Transdimensional and Hamiltonian Monte Carlo inversions of Rayleigh wave dispersion curves: A comparison on synthetic datasets

Mattia Aleardi[1], Alessandro Salusti[1,2], Silvio Pierini[1]

[1] University of Pisa, Earth Sciences Department, Via S. Maria 53, 56126, Pisa, Italy

[2] University of Florence, Earth Sciences Department, via. G. La Pira 4, 50121, Florence, Italy

Corresponding author: Mattia Aleardi: mattia.aleardi@unipi.it

ABSTRACT

We compare two Monte Carlo inversions that aim to solve some of the main problems of dispersion curve inversion: deriving reliable uncertainty appraisals, determining the optimal model parameterization, and avoiding entrapment in local minima of the misfit function. The first method is a transdimensional Markov Chain Monte Carlo that considers as unknowns the number of model parameters, that is the locations of layer boundaries, together with the *Vs* and the *Vp*/*Vs* ratio of each layer. A reversible jump Markov Chain Monte Carlo (rjMCMC) algorithm is used to sample the variable-dimension model space, while the adoption of a parallel tempering strategy and of a delayed rejection updating scheme improves the efficiency of the probabilistic sampling. The second approach is a Hamiltonian Monte Carlo (HMC) inversion that considers the *Vs,* the *Vp*/*Vs* ratio and the thickness of each layer as unknowns, whereas the best model parameterization (number of layer) is determined by applying standard statistical tools to the outcomes of different inversions running with different model dimensionalities. This work has a mainly didactic perspective and for this reason, we focus the attention to synthetic examples in which only the fundamental mode is inverted. We perform what we call semi-analytical and seismic inversion tests on 1D subsurface models. In the first case, the dispersion curves are directly computed from the considered model making use of the Haskell-Thompson method, while in the second case they are extracted from synthetic shot gathers. To validate the HMC and rjMCMC outcomes we analyse the

estimated posterior models and we also perform a sensitivity analysis in which we compute the model resolution matrices, posterior covariance matrices, and correlation matrices from the ensembles of sampled models. Our tests demonstrate that major benefit of the rjMCMC inversion is its capability of providing a parsimonious solution that automatically adjusts the model dimensionality. The downside of this approach is that many models must be sampled to guarantee accurate posterior uncertainty. Differently, less sampled models are required by HMC algorithm, but its limits are the computational effort related to the Jacobian computation, and the multiple inversion runs needed to determine the optimal model parameterization.

INTRODUCTION

Rayleigh wave measurements are highly sensitive to the S-wave velocity (*Vs*) and for this reason they are extensively used for geotechnical characterization or seismic site response studies (Socco and Strobbia 2004). Over the last years, the full-waveform inversion of surface waves is getting growing attention thanks to the increased computational power of modern parallel architectures (Gross et al. 2017; Xing and Mazzotti 2019). However, well-established methods still rely on dispersion curve inversion under the assumption of a 1D subsurface structure (Socco and Boiero 2008; Maraschini and Foti 2010; Cercato, 2011; Foti et al. 2018; Di Giulio et al. 2019).

The dispersion curve inversion is a highly non-linear and ill-conditioned problem. For this reason, it is crucial adopting inversion approaches that efficiently converge toward the global minimum, that allow for a straightforward introduction of a-priori model constrains, and that also provide reliable estimations of the uncertainties affecting the recovered solution (i.e. the estimation of the posterior probability density “PPD”). In this context, gradient-based inversion methods (i.e. Gauss-Newton, Steepest descent) exhibit fast convergence rates but limited capability to explore the parameter space, resulting in a final solution highly dependent on the initial model (Luke et al. 2003; Wathelet 2005; Kritikakis et al. 2014). Metaheuristic search algorithms (i.e. genetic algorithms, particle swarm optimization) exhaustively explore the model space but they usually require many forward

model evaluations and considerable computational effort to converge (Dal Moro et al. 2007; Picozzi and Albarello 2007; Sen and Stoffa 2013; Sajeva et al. 2014; Sajeva et al. 2017). In addition, both gradient-based and metaheuristic approaches do not satisfy the importance sampling principle and for this reason they hamper a reliable uncertainty appraisal (Tarantola, 2005). Markov Chain Monte Carlo (MCMC) algorithms exhibit global convergence capabilities and honour the importance sampling principle (Sen and Stoffa, 1996) but they usually require a higher computational effort with respect to linearized, gradient-based inversion algorithms.

Markov Chain Monte Carlo methods convert the inversion into a sampling problem in which the sampling density is proportional to the posterior (Sambridge and Mosegaard, 2002). The first stage of the MCMC sampling is the burn-in period in which the chain moves from a random starting model to a high-probability region. The second stage is often called the sampling stage in which the small fluctuations of the misfit value indicate that the MCMC algorithm reaches the stationary regime. Usually the samples accepted during the burn-in period do not accurately represent the target density and for this reason they are disregarded in the computation of the PPD.

Another issue of any geophysical inversion is determining the optimal model parameterization (i.e. number of layers) that guarantees a good compromise between model resolution and posterior uncertainty (Fernández-Martínez, 2015; Aleardi 2015; Menke 2018). A transdimensional MCMC inversion treats the number of model parameters as an unknown to be inferred from the data (Green 1995; Malinverno 2000; Sambridge et al. 2006). In the past years, transdimensional MCMC algorithms have been successfully applied to solve inverse problems especially at the earthquake seismology scale (Bodin and Sambridge, 2009; Bodin et al. 2012), whereas over the last few years their applications have also been extended to exploration and applied geophysics (Dadi et al. 2016; Ray et al. 2017; Mandolesi et al. 2018; Xiang et al. 2018; Galletti and Curtis 2018; Zhu and Gibson, 2018; Cho et al. 2018). In the context of surface waves inversion, it is known (e.g. Socco and Strobbia, 2004) that the chosen model parameterization (i.e. number of layers) heavily affects the outcomes of dispersion curve inversion. For this reason, it is desirable to use an inversion approach

79 that automatically adapts the underlying model parametrization and that produces solutions with an
80 appropriate level of complexity to fit the data to statistically meaningful levels. However, specific
81 MCMC recipes targeted to the problem at hand must be implemented to maintain the computational
82 effort of the Markov chain sampling affordable. More specifically, MCMC methods are primarily
83 affected by low acceptance rates and show high correlations between successively sampled models.
84 Hamiltonian Monte Carlo (HMC; Duane et al. 1987) algorithm was designed to circumvent these
85 critical issues of MCMC algorithms. The main benefits of HMC with respect to standard MCMC
86 algorithms are its rapid convergence toward the stationary regime, so that the burn-in phase is
87 drastically reduced, and its ability to make long jumps in the model space, so that the independence
88 of the sampled models is guaranteed. HMC exploits the derivative information of the target PPD to
89 focus on the most promising regions of the model space, that is regions containing more plausible
90 models. HMC was developed for problems in which the derivative of the target probability density
91 with respect to the model parameters can be computed quickly (MacKay, 2003; Bishop 2006; Neal
92 2011; Betancourt 2017). More in detail, HMC treats a model as the mechanical analogue of a
93 particle that moves from its current position (current model) to a new position (proposed model)
94 along a given trajectory (Muir and Tkalcic, 2015; Sen and Biswas 2017; Fichtner and Simuté, 2018;
95 Fichtner et al. 2019). The geometry of the trajectory is controlled by the kinetic energy (K), by the
96 mass of the particle, and by the potential energy (U) that is interpreted as the misfit function. Note
97 that the kinetic energy and the mass matrix are artificially introduced as auxiliary quantities and
98 allow for the inclusion of the derivative information of the misfit function into the sampling
99 framework. In case of optimal tuning of the hyperparameters, the acceptance rate in standard
100 MCMC sampling (such as for the random walk Metropolis) lies between 0.2-0.4 (Sambridge and
101 Moseggard, 2002), while in case of optimal tuning the acceptance rate of HMC is very close to 0.6
102 (Neal, 2011).
103 In this work we compare the transdimensional MCMC and HMC algorithms for inverting Rayleigh
104 wave dispersion curves and we discuss some of their main benefits and drawbacks. The

transdimensional MCMC has been already applied for inverting surface waves and some examples can be found in Bodin et al. (2012), Dettmer et al. (2012), and Galletti et al. (2016). Differently, this is the first application of HMC to dispersion curve inversion and to the best of our knowledge this is the first paper in which these two methods are extensively compared in solving this inverse problem. For this reason, we develop a mainly didactic perspective without including any methodological novelties to the rjMCMC and HMC approaches and by limiting the attention to synthetic data inversions in which only the fundamental mode is considered.

We apply a birth-death reversible jump MCMC algorithm (rjMCMC; Geyer and Møller, 1994), in which the amplitude of the variation of the number of dimensions is one (i.e. only one layer can be added or deleted to the current model). Both algorithms invert for the shear wave velocity (*Vs*) and the compressional wave velocity over the shear wave velocity (*Vp*/*Vs*) ratio. The rjMCM also considers the number of layers and the depth location of the layer interfaces as additional unknowns to be inverted for, whereas the HMC uses a fixed model parameterization but considers the layer thickness as un unknown parameter. To avoid biased parameter estimations, we treat the *Vp*/*Vs* ratio as an unknown, although this parameter exerts a small influence on the observed dispersion curves. The density has an even smaller influence on the observed dispersion curve than the *Vp*/*Vs* ratio, and for this reason this parameter is not inverted but fixed to a previously determined value.

We increase the efficiency of the implemented rjMCMC approach by applying a parallel tempering strategy (Sambridge, 2014) and a delayed rejection updating scheme (Bodin and Sambridge 2009). As previously mentioned, HMC requires the model parameterization (number of layers) as an input to the inversion. For this reason, we here estimate the most probable model dimensionality by making use of common statistical tools, such as the chi-squared (χ^2) probability and the Bayesian information criterion "BIC" (Schwartz 1978; Ando 2010; Sambridge et al. 2006), which are applied to models sampled in different inversion runs in which different model parameterizations are employed.

For both the rjMCMC and HMC we perform what we call semi-analytical and seismic inversion tests on 1D reference models. In the first case the observed data (dispersion curves) are semi-analytically derived from the true model, while in the seismic tests the dispersion curves are extracted from seismic shot gathers computed on the reference models by means of the reflectivity algorithm (Mallick and Fraser 1987). In all the following experiments we limit to consider the fundamental mode as the observed data, although it is known that higher modes are essentially to better constraints the solution in case of shear velocity inversions and/or high stiffness contrasts within the soil column (Feng et al. 2005; Luo et al. 2009; Cercato, 2011; Farrugia et al. 2016; Sajeva and Menanno 2017; Qiu et al. 2019). We return to this aspect in more detail in the discussion section. The extraction of the dispersion curves is always a very delicate step especially when higher modes are more energetic than the fundamental one (Xia et al. 2003; Luo et al. 2009; Boiero et al. 2011). In this work concerning synthetic data, we simply extract the observed data by picking the maxima of the fundamental mode in the frequency-phase velocity spectra. In both the semi-analytical and seismic inversions, the forward modelling is based on Haskell-Thompson (Haskell 1953) method that considers a stack of horizontal, homogeneous, layers. This method is also used in the semi-analytical tests to compute the observed data from the reference models.

For two inversion examples we quantitatively validate the rjMCMC and HMC outcomes by comparing the estimated marginal PPDs and by analysing some of the most popular sensitivity analysis tools such as the model resolution matrices, posterior model covariance matrices, and posterior correlation matrices (Menke, 2018), which are numerically computed from the ensemble of sampled models. In addition, we the relate these matrices and the estimated PPDs to the geometrical properties of the misfit function (L2 norm difference between observed and predicted dispersion curves).

In the following we start with a theoretical description of the rjMCMC and HMC approaches. Only the main mathematical aspects of the methods are described, while the interested reader is referred to pertinent references for further details. Then, we discuss the outcomes of the semi-analytical and

seismic inversion tests. Additional theoretical and practical insights into Monte Carlo inversion of geophysical data can be found in Mosegaard and Tarantola (1995), Sambridge and Mosegaard (2002), Mosegaard and Sambridge (2002), and Scalzo et al. (2019).

