated directly in model predictions, they are highly influential in attaining reliable parameter estimates. In Fig. 11a the standard deviation of the observational error term σ^2 is displayed, which is relatively centered around the target added Gaussian noise (signal to noise ratio, SNR=5).

a

b
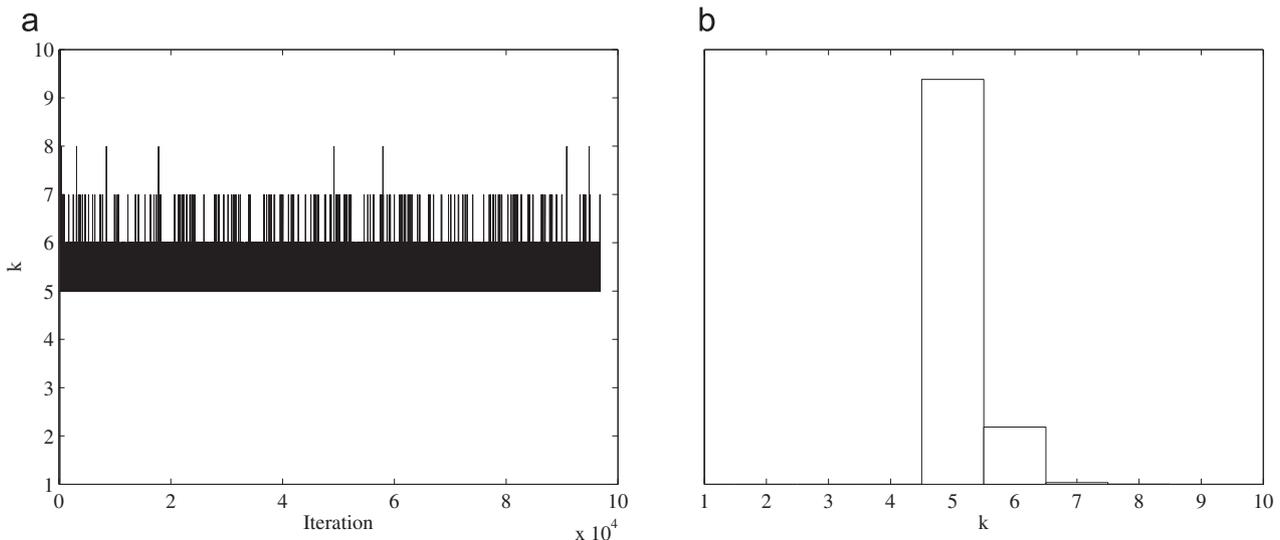
Fig. 8. Number of layers (k) in the partition model. (a) Posterior samples of k. (b) Posterior mass function of k.
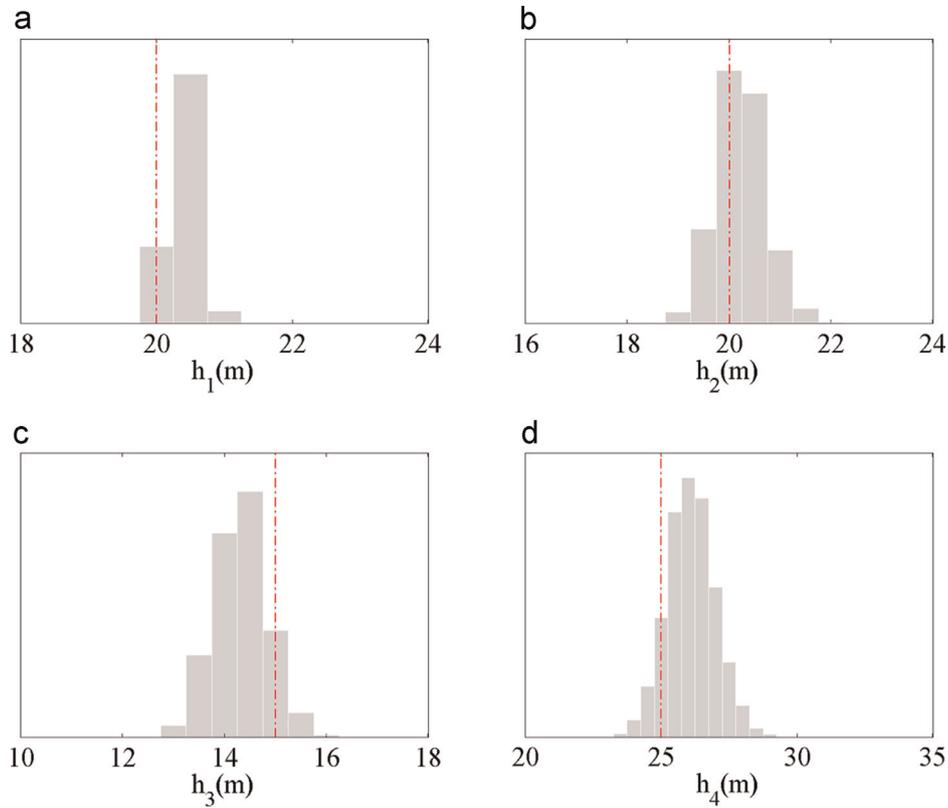
Fig. 9. Marginal posterior density of the layer thicknesses given $k=5$ and the corresponding target values (dashed line).

Fig. 10. Marginal posterior density of the layer velocities given $k=5$ and the corresponding target values (dashed line).
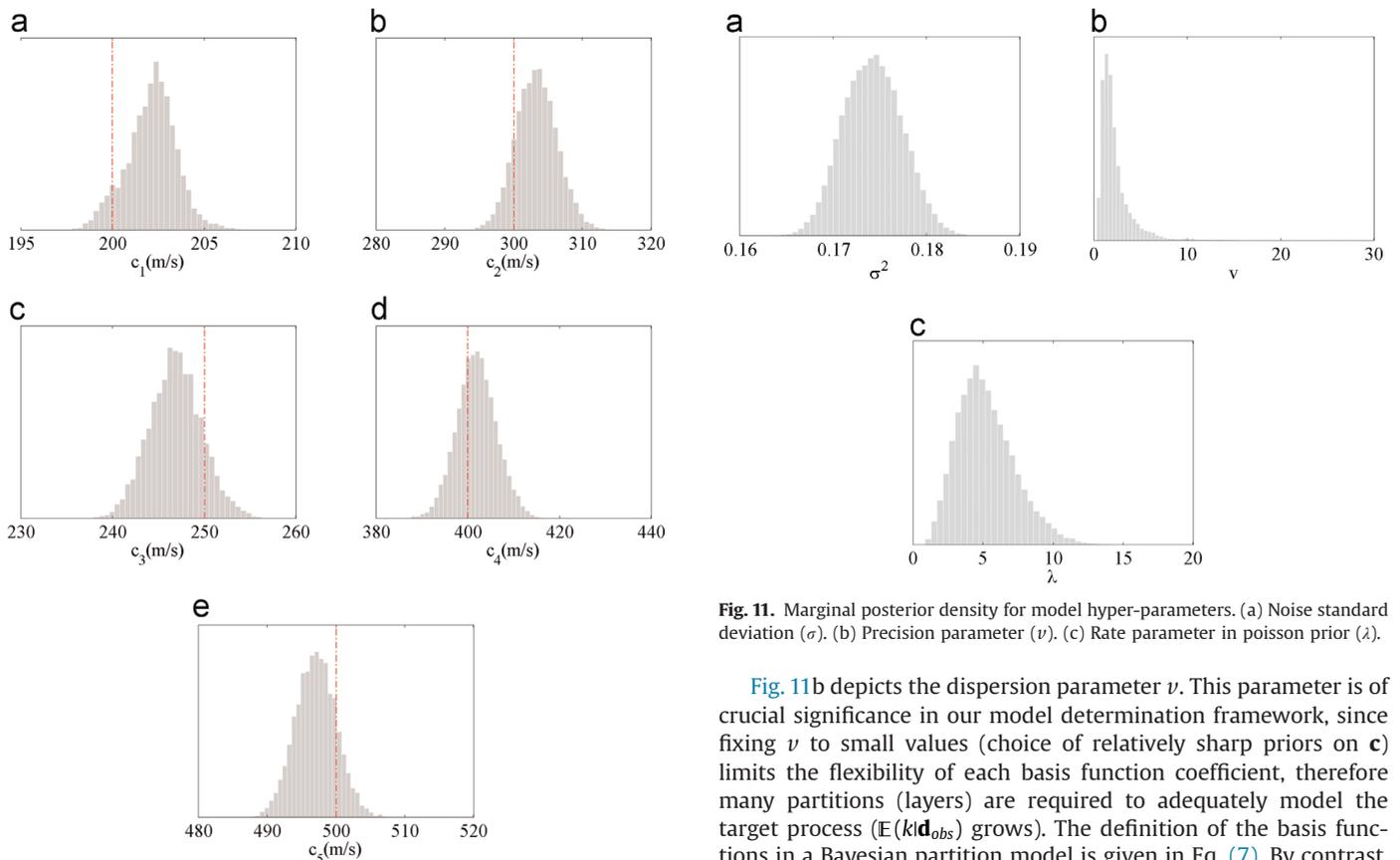
Fig. 11. Marginal posterior density for model hyper-parameters. (a) Noise standard deviation (σ). (b) Precision parameter (ν). (c) Rate parameter in poisson prior (λ).