METHODS

Birth-death reversible jump Markov Chain Monte Carlo

The Bayes theorem can be written as:

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d})}, \quad (1)$$

or

$$p(\mathbf{m}, n|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m}, n)p(\mathbf{m}|n)p(n)}{p(\mathbf{d})}, \quad (2)$$

where equation 1 represents the Bayesian equation for inverse problems with a fixed number of unknowns, while equation 2 is the Bayesian formulation in case of parameterizations with different number of unknowns *n*. In both equations 1 and 2, **d** is the *N*-dimensional observed data vector, and **m** is the *Q*-dimensional model parameter vector, whereas the left-hand side terms represent the target PPD that could be numerically estimated from the ensemble of models sampled by the MCMC algorithm after the burn-in phase. In a transdimensional MCMC, the Metropolis-Hasting rule determines the acceptance probability α, which is the probability to move from a model **m** with dimension *n* to a model $\mathbf{m}'$ with dimension n' at a given step of the chain:

$$\alpha = p(\mathbf{m}', n'|\mathbf{m}, n) = \min\left[1, \frac{p(\mathbf{m}', n')}{p(\mathbf{m}, n)} \times \frac{p(\mathbf{d}|\mathbf{m}', n')}{p(\mathbf{d}|\mathbf{m}, n)} \times \frac{q(\mathbf{m}, n|\mathbf{m}', n')}{q(\mathbf{m}', n'|\mathbf{m}, n)} \times |\mathbf{J}|\right], \quad (3)$$

where **J** is the Jacobian of the transformation from **m** to $\mathbf{m}'$ and is needed to account for the scale changes when the transformation involves a jump between models with different dimensions; ***q***() is the proposal distribution that defines the new model $\mathbf{m}'$ as a random deviate from a probability

177 distribution $q(\mathbf{m}'|\mathbf{m})$ conditioned only on the current model $\mathbf{m}$. Note that the proposal ratio term in

178 equation 3 vanishes if symmetric proposals (for example Gaussian proposals) are employed in a

179 fixed-dimensional model space (i.e. $\mathbf{m}$ and $\mathbf{m}'$ have the same dimensions). In addition, for the

180 birth-death rjMCMC algorithm the determinant of the Jacobian is equal to 1 and thus it can be

181 conveniently ignored in computing the acceptance ratio (Bodin and Sambridge 2009).

182 We consider a Dirichlet prior distribution for interface locations and bounded and independent

183 uniform priors for *Vs* and *Vp*/*Vs* ratio. We also assume a Gaussian distributed, uncorrelated noise

184 with data covariance matrix equal to $\mathbf{C}_d$, whereas the likelihood function is based on a L2 norm

185 misfit between observed and predicted data.

186 In particular, in the rjMCMC inversion the model vector $\mathbf{m}$ includes the *Vs* and *Vp*/*Vs* ratio (in the

187 following equations often indicated with γ for notational convenience) of the n layers and the vector

188 $\mathbf{z} = [z_1, z_2, \dots, z_{n-1}]$ expressing the location of the of the $n-1$ interfaces; $\mathbf{m} = [\mathbf{Vs}, \boldsymbol{\gamma}, \mathbf{z}] = [\mathbf{e}, \mathbf{z}]$.

189 The prior for model dimension is independent of the prior of layer properties and layer partitioning

190 (distribution of interfaces depths over the depth range of interest), so that the prior model can be

191 written as:

$$p(\mathbf{m}, n) = p(\mathbf{e}|n)p(\mathbf{z}|n)p(n), \qquad (4)$$

192

193 where $p(\mathbf{z}|n)$ is the prior on layer partitioning, $p(n)$ is the prior on the number of layers, whereas

194 $p(\mathbf{e}|n)$ represents the prior on *Vs* and γ. By assuming uncorrelated layer properties, equation 4

195 becomes:

$$p(\mathbf{m}, n) = p(\mathbf{Vs}|n)p(\boldsymbol{\gamma}|n)p(\mathbf{z}|n)p(n), \quad (5)$$

196

197 The prior $p(n)$ is defined as a uniform distribution bounded in the interval $[n_{min}, n_{max}]$:

$$p(n) = \begin{cases} \frac{1}{\Delta n}, & \text{if } n_{min} \leq n \leq n_{max} \\ 0, & \text{otherwise} \end{cases}, \quad (6)$$

198

199 where $\Delta n = n_{max} - n_{min}$. A uniform bounded prior for the interface depth within the partition is

200 given by the Dirichlet distribution (Steininger et al. 2013):

$$p(\mathbf{z}|n) = n!\, z_{max}^{-n}, \quad (7)$$

201

202 where z_{max} is the maximum allowed depth position of an interface.

203 We define the prior for each layer property (*Vs* and γ) to be a bounded uniform distribution

204 between $[e_{min}, e_{max}]$:

$$p(e|n) = \begin{cases} \prod_{i=1}^{n} p(e_i|n) = \prod_{i=1}^{n} \frac{1}{\Delta e} = \frac{1}{(\Delta e)^n}, & \text{if } e_{max} \leq e \leq e_{max} \\ 0, & \text{otherwise} \end{cases}, \quad (8)$$

205

206 where e generically represents both *Vs* and γ with $\Delta e = e_{max} - e_{min}$.

207 The assumption of uncorrelated *Vs* and *Vp*/*Vs* ratio is an oversimplification that results in an
208 augmented null-space of solutions and in an decreased convergence of the sampling. However, in
209 the following applications the limited number of model parameters (always less than 12) guarantees
210 the rapid convergence of the algorithm even though an oversimplified prior model is used.

211 If we assume Gaussian distributed, uncorrelated noise with a covariance matrix $\mathbf{C}_d$, the likelihood

212 function takes the following form:

$$p(\mathbf{d}|\mathbf{m},n)=\frac{1}{\sqrt{(2\pi)^{L}|\mathbf{C}_d|}}\exp\left(-\frac{\varphi(\mathbf{m},n)}{2}\right)\propto\exp\left(-\frac{\varphi(\mathbf{m},n)}{2}\right),\quad(9)$$

where L is the number of data points, whereas $\varphi(\mathbf{m},n)=||\mathbf{C}_d^{-1/2}(\mathbf{d}-\mathbf{d_{pre}})||_2^2$ is the weighted L2 norm difference between observed and predicted data.

To increase the computational efficiency of the algorithm, we employ a parallel tempering strategy (Dosso et al. 2012; Sambridge, 2014) in which multiple and interactive chains are simultaneously run at different temperature levels T=[T_1,T_2,…,T_{max}]. High-temperature chains ensure a wide exploration of the model space, whereas low-temperature chains exploit the high-probability regions. According to stochastic criteria, swaps of models are allowed between chains at different temperatures, and in this context the high temperature chains ensure that low-temperature chains access all the high probability regions while maintaining an efficient exploitation capability. Only the models collected at T=1 are considered in the computation of the PPD because the models collected at T>1 sample a biased posterior probability density function. In this context the likelihood ratio of equation 3 becomes:

$$\frac{p(\mathbf{d}|\mathbf{m}',n',T)}{p(\mathbf{d}|\mathbf{m},n,T)}=\exp\left(-\frac{\varphi(\mathbf{m}',n')-\varphi(\mathbf{m},n)}{2T}\right).\quad(10)$$

In the approach to parallel tempering applied here, the two chains for a proposed swap are chosen at random from all chains at each iteration, and these chains are allowed to exchange their current models (or equivalently their current temperature levels). To derive the interchange probability let us consider two independent Markov chains with temperature T_i and T_j with current models $\mathbf{m}_i$ and $\mathbf{m}_j$, respectively. At a given Monte Carlo step the two chains exchange their models with a probability equal to:

$$p(i,j) = \min\left[1, \exp\left[\left(\frac{1}{T_i} - \frac{1}{T_j}\right)\left(\varphi(\mathbf{m}_i) - \varphi(\mathbf{m}_j)\right)\right]\right], \quad (11)$$

and if the swap is accepted:

$$(\mathbf{m}_i, T_i), (\mathbf{m}_j, T_j) \rightarrow (\mathbf{m}_j, T_i), (\mathbf{m}_i, T_j). \quad (12)$$

In the dispersion curve inversion, the spread of the PPD is influenced by the model parameter illumination that, for example, decreases as the depth of investigation increases. In other terms, the optimal variance of the proposal for a given model parameter (e.g. *Vs* and *Vp*/*Vs*) is expected to vary with depth. For this reason, we increase the efficiency of the implemented rjMCMC sampling by using a delayed rejection scheme: after a rejected perturbation, a second attempt is made with a different proposal that could be theoretically dependent or independent from the previously rejected model. If the second move is independent from $\mathbf{m}'$ we can simply write (Bodin and Sambridge 2009):

$$p(\mathbf{m}''|\mathbf{m}) = \min\left[1, \frac{p(\mathbf{m}''|\mathbf{d})}{p(\mathbf{m}|\mathbf{d})} \frac{q_1(\mathbf{m}'|\mathbf{m}'')}{q_1(\mathbf{m}'|\mathbf{m}')} \frac{(1 - p(\mathbf{m}'|\mathbf{m}''))}{(1 - p(\mathbf{m}'|\mathbf{m}))}\right]. \quad (13)$$

In practice we derive $\mathbf{m}''$ by perturbing $\mathbf{m}$ according to a second proposal distribution q_2 characterized by a reduced variance with respect to the first proposal q_1. To make the computation of the $p(\mathbf{m}''|\mathbf{m})$ straightforward**,** the delayed rejection scheme is here applied only to the fixed-dimension moves (see the following discussion). This strategy automatically adapts the characteristics of the proposal distribution to the spread of the PPD along different directions in the parameter space.

For even iteration numbers, we apply the following perturbation scheme:

- Property move: Randomly pick one layer and perturb its *Vs* and *Vp*/*Vs* values according to a Gaussian proposal distribution with a null mean value and a previously selected variance.

For odd iteration numbers we select one of the following updating strategies with equal probability:

- Interface move: Randomly pick one interface and slightly perturb its position using a Gaussian proposal with a zero mean and a previously selected variance. This move guarantees a small variation in the likelihood between the current and the candidate models. However, to prevent the creation of a too thin layer, if an interface is moved too close to another interface, the move is rejected.
- Birth move: Add a new interface to the model at a given vertical location. Then, pick the *Vs* and *Vp*/*Vs* values for the newly created layer from the corresponding prior models.
- Death move: choose at random one interface and remove it from the model. Then, randomly select the layer above or below the deleted interface and assign its properties to the whole vertical interval pertaining to the deleted layer.

It can be demonstrated that the previously described MCMC recipe results in an acceptance ratio for all the moves equal to (see Dosso et al. 2014 and Xiang et al. 2018 for a complete mathematical derivation):

$$p(\mathbf{m}', n'|\mathbf{m}, n) = \min\left[1, \frac{p(\mathbf{d}|\mathbf{m}', n')}{p(\mathbf{d}|\mathbf{m}, n)}\right]. \qquad (14)$$

In the following examples we consider a uniform prior for *Vs* defined over [50 m/s, 400 m/s], a uniform prior for *Vp*/*Vs* ratio defined over [1.8, 5.8], whereas the number of layers is bounded between [2, 5], for a maximum allowed depth position for an interface equal to 20 m. The variances of the proposals are properly set to have an acceptance probability between 0.2-0.4 for the chains running at $T = 1$. For each chain the starting models are randomly generated from the prior. In all the following experiments, we employ 20 tempered chains, with 10 chains at *T*=1 and the remainder

with logarithmically distributed temperature values between 1 and 100. The highest temperature values are set to obtain an acceptance probability around 0.6-0.7 for the corresponding chain. The number of iterations is set to 3000 with a burn-in period of 300. Figure 1 represents the flow-chart of the rjMCMC algorithm.