Fig. 11b depicts the dispersion parameter ν. This parameter is of crucial significance in our model determination framework, since fixing ν to small values (choice of relatively sharp priors on $\mathbf{c}$) limits the flexibility of each basis function coefficient, therefore many partitions (layers) are required to adequately model the target process ($\mathbb{E}(k|\mathbf{d}_{obs})$ grows). The definition of the basis functions in a Bayesian partition model is given in Eq. (7). By contrast, large ν (relatively diffuse prior on $\mathbf{c}$) results in a more flexible regression function posterior mean $\tilde{c}(z)$ (see Eq. (6)), which can

a

b

Fig. 12. Prior and posterior model predictions. (a) Superimposed $5e2$ likely prior soil models. (b) Superimposed $5e3$ likely posterior soil models.

accommodate wilder oscillations in its behavior. Hence, fewer basis functions are needed to reflect the true underlying process ($\mathbb{E}(k|\mathbf{d}_{obs})$ becomes increasingly small), as each basis function has more degrees of freedom. Notice that here we did not choose to set up a fixed value for v, rather this parameter is considered as a random variable (Eq. (13c)), and its value is deduced from the data such that the marginal likelihood is maximized. The marginal posterior density of the rate parameter λ in the Poisson prior (Eq. (13e)) is provided in Fig. 11c. We can see that the Bayesian point estimate for λ is closely approximated by $\hat{\lambda} \approx 5$. This parameter is the mean of the Poisson prior equation (13e), which reflects the numbers of layers k accommodated in the model $c(z)$. Fig. 12 demonstrates the essence of Bayesian updating and uncertainty reduction as a result of introducing the experimental observations. Fig. 12a presents 5×10^2 superimposed likely prior soil models (Eq. (6)), with the coefficients of each curve drawn directly from the definition of the priors (Eq. (13)). These curves show the state of minimum knowledge about the subsurface structure. No stratification and velocity measure is discernible at this initial state. Fig. 12b depicts 5×10^3 posterior soil model realizations, which mimic accurately the general trend of the target process.

Fig. 13 quantifies the observations of the previous figure. The posterior mean soil profile $\tilde{c}(z)$ (black solid curve) is illustrated together with 95% credible intervals for the posterior predictions (dark shaded area). The prior credible region is also included in the figure (light shaded area), which occupies the entire space (and extends to infinity). The mean posterior prediction of the displacement time history response of the media $v(z = 0, t)$ is pictured in Fig. 14. The figure also provides a measure of uncertainty around the posterior mean estimate $\tilde{v}(0, t)$. This plot accentuates the high fidelity of the posterior estimates to the experimental observations.