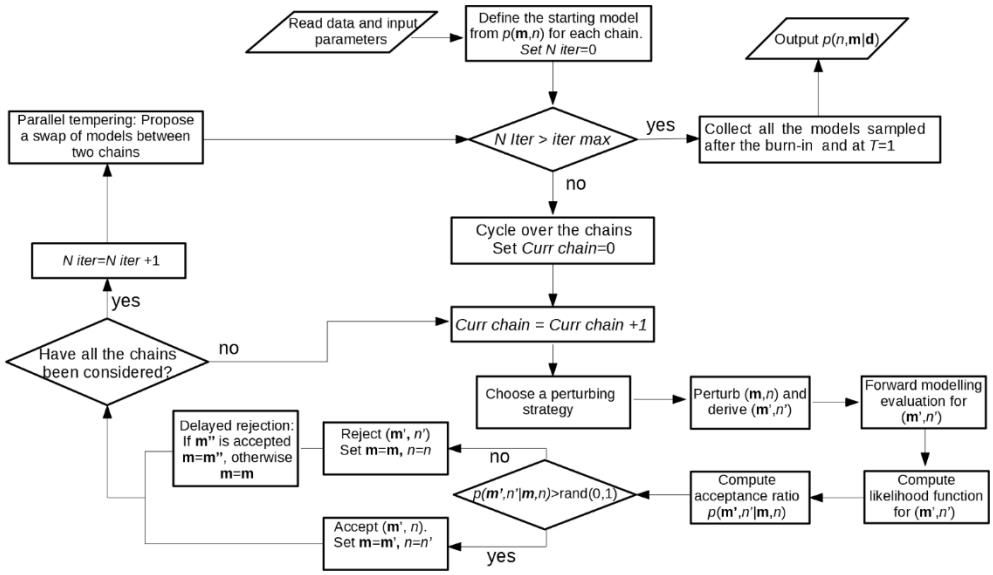


Figure 1: A schematic representation of the rjMCMC inversion. N iter, iter max, and Curr chain identify the current iteration number, the maximum number of iterations, and the considered chain, respectively.

Hamiltonian Monte Carlo

The Hamiltonian Monte Carlo is another method to numerically solve the Bayes theorem (equation 1). HMC improves upon random walk Metropolis by proposing states that are distant from the current one, but nevertheless have a high probability of acceptance (Duane et al. 1987; Neal, 1996; Neal 2011; Betancourt 2017). These distant proposals are found by numerically simulating the Hamiltonian dynamics, in which a model is viewed as a moving particle with a physical state uniquely determined by the position and momentum vectors: These vectors define the phase space. HMC samples an auxiliary distribution defined over the 2Q-dimensional phase space, from which

samples of the posterior can be obtained by ignoring the momentum space component. The particle trajectory in the phase space is determined by the potential energy (*U*), the kinetic energy (*K*) and the mass matrix (**M**). The potential energy is the negative natural logarithm of the posterior (see equation 1) or in other terms is the misfit function associated to the inverse problem. In this context more plausible models with large values of the posterior are associated to low potential energies. Generally speaking, the potential energy is given by:

$$U(\mathbf{m}) = -\ln\big(p(\mathbf{m}|\mathbf{d})\big). \quad (15)$$

HMC determines the kinetic energy by introducing an auxiliary variable (momentum variable) **p** defined over a *Q*-dimensional space. It is usually assumed that the auxiliary momentum variable has a multivariate normal distribution with zero mean and a covariance matrix equal to the so-called mass matrix:

$$K(\mathbf{p}) = N(\mathbf{p}; 0, \mathbf{M}), \quad (16)$$

where *N* represents the Gaussian distribution, and **M** is the $Q \times Q$ mass matrix that must be accurately set to ensure the convergence of the HMC algorithm (see Fichtner et al. 2019). The vectors **p** and **m** define the 2*Q*-dimensional phase space, whereas the Hamiltonian $H(\mathbf{p}, \mathbf{m})$ is the total energy of the particle:

$$H(\mathbf{p}, \mathbf{m}) = U(\mathbf{m}) + K(\mathbf{p}). \quad (17)$$

After defining the kinetic and potential energies, the Hamiltonian dynamics can be simulated. In this context, the model **m** moves through the 2*Q*-phase space according to Hamilton's equations:

$$\frac{d\mathbf{m}_i}{dt} = \frac{\partial K}{d\mathbf{p}_i}, \qquad \frac{d\mathbf{p}_i}{dt} = -\frac{\partial U}{d\mathbf{m}_i}, \quad \text{with } i = 1, 2, \ldots, Q, \quad (18)$$

where *t* indicates the artificially introduced time variable. Note that the rightmost term of equation 18 contains the partial derivative of the potential energy (i.e. the misfit function) with respect to the

315 considered model $\mathbf{m}$. This makes it possible introducing information about the gradient of the misfit

316 function into the Monte Carlo sampling framework.

317 For each current model $\mathbf{m}$, and for each iteration, HMC executes the following steps:

318 1. Randomly draw the Q momenta $\mathbf{p}$ from the normal distribution $N(\mathbf{p}; 0, \mathbf{M})$;
319 2. Derive the proposed model $\mathbf{m}(t)$ and the new momenta $\mathbf{p}(t)$ by solving Hamilton's
320 equations (18) for a given propagation time t. In this work we use the leap-frog method as
321 the numerical integration method (Betancourt, 2017);
322 3. Accept the proposed model with probability α:

323 $$\alpha = \min\left[1, \frac{\exp\left[-H(\mathbf{p}(t), \mathbf{m}(t))\right]}{\exp\left[-H(\mathbf{p}, \mathbf{m})\right]}\right]. \quad (19)$$

324 If accepted, the proposed $\mathbf{m}(t)$ point constitutes the starting model for the next trajectory
325 ($\mathbf{m} = \mathbf{m}(t)$). Otherwise, the current model $\mathbf{m}$ is again used as the starting point in the
326 following iteration;
327 4. Return to step 1.
328

329 Note that standard MCMC algorithms explore model space only slowly compared to HMC, because
330 in HMC all model parameters are updated at each iteration, so that long distances in phase space
331 can be traversed with a single move. This promotes a high level of acceptance and independence of
332 the sampled models. Differently, in standard MCMC only subsets of parameters are updated in each
333 iteration. This maintains a relatively high level of acceptance but at the expense of a high degree of
334 correlation between successively sampled models. For a more detailed comparison of MCMC and
335 HMC see, for example, Fichtner et al. (2019).
336 The model vector in the implemented HMC approach includes the *Vs*, the *Vp*/*Vs* ratio and the layer
337 thicknesses h of each layer: $\mathbf{m} = [\mathbf{Vs}, \boldsymbol{\gamma}, \mathbf{h}]$. We use equation 16 to define the momentum

distribution. Since it is known that standard HMC exhibits poor performance in sampling multimodal target densities (see the discussion section), we simply employ a Gaussian prior model. In this context the prior model can be compactly written as follows:

$$p(\mathbf{m}) = N\left(\mathbf{m};\ \mathbf{m}_{prior}, \mathbf{C}_m\right), \quad (20)$$

where $\mathbf{m}_{prior}$ and $\mathbf{C}_m$ are the a-priori mean vector and model covariance matrix that can be derived, for example, from previous knowledge about the investigated area. In the following inversion examples, we assume an uncorrelated and depth independent prior model (i.e. $\mathbf{C}_m$ is diagonal).

In this work the potential energy is defined as:

$$U(\mathbf{m}) = -\ln\left(p(\mathbf{m}|\mathbf{d})\right)$$

$$= \frac{1}{2}\left[\left(\mathbf{d} - G(\mathbf{m})\right)^T \mathbf{C}_d^{-1}\left(\mathbf{d} - G(\mathbf{m})\right) + \left(\mathbf{m} - \mathbf{m}_{prior}\right)^T \mathbf{C}_m^{-1}\left(\mathbf{m} - \mathbf{m}_{prior}\right)\right], \quad (21)$$

The partial derivative of equation 21 with respect to the model parameter vector $\mathbf{m}$ is equal to:

$$\frac{\partial U(\mathbf{m})}{\partial \mathbf{m}} = \mathbf{J}\mathbf{C}_d^{-1}\left(\mathbf{d} - G(\mathbf{m})\right) + \mathbf{C}_m^{-1}\left(\mathbf{m} - \mathbf{m}_{prior}\right), \quad (22)$$

In equations 21 and 22, G is the non-linear forward modelling that computes the dispersion curves from the current model (i.e. the Haskell-Thompson method), $\mathbf{C}_d$ is the data covariance matrix that expresses the error affecting the observed dispersion curve (under the assumption of a Gaussian distribution). $\mathbf{J}$ is the Jacobian matrix (also called Fréchet derivative matrix) that we compute with a central finite difference (FD) approach (Aster et al. 2018). We use the FD approach for its straightforward implementation but if needed other more sophisticated and more efficient methods can be used (Sen and Roy, 2003).

A proper setting of the mass matrix (**M**) is of crucial importance in any HMC sampling (see for example Fichtner et al. 2019) because it serves as a tuning parameter to adjust the speed with which the phase space is traversed, or in other words it is used to decorrelate the target distribution (Betancourt, 2017). In practice, a proper setting of this matrix maximizes the independency of the sampled models, while preventing the exploration of similar model space regions. The optimal setting of the mass matrix is strongly case dependent, but typical applications use a diagonal **M** defined as a scalar multiple of the identity matrix. However, this strategy often hampers an efficient sampling of the parameter space. In this work, the mass matrix is computed as a local approximation (around the considered model) of the inverse of the posterior covariance matrix (see Fichtner et al. 2019):

$$\mathbf{M} = \mathbf{J}^T \mathbf{C}_d^{-1} \mathbf{J} + \mathbf{C}_m^{-1}, \quad (23)$$

In the leapfrog method, we update the momentum and position variables sequentially. First, we simulate the momentum dynamics (changing momentum) by $\delta/2$ time units, then simulating the position dynamics (moving in model space) for full δ time units, then again completing the momentum simulation for the remaining half-time units, $\delta/2$, so that the momentum and model perturbations can be at the same time of δ time units. This process is repeated for a total of L times after which the algorithm reaches a new state. The leapfrog method integration is described by the following equations:

$$p_i(t + \delta/2) = p_i(t) - \frac{\delta}{2} \times \frac{\partial U}{\partial m_i}|_t, \quad (24)$$

$$m_i(t + \delta) = m_i(t) + \delta \times \frac{\partial K}{\partial p_i}|_{t+\frac{\delta}{2}}, \quad (25)$$

$$p_i(t + \delta) = p_i\left(t + \frac{\delta}{2}\right) - \frac{\delta}{2} \times \frac{\partial U}{\partial m_i}|_{t+\delta}, \quad (26)$$

where t is the time variable. Although, many other integrator methods exist (see for example Blanes et al. 2014), here we use the leapfrog method for its easy implementation.

In addition to the mass matrix, also the choice of L and δ plays an important role in the efficiency of the sampling and for this reason these parameters need to be tuned properly to get the desired acceptance rate. In particular, they determine the sampling trajectory in the augmented model space. A too short trajectory generates proposal models very close to the current position, thus slowing down the exploration of the model space. In contrast, if the trajectory is too long the algorithms revisits points in model space that have been already sampled, thereby wasting computing time. To avoid locking in periodic trajectories we follow the strategy discussed in Mackenze (1989) and in each HMC iteration we draw the L parameter randomly from a previously defined uniform distribution (see the discussion section for more details). More theoretical details about HMC inversion can be found in Neal (2011) and Betancourt (2017).

Differently from the rjMCMC, the model dimensionality (layer number) is fixed in an HMC inversion run. For this reason, we propose to determine the most probable number of layers by performing different HMC inversions with different parameterizations. Then, standard statistical tools, such as the χ^2 probability and the BIC, can be used to determine the most probable model dimensionality. In the following examples we use a single HMC chain running for 3000 iterations, for a burn-in period of 30 samples. Similarly to the rjMCMC, the starting points for the HMC inversions are randomly generated from the prior distribution. Figure 2 shows the schematic flow-chart of the implemented HMC inversion.

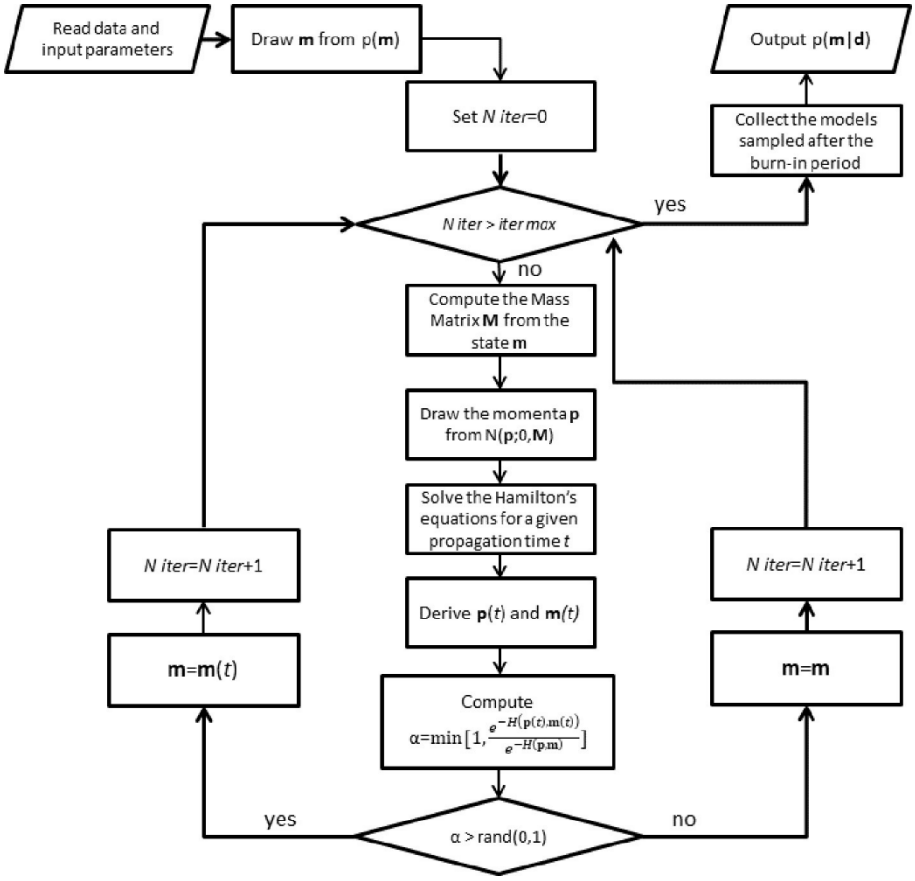


Figure 2: The schematic workflow of the HMC inversion. N iter and iter max identify the current iteration number and the maximum number of iterations, respectively.

SEMI-ANALYTICAL INVERSIONS

We represent in Figure 3 the prior models for the rjMCMC and HMC algorithms. Note that these priors are depth independent and different for the two inversion approaches: uniform for rjMCMC and Gaussian for HMC.

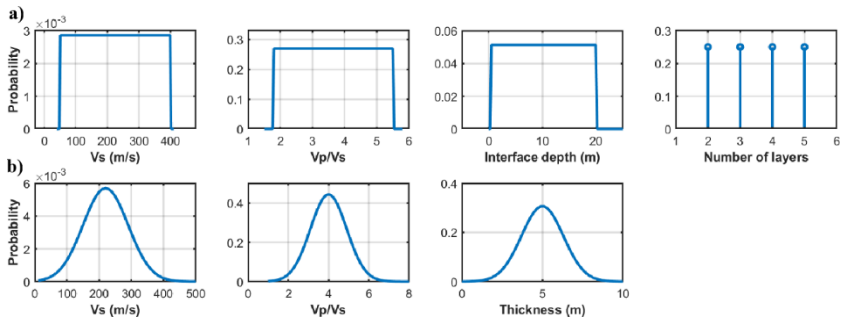


Figure 3: a) Prior distributions of Vs, Vp/Vs ratio, interface depth, and number of layers for the rjMCMc algorithm. b) Prior distributions for the HMC algorithm projected onto the Vs, Vp/Vs, and layer thickness axes.

The use of uniform and Gaussian priors highly simplifies the derivation of the acceptance ratio in the rjMCMC and of the posterior model in the HMC inversions, respectively. From a theoretical point of view the use of different priors complicates a direct comparison of the posterior models provided by the two inversion algorithms. Indeed, the posterior tends to the prior as the likelihood function becomes less informative. Despite this difference, we will see that the two methods provide PPD estimations congruent with the likelihood function and with the considered prior models. In particular, they yield similar posterior estimations for the model parameters better constrained by the data. For detailed investigation of the influence played by the prior model in a Bayesian inversion we refer the reader to Malinverno and Briggs (2004), and Theune et al. (2010).

Transdimensional Markov Chain Monte Carlo inversions

In this section we consider two schematic subsurface models with 2 and 4 layers, respectively. The frequency range of the data lies in the interval [4-30 Hz], whereas the noise standard deviation is assumed known and fixed to a value of 5 for all the frequencies.

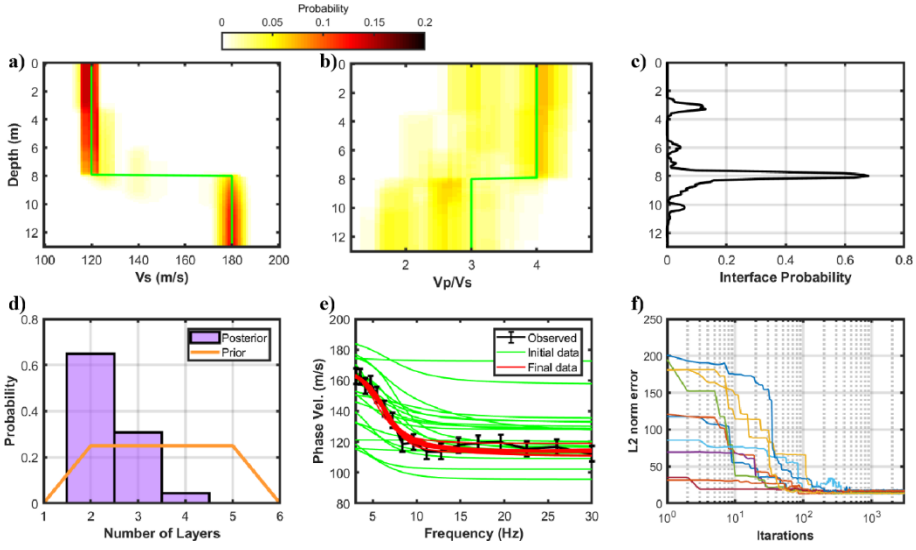


Figure 4: Semi-analytical inversion results provided by the rjMCMC algorithm for the 2-layer model. a) True Vs model (green line) and marginal PPD for Vs (colour scale). b) True Vp/Vs model (green line) and marginal PPD for Vp/Vs ratio (colour scale). c) Marginal PPD for interface location. d) Prior and posterior probability density functions for the number of layers. e)

Comparison between the observed noisy data (black line), the data generated on the starting models (green lines) and the data generated on the last sampled models (red line). The black bars represent the noise standard deviation. f) Example of evolution of the L2 norm misfit for 10 chains out of 20.

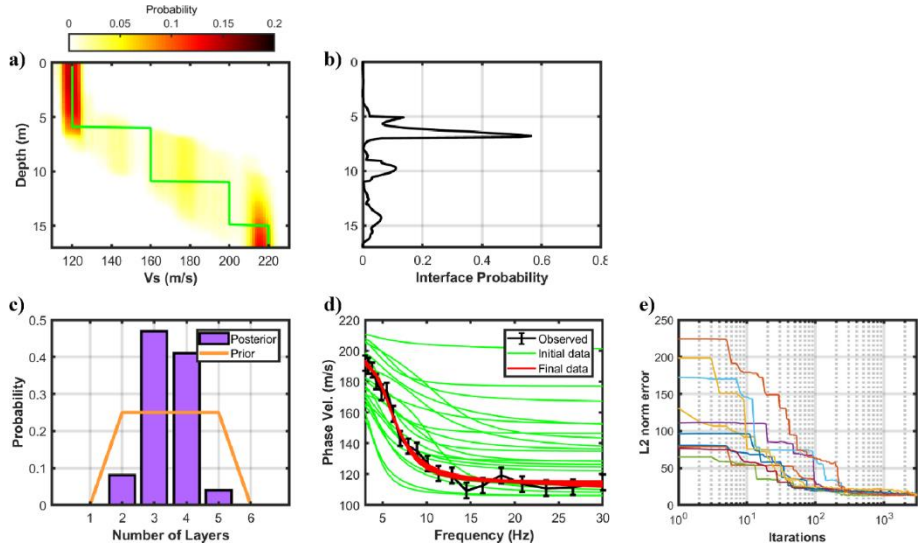


Figure 5: Semi-analytical inversion results provided by the rjMCMC algorithm for the 4-layer model. a) True Vs model (green line) and marginal PPD for Vs (colour scale). b) Marginal PPD for interface location. c) Prior and posterior probability density functions for the number of layers. d) Comparison between the observed noisy data (black line), the data generated on the starting models (green lines) and the data generated on the last sampled models (red line). The black bars represent the noise standard deviation. e) Example of evolution of the L2 norm misfit for 10 chains out of 20.

We start from a very schematic model constituted by two layers separated by an interface located at 8 m depth. The marginal PPDs of *Vs* and interface location show that the inversion recovers the interface positions, the *Vs* of the first layer and that of the half-space (Figure 4). Conversely, as expected, the estimated *Vp*/*Vs* ratio is affected by much greater ambiguity. For this reason, in many of the following examples the estimated *Vp*/*Vs* ratio is not discussed. The algorithm has reliably sampled the PPD for the number of layers and yields a maximum-a-posteriori (MAP) solution that

matches the correct value of 2. The data comparison in Figure 5d demonstrates that the algorithm successfully converges toward the so-called equivalence region of solutions. Finally, Figure 4f proves that the rjMCMC reaches the stationary regime after 200 iterations approximately, when the L2 errors of the different chains start to oscillate around stable values.

The second example considers a 4-layer model (Figure 5). As expected, we observe an increase of uncertainties with respect to the previous test, related to the increased ill-conditioning of the inversion procedure. The *Vs* of the shallowest layer and the position of the shallowest interface are recovered with high accuracy, while the uncertainties on the estimated *Vs* and on the interface positions increase for the deepest layers and deepest interfaces. The uncertainties affecting the model parameterization (Figure 5c) increase with respect to the previous example. In this case both the 3-layer and the 4-layer parameterizations exhibit similar posterior probability values, although the MAP solution erroneously indicates that 3 is the most probable number layers, thus underestimating the correct number of layers (4). This proves that in cases of more than 3 layers and if we consider the fundamental mode only, the estimation of the number of layers becomes a severely ill-conditioned problem. Differently, the inversion still provides satisfactory shear wave velocity predictions. Figure 5d proves that the models sampled by the rjMCMC algorithm successfully predict the observed data, whereas the evolution of the L2 norm data misfit (Figure 5e) shows that the stationary regime is reached after 300 iterations, approximately.

Hamiltonian Monte Carlo inversion

In this section we again consider the 2- and 4-layer models already used in the rjMCMC inversion. In the 2-layer model example our aim is to compare the uncertainties affecting the final solution when the dispersion curves lie in different frequency bands. In the first test the dispersion curve extends over [3-30 Hz], whereas in the second case the observed data lie in the interval [6-30 Hz]. Both examples consider a correct number of layers equal to 2. In Figure 6 we represent the results

for the 3-30 Hz example. The marginal PPDs of *Vs* and interface location show that the inversion correctly and accurately estimates the true model. Differently the *Vp*/*Vs* ratio is not recovered, as demonstrated by a posterior distribution still very similar to the prior with a depth-independent MAP value equal to 4 (see Figure 6b). Note the fast convergence rate and that less than 20 iterations are needed to reach the stationary regime. The data comparison in Figure 6e demonstrates that the algorithm successfully predicts the observed data. Figure 7 shows the results for the 6-30 Hz example. We observe that the *Vs* of the first layer has been recovered with the same accuracy of the previous example. Differently, the position of the interface and particularly the velocity of the deepest layer are now estimated with higher uncertainties. This is mainly related to the lack of low frequencies in the observed data that are crucial to constraint the *Vs* of the deepest layer (Socco and Strobbia, 2004).