10. Conclusions

This paper introduces a probabilistic calibration approach via a Bayesian formulation for the solution of inverse problems, defined by the random field characterization of heterogeneous media, for an acoustic one-dimensional velocity field with horizontally layered structure. A self-regularized varying structure model is formulated based on the notion of Bayesian partition models in order to parameterize the acoustic wave velocity random field. The method offers a reduced dimensional inversion technique by dividing the velocity random field into an unknown number of soil
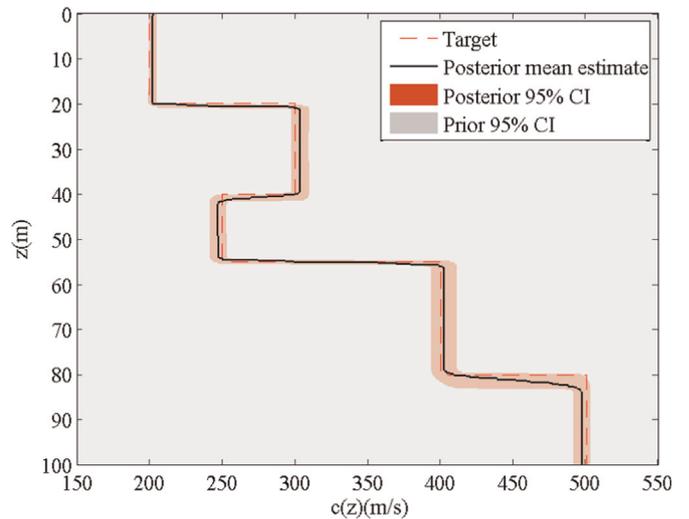
Fig. 13. Posterior mean estimate together with 95% credible intervals for the mean posterior and the mean prior.
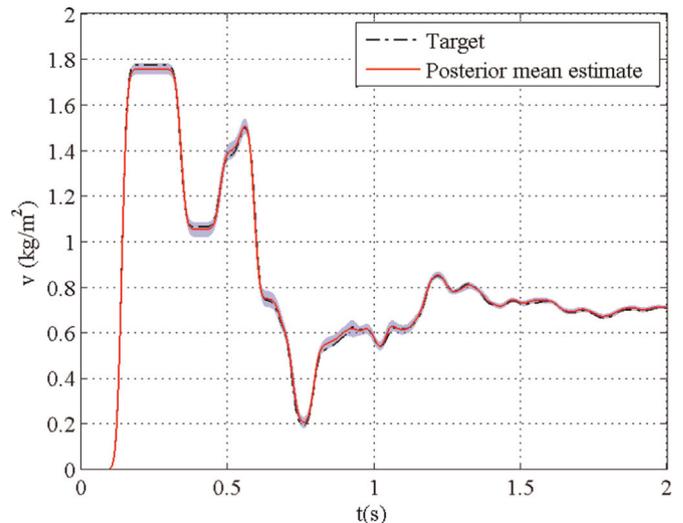
Fig. 14. Posterior displacement prediction together with 95% credible intervals around the mean.
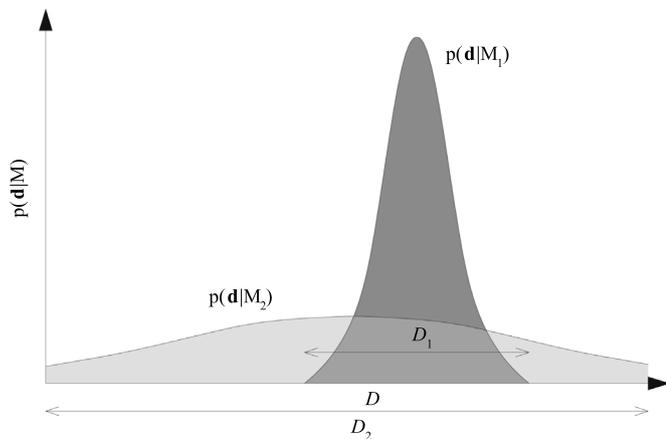
Fig. 15. A Schematic presentation of Bayesian Occam's razor.

layers within a certain depth interval. Number of layers, their velocities and thicknesses are inferred, conditioned on the observations. The reward of the approach is that the explicit regularization of the inverted profile by global damping procedures or even through imposition of priors, which carry smoothness constraints (and might introduce subjectivity to the inference process), is not required. The reversible jump MCMC algorithm was implemented to carry out the simulation of the resulting varying dimensional posterior density. The provided synthetic case indicates significant functionality of the inversion scheme to retrieve the benchmark subsurface profile.

Acknowledgments

Appendix A. Bayesian Occam's Razor

A simple explanation of why the Bayesian model selection adheres to the concept of parsimony is presented in Fig. 15. This figure illustrates the Bayesian embodiment to the latter concept [21,30,8]. The horizontal axis presents the data space, and the vertical axis shows the measure of the marginal likelihood. $\mathcal{M}_1$ and $\mathcal{M}_2$ are two competing explanations of a same process. Model $\mathcal{M}_1$ is the simpler theory, and $\mathcal{M}_2$ is the more complex one. The simple model is only capable of reaching a limited subdomain in the data space $\mathcal{D}_1$, whereas the more complex model is able to embrace a wider space due to its flexibility. As both $p(\mathbf{d}|\mathcal{M}_1)$ and $p(\mathbf{d}|\mathcal{M}_2)$ are integrated to one over the data space, if the observed data lies in $\mathcal{D}_1$ which is accessible by both the models, then $\mathcal{M}_1$ is favored over $\mathcal{M}_2$, as it assumes higher probability in this region.