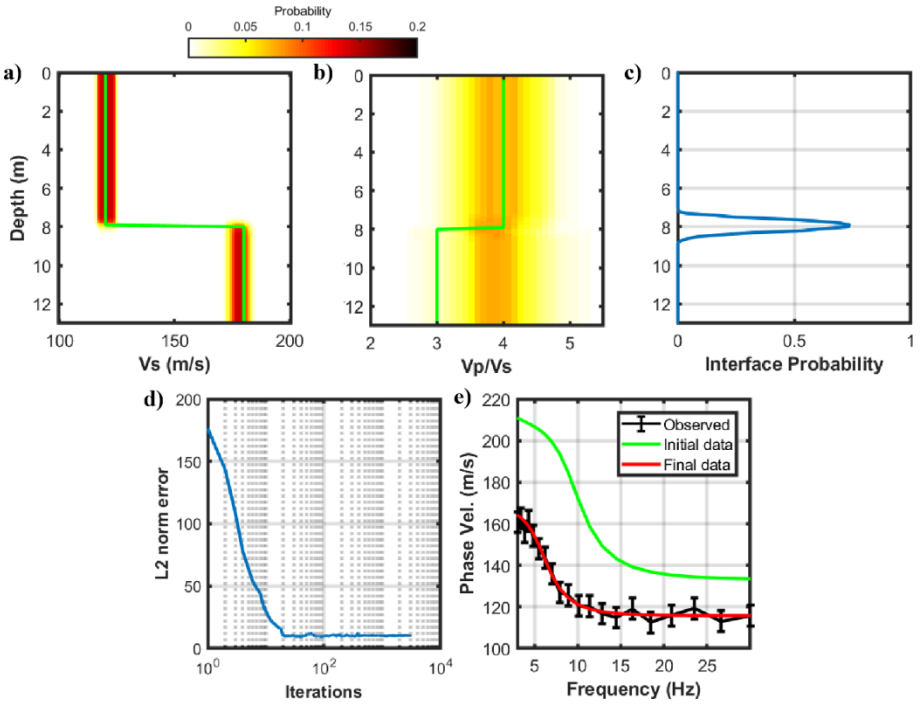


Figure 6: Semi-analytical inversion results for the 2-layer model provided by the HMC algorithm in the frequency range [3-30] Hz. a) True Vs model (green line) and marginal Vs PPD (colour scale). b) True Vp/Vs model (green line) and marginal Vp/Vs PPD (colour scale). c) Marginal PPD for interface location. d) Evolution of the L2 norm misfit. e) Comparison between the observed noisy

data (black line), the data generated on the starting model (green line) and the data generated on the last sampled model (red line). The black bars represent the noise standard deviation.

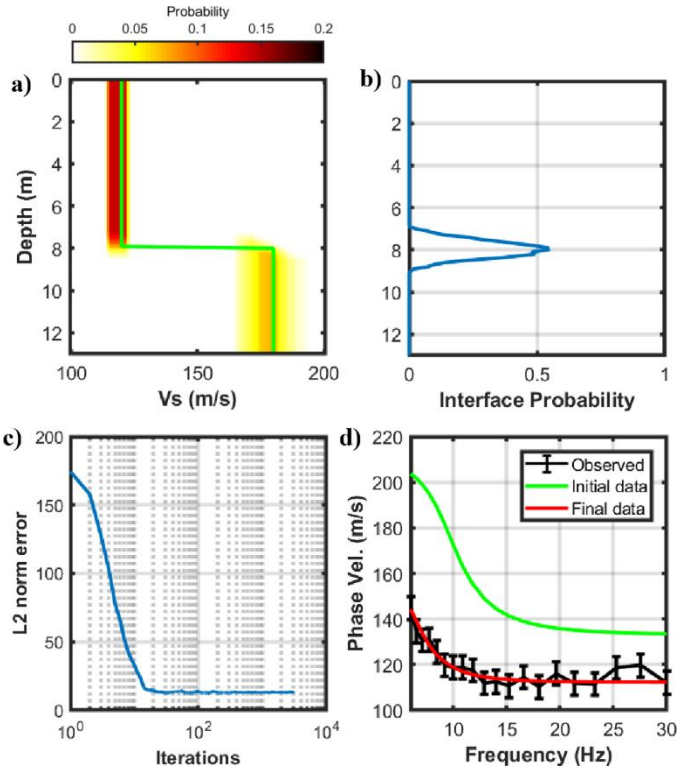


Figure 7: Semi-analytical inversion results for the 2-layer model provided by the HMC algorithm in the frequency range [6-30] Hz. a) True Vs model (green line) and marginal Vs PPD (colour scale). b) Marginal PPD for interface location. c) Evolution of the L2 norm misfit. d) Comparison between the observed noisy data (black line), the data generated on the starting model (green line) and the data generated on the last sampled model (red line). The black bars represent the noise standard deviation.

We now perform different inversion runs with different model parameterizations (i.e. different number of layers) and considering a frequency band between 4 and 30 Hz. We compute the observed data on the 4-layer model already considered in the rjMCMC inversion (Figure 5), and we invert by considering a 4- and a 3-layer parameterization. Figure 8 shows that if the correct parameterization is employed, the inversion is still able to accurately predict the *Vs* of the two shallowest layers and to correctly locate the shallowest layer boundary. As expected, the shear wave

velocities of the intermediate layers and the position of the deepest interface are recovered with high uncertainties. If we consider the 3-layer parameterization, we are still able to accurately predict the velocity of the two shallowest layers and to correctly locate the positions of the two shallowest interfaces but the position of the deepest interface and the *Vs* of the third layer are not recovered.

	$p(\chi^2)$	BIC
2-layer parameterization	0.0003	38.93
3-layer parameterization	0.0386	21.93
4-layer parameterization	0.0199	32.30

Table 1: $p(\chi^2)$ and BIC values for the semi-analytical test resulting from different model parameterizations.

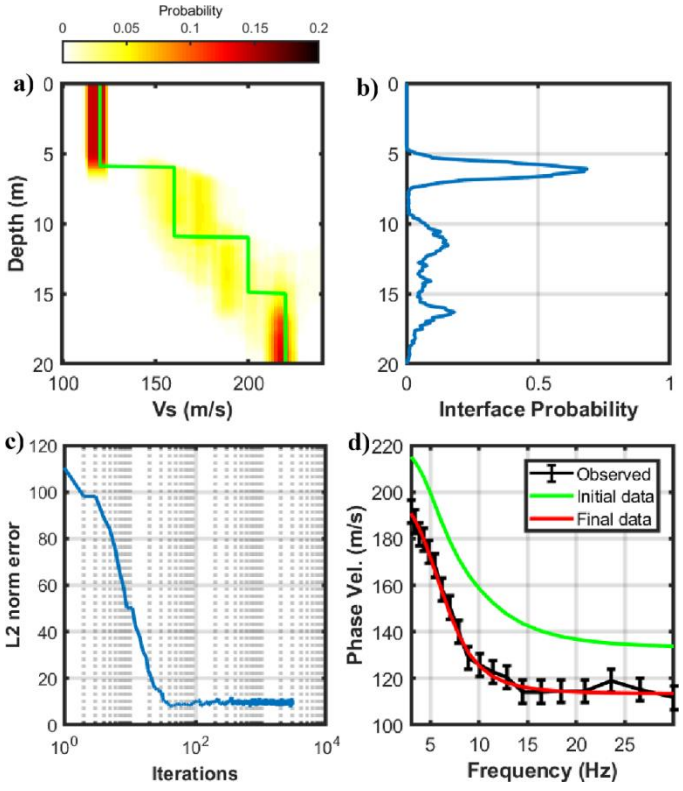


Figure 8: Semi-analytical inversion results for the 4-layer model provided by the HMC algorithm for a 4-layer parameterization. a) True Vs model (green line) and marginal Vs PPD (colour scale). b) Marginal PPD for interface location. c) Evolution of the L2 norm misfit. d) Comparison between the observed noisy data (black line), the data generated on the starting model (green line) and the

data generated on the last sampled model (red line). The black bars represent the noise standard deviation.

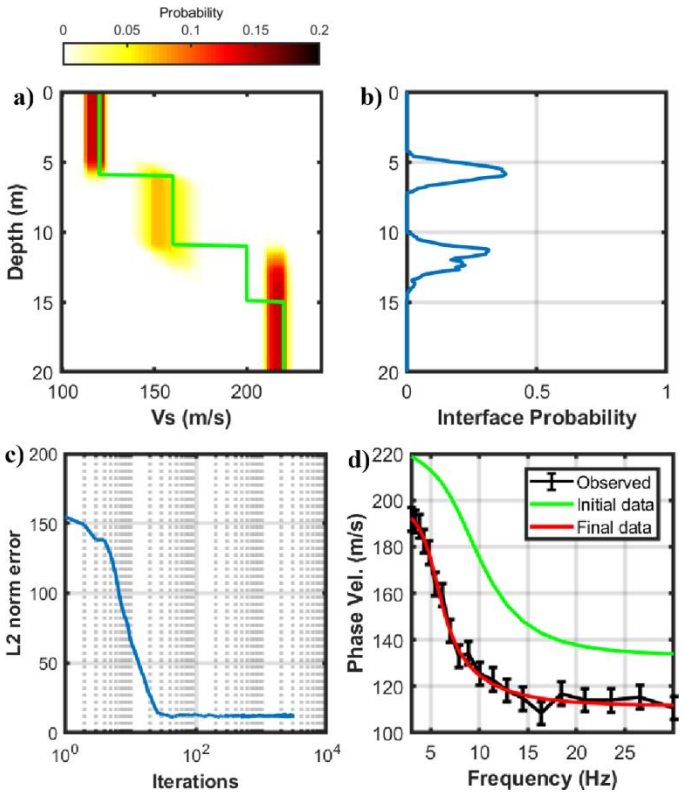


Figure 9: As in Figure 8 but for a 3-layer parameterization.

Table 1 shows the $p(\chi^2)$ and BIC values obtained in this semi-analytical test when different model parameterizations are employed. The higher $p(\chi^2)$ and lower BIC values for the 3-layer parameterization indicate that this is the most appropriate in this case. This result still proves that estimating the correct number of layers for a 4-layer model from the fundamental model only is a hopelessly ill-conditioned problem. Despite this limitation the *Vs* profile is still retrieved with reasonable accuracy. Finally, in both Figures 8c and 9c note the fast convergence of the HMC algorithm toward the stationary regime that is reached just after 20 iterations.

Sensitivity analysis of the inversion results

In this section we discuss in more detail the statistical properties of the ensemble of models sampled by the HMC and rjMCMC algorithms. For brevity we limit to a single semi-analytical test running on the previously considered 2-layer model and employing observed data lying in the same frequency range (6-30 Hz). All the algorithm hyperparameters are the same previously used in the 2-layer examples. First, we compute the misfit function of the inversion procedure that is the L2 norm difference between observed and predicted dispersion curves. We remind that this misfit function is directly related to the likelihood function of the Bayesian inversion through equation 9. Figure 10 shows some examples of 2D projections of the misfit function, in which the elongated valleys of minimum represent directions in the model space, and then model parameter combinations, that weakly affect the observed data. In other terms the elongated valleys are aligned with the eigenvectors in model space associated with the smallest singular values of the inversion kernel (Fernandez Martinez et al. 2012).

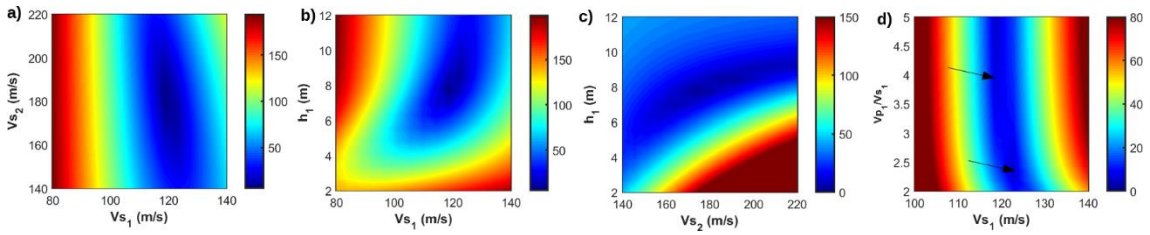


Figure 10: Some projections of the L2 norm misfit function onto different 2D sections. The subscripts 1 and 2 refer to the first (shallowest) and second (deepest) layer, respectively, whereas h indicates the layer thickness. The black arrows in d) indicate two local minima.

As expected, the contours lines of the misfit function are very different from elliptic cylinders or ellipses as it would be in a linear inverse problem (Fernandez Martinez et al. 2012). From the inspection of Figure 10 we note that the dispersion curve is mostly influenced by the *Vs* of the shallowest layer (Vs_1) , followed by the *Vs* of the deepest layer (Vs_2), and by the thickness of the first layer (h_1), while the *Vp*/*Vs* ratio plays a negligible role in determining the measured data. From the analysis of the orientation of the misfit function we can claim that the shear wave velocities of

551 the two layers independently influence the data since the axes of the global minimum valley are
552 almost parallel to the model parameter axes (Figure 10a). In Figure 10b if we focus the attention to
553 the neighbourhood of the global minimum, we see that the *Vs* and the thickness of the first layer are
554 positively correlated. The orientation of minimum valley in Figure 10c indicates a strong positive
555 correlation between the *Vs* of the second layer and the thickness of the first one. In other terms, to
556 keep the data unchanged, an increase of the thickness of the first layer must be compensated by an
557 increase of the shear wave velocity of the underlying layer. In any case all these 3 parameters (Vs_1,
558 Vs_2 and h_1 in Figure 10) are informed by the data and none of them lies in the null space of
559 solutions. This means that the posterior model for these parameters will be mainly influenced by the
560 likelihood function instead by the a-priori assumptions. For this reason, in case of an accurate
561 sampling we expect that the geometrical characteristics of the misfit function influence the
562 statistical properties of the estimated PPDs. On the contrary. we observe in Figure 10d that the
563 *Vp*/*Vs* ratio is not illuminated by the data as the misfit function shows a very elongated valley with
564 two minima: one local minimum for a *Vp*/*Vs* ratio equal to 2.4, and the global minimum
565 corresponding to *Vp*/*Vs* ratio of the reference model equal to 4. In this case the two axes of the
566 minimum valley are aligned with the two considered model parameters and this means that the *Vs*
567 and the *Vp*/*Vs* of the first layer independently influence the observed dispersion curve. Similar
568 conclusions could have been drawn about the resolvability of the *Vp*/*Vs* ratio of the second layer
569 (not shown here for brevity). For this reason, we conclude that the *Vp*/*Vs* ratio lies in the null space
570 of solutions and for this reason we expect that the associated PPDs will be mainly determined by
571 the a-priori assumptions in case of an informative prior model (i.e. the Gaussian prior used by the
572 HMC algorithm), or by the likelihood function for an uninformative prior (i.e. the uniform prior
573 used by rjMCMC).

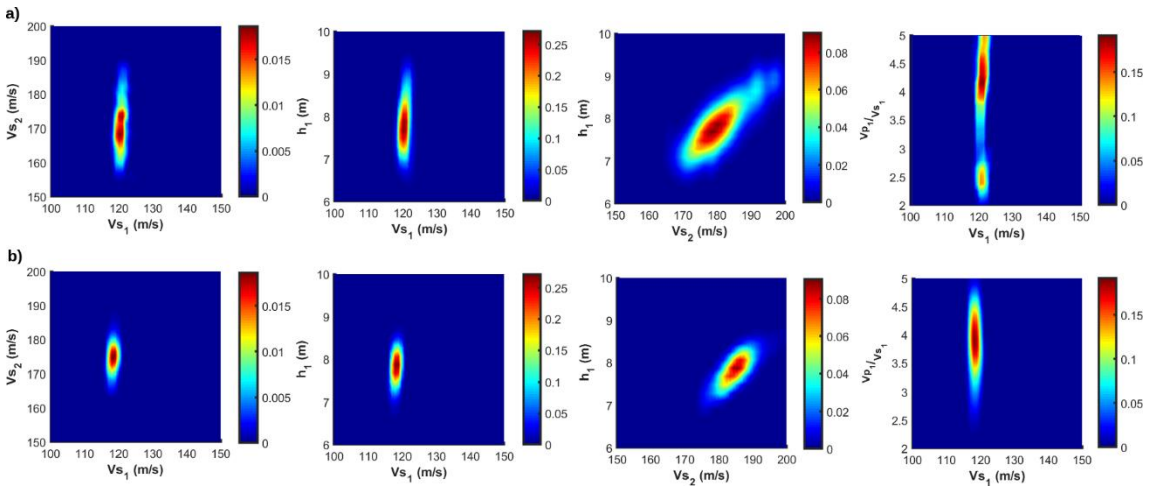


Figure 11: 2D Marginal PPDs estimated by the rjMCMC (a) and by the HMC (b) algorithms.

Figure 11 shows 2D marginal PPDs projected onto the same sections previously considered in Figure 10. Both the posterior models estimated by the rjMCMC and HMC algorithms confirm that the *Vs* of the first layer is the parameter mostly illuminated by the data, followed by the *Vs* of the second layer and the thickness of the first layer. As expected, the posterior estimations of these parameters are mainly driven by the likelihood function and for this reason both HMC and rjMCMC provide similar PPD predictions, although the informative, Gaussian prior adopted by HMC often gives narrower posterior estimations with respect to rjMCMC. Differently the estimated PPDs for the *Vp*/*Vs* ratio of the first layer are substantially different, because mainly influenced by the prior model. Indeed, the HMC provides a Gaussian posterior centred on the prior mean value of 4, while the uninformative prior used by the rjMCMC results in a posterior model very similar to the likelihood. In this case, note that two local maxima of the posterior model (Figure 11a rightmost plot) correspond to the two local minima of the misfit function (Figure 10d). The model parameter correlation can also be correctly inferred from the sampled models. Indeed Figure 11 clearly shows the independence of the *Vs* of the two layers, the weak positive correlation between the *Vs* and the thickness of the first layer, the independence of the *Vs* and the *Vp*/*Vs* ratio of the firs layer, and the strong positive correlation between the *Vs* of the second layer and the thickness of the underlying layer.

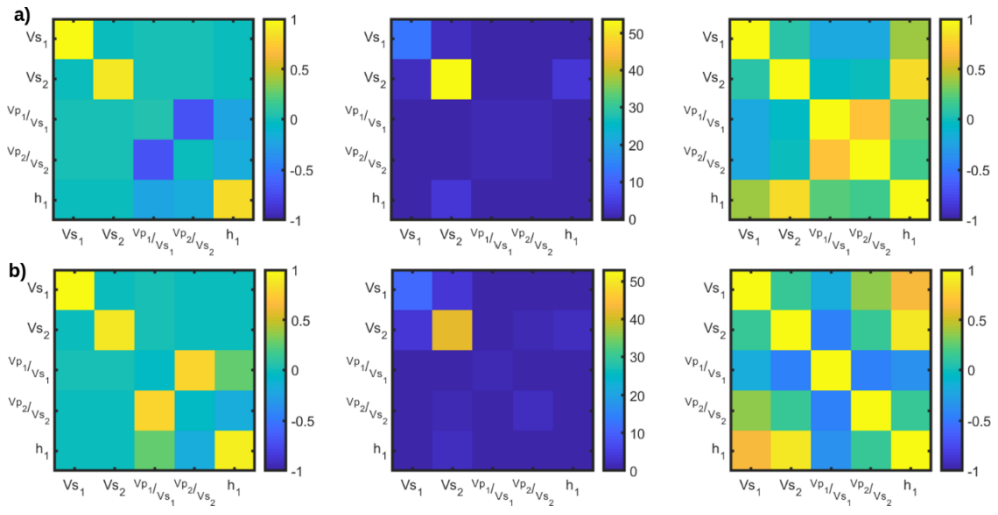


Figure 12: From left to right we represent: non-dimensional model resolution matrix, posterior model covariance matrix, and posterior model correlation matrix. a) and b) represent the results for the rjMCMC and HMC algorithms, respectively.

For a further quantitative quantification of the results provided by the two sampling methods we use the ensemble of sampled models to compute some of the most popular sensitivity analysis tools such as the model resolution matrix and the posterior model covariance matrix. Since the inversion considers different parameter types (seismic velocity and thickness) the off-diagonal elements of the model resolution matrix and of the posterior covariance matrix become dimensionally dependent, thus complicating their visual inspection. For this reason, we resort to scale these matrices with respect to the prior and posterior standard deviation of the model parameters, thus deriving the non-dimensional model resolution matrix and posterior correlation matrix (see Sambridge, 1999). For example, the values corresponding to the *i*-th row and *j*-th column of the non-dimensional model resolution matrix is derived as:

$$\hat{R}_{i,j} = R_{i,j}\frac{\sigma_i}{\sigma_j} \quad , \quad (27)$$

where σ_i and σ_j are the a-priori standard deviations of the *i*-th and *j*-th model parameter respectively, whereas R and $\hat{R}$ are the standard and non-dimensional model resolution matrices, respectively. The posterior correlation matrix C can be computed as follows:

$$C_{i,j} = \frac{C^M{}_{i,j}}{\sqrt{C^M_{i,i} C^M_{j,j}}} \quad , \quad (28)$$

where $\boldsymbol{C}^M$ is the posterior model covariance matrix. Figure 12 demonstrates that the rjMCMC and HMC provide similar non-dimensional model resolution matrices, posterior covariance matrices and posterior model correlation matrices especially for the parameters better illuminated by the data (*Vs* and layer thickness). We observe an almost perfect resolution on the *Vs* of the two layers and a slightly lower resolution for the layer thickness, while the *Vp*/*Vs* ratios are not resolved. Due to the different parameter dimensionality the analysis of the posterior covariance matrix only reveals that the *Vs* of the first layer is more accurately estimated than that of the second layer. The posterior model correlation matrices show the almost null correlation among the shear wave velocities of the two layers, the weak positive correlation between the *Vs* and the thickness of the first layer, the null correlation between the *Vp*/*Vs* ratio and the *Vs* of the first layer, and the strong positive correlation between the *Vs* of the second layer and the thickness of the first one.

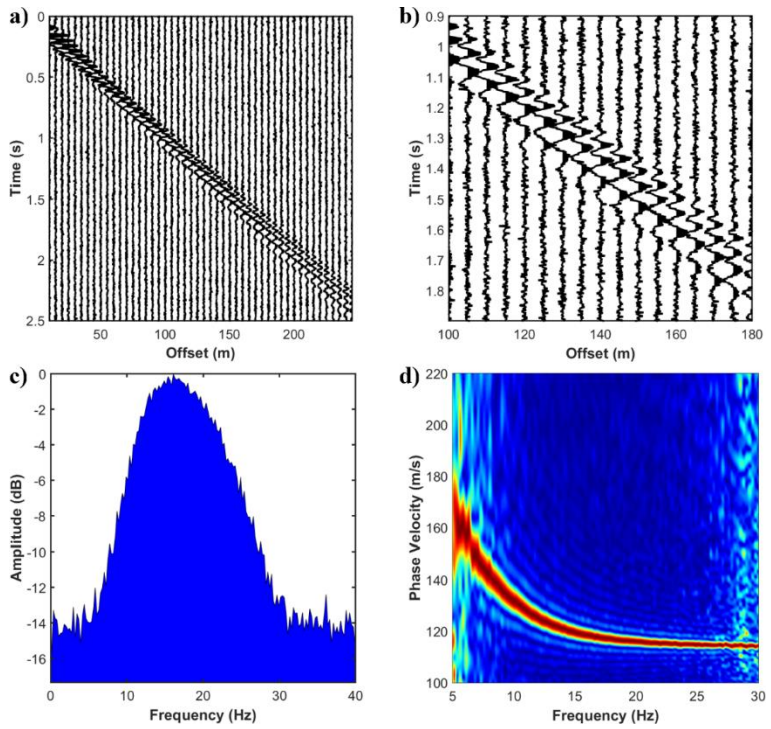
For the sake of clarity, we point out that for the HMC algorithm the matrices shown in Figure 12 have been derived from the ensemble of models sampled after the burn-in period. Differently, not all the models sampled by rjMCMC have the same dimensionality and for this reason the matrices shown in Figure 12 have been computed from the sampled models with two layers, that corresponds to the estimated MAP solution for the layer number.

Finally, this analysis confirms that although the rjMCMC and HMC algorithms use different model parameterizations and prior models, they not only achieve comparable and congruent estimates of

Vs, and interface positions but also provide the same insight into the model parameter correlation and resolution, at least for the parameters better constrained by the data. Indeed, for these parameters the posterior sampling is mainly driven by the likelihood function instead of by the a-priori constraints.