References

[1] N.P. Agostinetti, A. Malinverno, Receiver function inversion by transdimensional Monte Carlo sampling, Geophys. J. Int. (181) (2010) 858–872. http://dx.doi.org/10.1111/j.1365-246X.2010.04530.x.

[2] H. Akaike, A new look at the statistical model identification, IEEE Trans. Autom. Control (19) (1974) 716–723.

[3] J.O. Berger, L.R. Pericchi, Accurate and stable Bayesian model selection: the median intrinsic Bayes factor, Sankhya Ser B (60) (1998) 1–18.

[4] B.P. Carlin, S. Chib, Bayesian model choice via Markov chain Monte Carlo methods, J. R. Stat. Soc. B Methodol. 57 (3) (1995) 473–484.

[5] S. Chib, I. Jeliazkov, Marginal likelihood from the Metropolis–Hastings output, J. Am. Stat. Assoc. 96 (453) (2001) 270–281.

[6] D.G.T. Denison, B.K. Mallick, A.F.M. Smith, Automatic Bayesian curve fitting, J. R. Stat. Soc. B Methodol. (60) (1998) 333–350.

[7] Y. Fan, S. Sisson, Reversible jump MCMC, in: Handbook of Markov Chain Monte Carlo, Chapman and Hall/CRC, Boca Raton, FL, 2011.

[8] D.G.T. Denison, C.C. Holmes, B.K. Mallick, A.F.M. Smith, Bayesian methods for nonlinear classification and regression, in: Wiley Series in Probability and Statistics, UK 2002.

[9] D.G.T. Denison, N.M. Adams, C.C. Holmes, D.J. Hand, Bayesian partition modelling, Comput. Stat. Data Anal. 38 (4) (2002) 475–485.

[10] J. Dettmer, S.E. Dosso, C.W. Holland, Model selection and Bayesian inference for high-resolution seabed reflection inversion, J. Acoust. Soc. Am. 125 (2) (2009) 706–716.

[11] J. Dettmer, S.E. Dosso, C.W. Holland, Trans-dimensional geoacoustic inversion, Geophys. J. Int. (128) (2010) 3393–3405.

[12] S.E. Dosso, Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach, J. Acoust. Soc. Am. 111 (129) (2002), http://dx.doi.org/10.1121/1.1419086.

[13] S.E. Dosso, J. Dettmer, Bayesian matched-field geo-acoustic inversion, Inverse Probl. (27) .

[14] A.J.W. Duijndam, Bayesian estimation in seismic inversion Part I: principles, Geophys. Prospect. (36) (1988) 878–898.

[15] I. Epanomeritakis, V. Akcelik, O. Ghattas, J. Bielak, A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion, Inverse Probl. 24 (034015) (2008), http://dx.doi.org/10.1088/0266-5611/24/3/034015.

[16] A.J.W. Duijndam, Bayesian estimation in seismic inversion Part II: uncertainty analysis, Geophys. Prospect. (36) (1988) 899–918.

[17] W.P. Gouveia, J.A. Scales, Bayesian seismic waveform inversion: parameter estimation and uncertainty analysis, J. Geophys. Res. Solid Earth 1 (103) (1998) 2759–2779.

[18] P.J. Green, Reversible jump Markov Chain Monte Carlo Computation and Bayesian model determination, Biometrika 82 (4) (1995) 711–732.

[19] C.C. Holmes, D.G.T. Denison, S. Ray, B.K. Mallick, Bayesian prediction via partitioning, J. Comput. Graph. Stat. (14) (2005) 811–830.

[20] C.F. Huang, P. Gerstoft, W.S. Hodgkiss, Uncertainty analysis in matched-field geoacoustic inversions, J. Acoust. Soc. Am. 119 (197) (2006).

[21] W.H. Jefferys, J.O. Berger, Ockham's Razor and Bayesian analysis, Am. Sci. (80) (1992) 64–72.

[22] J.W. Kang, L.F. Kallivokas, Mixed unsplit field perfectly matched layers for transient simulations of scalar waves in heterogeneous domains, Comput. Geosci. (14) (2010) 623–648.

[23] J.W. Kang, L.F. Kallivokas, The inverse medium problem in 1D PML-truncated heterogeneous semi-infinite domains, Inverse Probl. Sci. Eng. 18 (6) (2010).

[24] R.E. Kass, L. Wasserman, A reference Bayesian test for nested hypothesis and its relation-ship to the Schwarz criterion, J. Am. Stat. Assoc. 90 (431) (1995) 928–934.

[25] R.E. Kass, A.E. Raftery, Bayes factors, J. Am. Stat. Assoc. 90 (430) (1995).

[26] R.E. Kass, L. Wasserman, The selection of prior distributions by formal rules, J. Am. Stat. Assoc. 91 (453) (1996) 1343–1370.

[27] M.C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, J. R. Stat. Soc.: Ser. B Stat. Methodol. 63 (3) (2001) 425–464.

[28] P.S. Koutsourelakis, A multi-resolution, non-parametric, Bayesian framework for identification of spatially-varying model parameters, J. Comput. Phys. 228 (17) (2009) 6184–6211.

[29] D.V. Linley, A statistical paradox, Biometrika (44) (1957) 187–192.

[30] D.J.C. MacKay, Probable networks and plausible predictions—a review of practical Bayesian methods for supervised neural networks, Comput. Neural Syst. (6) (1995) 1056–4062.

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

[32] Z. Medina-Cetina, Probabilistic calibration of a soil model (Doctoral dissertation), The Johns Hopkins University, 2006.

[33] Z. Medina-Cetina, S. Esmailzadeh, J.W. Kang, L. Kallivokas, Bayesian inversion of heterogeneous media: introducing the next generation of integrated studies for offshore site investigations, in: Proceedings of the Offshore Technology Conference 2014, http://dx.doi.org/10.4043/24200-MS.

[34] S. Esmailzadeh, Z. Medina-Cetina, Joint states of information from different probabilistic geo-profile reconstruction methods, Georisk: Assess. Manage. Risk Engineered Syst. Geohazard. 8(3), 2014, 171–191. http://dx.doi.org/10.1080/17499518.2014.939411.

[35] B.J. Minsley, A trans-dimensional Bayesian Markov chain Monte Carlo algorithm for model assessment using frequency-domain electromagnetic data, Geophys. J. Int. (187) (2011) 252–272.

[36] S.W. Na, L.F. Kallivokas, Continuation schemes for shape detection in inverse acoustic scattering problems, Comput. Model Eng. Sci. 35 (1) (2008) 73–90.

[37] C.P. Robert, G. Casella, Monte Carlo Statistical Methods, second ed., Springer texts in Statistics, NY, USA, 2004.

[38] C.P. Robert, The Bayesian Choice from Decision-theoretic Foundation to Computational Implementation, second ed., Springer texts in Statistics, NY, USA, 2007.

[39] M. Sambridge, K. Gallagher, A. Jackson, P. Rickwood, Transdimensional inverse problems, model comparison and the evidence, Geophys. J. Int. (167) (2006) 528–542.

[40] M. Sambridge, T. Bodin, K. Gallagher, H. Tkalcic, Transdimensional inference in the geosciences, Philos. Trans. R. Soc... A (371) (2013), http://dx.doi.org/10.1098/rsta.2011.0547.

[41] J.A. Scales, L. Tenorio, Tutorial: prior information and uncertainty in inverse problems, Geophysics 66 (2) (2001) 389–397.

[42] G. Schwarz, Estimating the dimension of a model, Ann. Stat. (6) (1978) 461–464.

[43] A. Tarantola, Inverse Problem Theory and Methods for Model Parameter Estimation, SIAM, Philadelphia, USA, 2005.

[44] A. Tikhonov, Solution of incorrectly formulated problems and the regularization method, Sov. Math. Dokl. 4 (1963) 1035–1038.

[45] T.J. Ulrych, M.D. Sacchi, A. Woodbury, A Bayesian tour of inversion, Tutor. Geophys. (66) (2001) 55–69.

[46] Q.H. Vuong, Likelihood ratio tests for model selection and non-nested hypotheses, Econometrica 57 (2) (1989) 307–333.