SEISMIC INVERSIONS

In the following experiments the observed dispersion curves have been picked on the frequency-phase velocity spectra derived from synthetic shot gathers. For the sake of conciseness, we limit the attention to two examples pertaining to two different 3-layer reference models. The first example concerns a subsurface model with a steady increase of *Vs* with depth. The second example is more challenging because it considers a model in which the *Vs* of the second layer is lower than that of the first layer. This velocity inversion complicates the picking of the fundamental mode because higher modes become more energetic than the fundamental mode at high frequencies. In both cases we employ the reflectivity algorithm as the forward modelling, where the source signature is a 15-Hz Ricker wavelet with a sampling interval of 2 ms. We simulate an off-end acquisition geometry with a minimum offset of 10 m and 48 receivers equally spaced of 5 m, resulting in a maximum offset of 245 m. Before the picking procedure we add to the synthetic seismic data, Gaussian random noise with a standard deviation equal to the 100 % of the standard deviation of the noise-free seismic dataset. As a result, differently from the previous inversion tests, the noise standard deviation contaminating the observed dispersion pattern is not constant over the frequency range but decreases as the frequency increases. For this reason, the diagonal entries of the data covariance matrix are now related to the ambiguity affecting the picking of the dispersion curve. The prior distributions are the same previously used in the semi-analytical tests.



654

655 *Figure 13: Data associated to the first 3-layer reference model used in the seismic tests. a) Noise*
656 *contaminated shot gather. b) Close-up of a). c) Fourier amplitude spectra of a). d) Phase velocity*
657 *spectra derived on a). In d) blue and red colors code low and high amplitude values, respectively.*

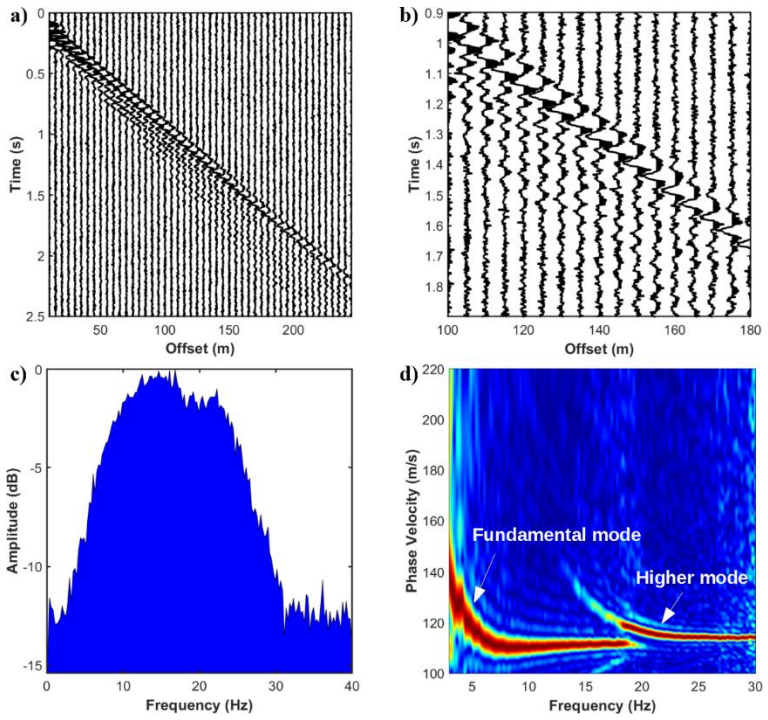


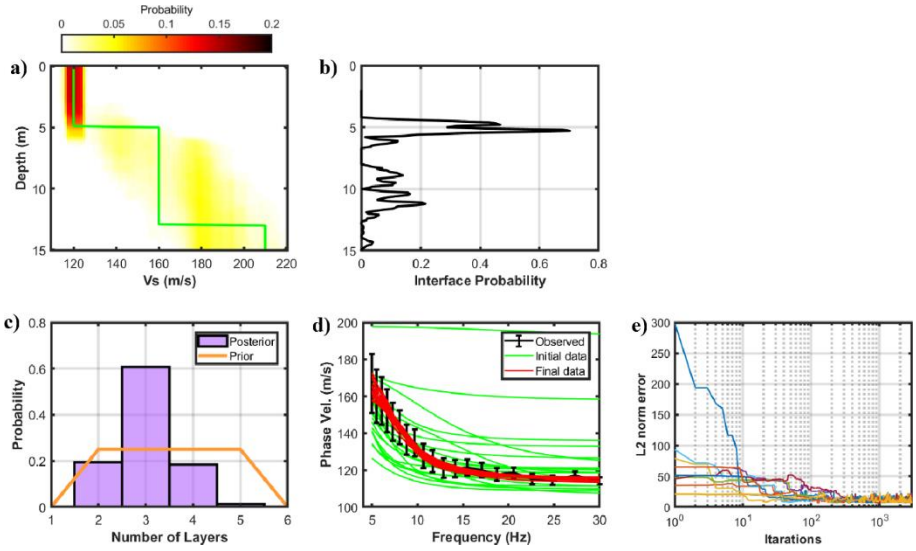
Figure 14: Data associated to the second 3-layer reference model. a) Noise-contaminated shot gather. b) Close-up of a). c) Fourier amplitude spectra of a). d) Phase velocity spectra derived on a). In d) blue and red colors code low and high amplitude values, respectively.

Figures 13 and 14 show the shot gathers, their Fourier amplitude spectra and the frequency-phase velocity spectra. In Figure 13 note that the steady *Vs* increase with depth makes it possible an accurate identification of the fundamental mode over the entire frequency range [5-30 Hz], but especially at high frequencies. Differently in Figure 14, the velocity inversion generates higher modes that become dominant for frequencies higher than 20 Hz. For this reason, in this case the rjMCMC and HMC inversions run in a narrower frequency band ranging from 5 to 20 Hz.

Transdimensional Markov Chain Monte Carlo inversion

Figure 15 shows the rjMCMC inversion results for the first 3-layer model. The *Vs* of the shallowest layer and the position of the first interface are well resolved. As expected, the uncertainties rapidly increase with depth. The overall loss of accuracy on the *Vs* and interface position with respect to the semi-analytical inversions is obviously produced by the loss of low frequencies (frequency lower

674 than 5 Hz) and by the higher uncertainties contaminating the picked dispersion curve at low
675 frequencies. Figures 15c and 15d prove that the algorithm, even in this not favourable scenario,
676 correctly identifies the correct number of layers and produces final dispersion curves that match the
677 observed data. The evolution of the L2 norm error shows that the stationary regime is reached after
678 200 iterations, approximately.



679

680 *Figure 15: Seismic inversion results for the first 3-layer model. a) True Vs model (green line) and*
681 *marginal PPD for Vs (colour scale). b) Marginal PPD for interface location. c) Prior and posterior*
682 *probability density functions for the number of layers. d) Comparison between the picked*
683 *dispersion curve (black line), the data generated on the starting models (green line), and the data*
684 *generated on the last sampled models (red line). The black bars represent the noise standard*
685 *deviation. e) Example of evolution of the L2 norm misfit for 10 chains out of 20.*

686

687 Figure 16 displays the results for the second 3-layer model. We note that in spite of the more
688 complex dispersion pattern generated by the shear-wave velocity inversion, the algorithm correctly
689 estimates the velocity of the first two shallowest layers, and the velocity inversion occurring at 6 m
690 depth, while the uncertainties significantly increase for the *Vs* of the deepest layer. The layer
691 interfaces are still accurately located, and the number of layers is correctly determined. The

comparison of the picked and finally sampled dispersion curves confirms the convergence of the algorithm to the stationary regime that is attained after 200 iterations, approximately.

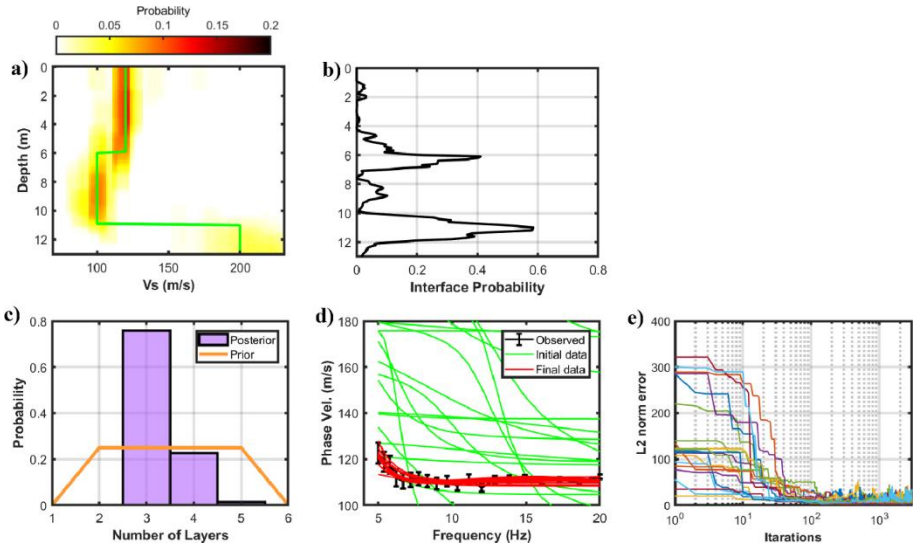


Figure 16: As in Figure 15 but for the second 3-layer model.

Hamiltonian Monte Carlo inversion

We now discuss the results obtained by the HMC algorithm on the data generated from the two considered 3-layer models. Figure 17 represents the outcomes of the first example without the velocity inversion. If we adopt a 3-layer parameterization we accurately predict the *Vs* of the shallowest layer and the position of the first interface. The accuracy of the results rapidly decreases as the depth increases, although the shear velocity of the second layer is still predicted with acceptable accuracy, while the *Vs* of the third layer and the position of the second interface are slightly underpredicted. This is obviously related to the lack of low frequencies in the data and to the severe noise contamination at low frequencies that produces high uncertainties in the picking phase (see Figure 14). However, the overall *Vs* trend and the positions of the layer interfaces are still recovered with reasonable accuracy. The evolution of the data misfit shows the rapid convergence of the HMC algorithm toward the stationary regime, that is attained in less than 30 iterations. Again, the final predicted data shows a good match with the picked dispersion curve.

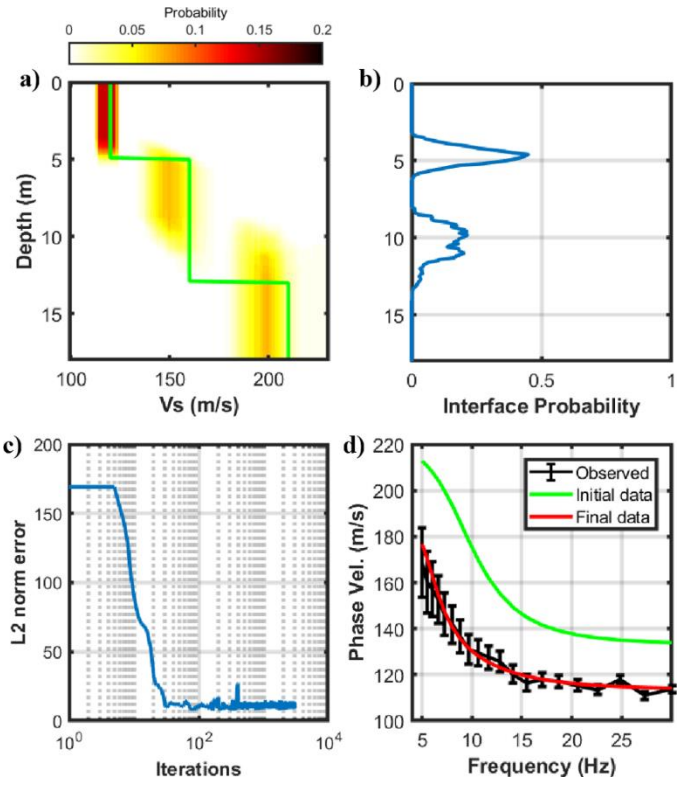


Figure 17: Seismic inversion results for the first 3-layer model and considering a 3-layer parameterization. a) True Vs model (green line) and marginal Vs PPD (colour scale). b) Marginal PPD for interface location. c) Evolution of the L2 norm misfit. d) Comparison between the picked dispersion curve (black line), the data generated on the starting model (green line) and the data generated on the last sampled model (red line). The black bars represent the noise standard deviation.

Similarly to the semi-analytical inversion tests, we now rerun the HMC inversion on the same data but employing a 4-layer parameterization. The results are shown in Figure 18. The *Vs* of the shallowest layer and the position of the first interface are accurately predicted and also the *Vs* of the second layer is well recovered, while high uncertainties affect the estimated positions of the deepest interfaces and the *Vs* of the deepest layer. Spurious interfaces are predicted by the algorithm below 15 m depth. Similarly to the previous example, less than 30 iterations are enough to converge toward the stationary regime in which the predicted data closely match the dispersion curve extracted from the seismic shot gather.

Table 2 demonstrates that our inversion workflow correctly identifies the correct number of layers in this first seismic inversion and proves that the statistical tools we use ($p(\chi^2)$ and BIC) could constitute a valid help to select the optimal model parameterization.

	$p(\chi^2)$	BIC
2-layer parameterization	0.0092	38.93
3-layer parameterization	0.0342	25.20
4-layer parameterization	0.0105	31.30

Table 2: $p(\chi^2)$ and BIC values for the first seismic inversion resulting from different model parameterizations.

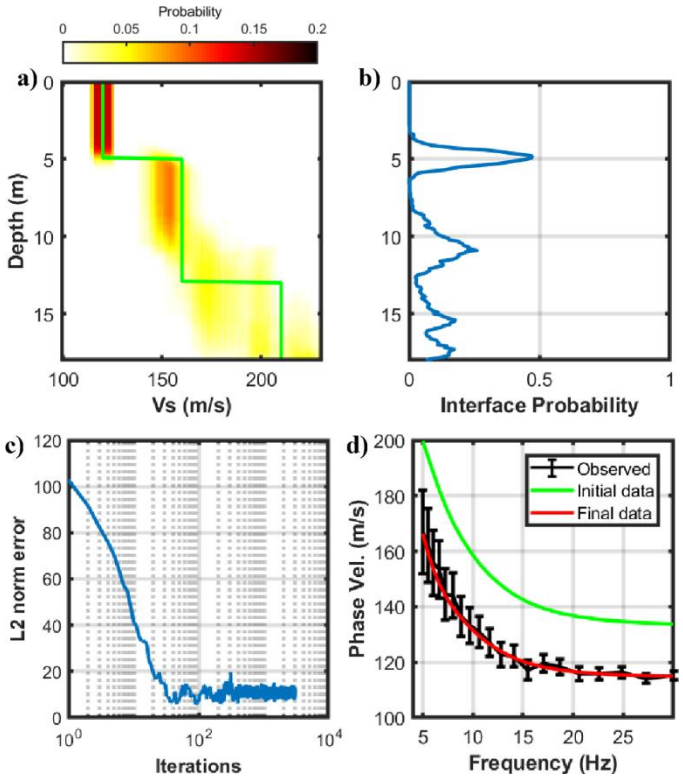


Figure 18: As in Figure 17 but for a 4-layer parameterization.

If we compare Figure 17 with Figure 15, we note that HMC provides more accurate predictions (affected by lower uncertainties) with respect to the rjMCMC algorithm, especially for the *Vs* of the two deepest layers and the position of the deepest interface. Probably, conversely to the uninformative (i.e. uniform) priors used by the rjMCMC, the Gaussian prior employed by the HMC

method efficiently contributes to reduce the ensemble of possible solutions. In this case the a-priori constraints play a crucial role in determining the posterior model due to the ill-conditioning of the inverse problem that is mainly related to the narrow frequency band and to the noise contamination at low frequency.

Figure 19 displays the results for the second model and for a 3-layer parameterization. The *Vs* of the two shallowest layers are well resolved, and despite the velocity inversion the algorithm accurately predicts a low velocity layer located between 6 m and 11 m depth. The uncertainties in the estimated *Vs* rapidly increase as the depth increases and the velocity of the deepest layer is not well recovered with a MAP solution that significantly underestimates the actual velocity value. Figure 19b demonstrates that the HMC algorithm correctly identifies the position of the interfaces, whereas Figure 19c shows that in 30 iterations the algorithm converges toward the stationary regime, after which the observed data is fairly reproduced by the sampled models (Figure 19d).

	$p(\chi^2)$	BIC
2-layer parameterization	8×10^{-7}	48.92
3-layer parameterization	0.076	36.61
4-layer parameterization	0.021	41.42

Table 3: $p(\chi^2)$ and BIC values for the second synthetic test resulting from different model parameterizations.

Figure 20 illustrates the results obtained on the same dataset but with a 2-layer parameterization. The velocity of the shallowest layer is still accurately retrieved, but the *Vs* of the deepest layers and the interface locations are not correctly estimated and show MAP solutions affected by significant biases. The evolution of the L2 norm error value demonstrates that in this case more iterations are needed to converge (for this reason we adopt only for this test a burn-in period of 150 samples), whereas the data generated on the final sampled model show some discrepancies with respect to the picked dispersion curve especially at low and high frequencies. Again, the $p(\chi^2)$ and BIC values

listed in Table 3 confirm that the proposed HMC workflow correctly estimates the actual number of layers.

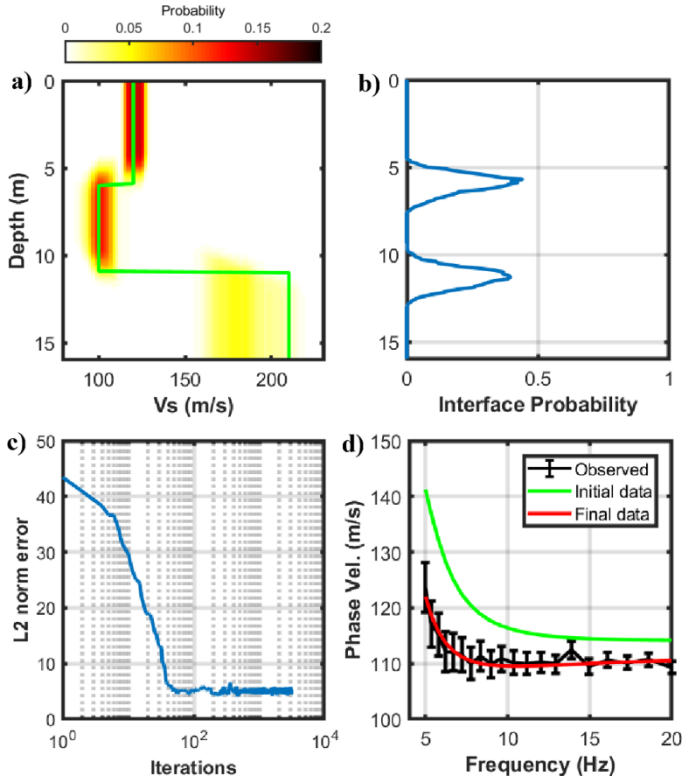


Figure 19: Seismic inversion results for the second 3-layer model and considering a 3-layer parameterization. a) True Vs model (green line) and marginal Vs PPD (colour scale). b) Marginal PPD for interface location. c) Evolution of the L2 norm misfit. d) Comparison between the picked dispersion curve (black line), the data generated on the starting model (green line) and the data generated on the last sampled model (red line). The black bars represent the noise standard deviation.

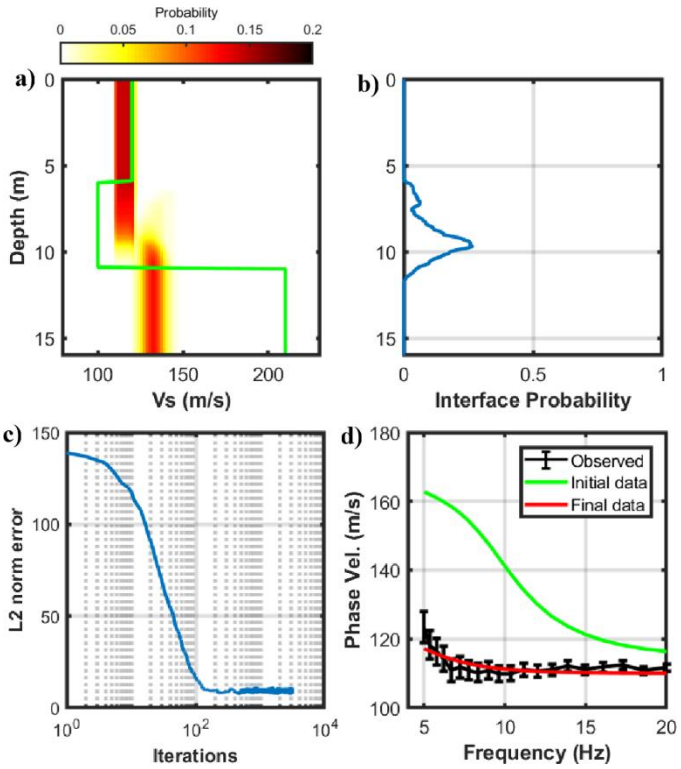


Figure 20: As in Figure 19 but for a 2-layer parameterization.

If we compare Figure 16 and Figure 19, we note that the rjMCMC and HMC algorithms provide congruent and comparable PPDs for the shear wave velocity and interface locations. This means that the posterior model for these parameters is mainly influenced by the observed data than by the a-priori constraints infused into the inversion framework. This aspect is deepened in the following section.

Sensitivity analysis of the inversion results

We now discuss the results of the sensitivity analysis for the second 3-layer model. Differently from the semi-analytical examples, we limit our discussion to the model parameters that can be inferred from the data (i.e. the *Vp*/*Vs* ratio is not discussed). Figure 21 shows some 2D projections of the L2 norm misfit function. We observe that the shear wave velocities of the first two layers (Vs_1 and Vs_2) are the parameters that mostly influence the dispersion curve, followed by the layer thicknesses (h_1 and h_2) and by the *Vs* of the deepest layers (Vs_3). Between the considered parameters, the velocity of the third layer is the one that plays the minor role in determining the dispersion pattern.

From the orientation of the global minimum valleys we observe positive correlations between the *Vs* and the thickness of the second layer and between the *Vs* of the third layer and the thickness of the second one (Figures 21b and 21c), while the shear wave velocities and the thicknesses of the first two shallowest layers are negatively correlated (Figures 21a and 21d).

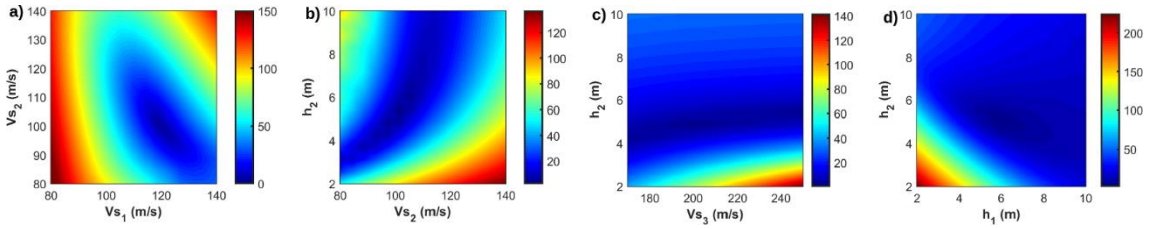


Figure 21: Some projections of the L2 norm misfit function onto different 2D sections. The subscripts 1 and 2 refer to the first (shallowest) and second (deepest) layer, respectively, whereas h indicates the layer thickness.

Figure 22 shows 2D marginal PPDs estimated by the HMC and rjMCMC algorithms and projected onto the same sections previously considered in Figure 21. For both the rjMCMC and HMC inversions emerges that the velocities and the thicknesses of the first and second layer are well recovered, while higher uncertainties affect the estimated *Vs* of the deepest layer. We also observe a positive correlation between Vs_2 and h_2, and between Vs_3 and h_2, while negative correlations characterize the marginal PPDs projected onto the Vs_1-Vs_2 and h_1-h_2 planes.

The non-dimensional model resolution matrices (left hand side of Figure 23) confirm that the *Vs* of the two shallowest layers are the best resolved model parameters, while lower resolutions characterize the *Vs* of the third layer and the layer thicknesses. The mutual parameter correlations expressed by the 2D projections of the marginal PPDs can be also inferred from the inspection of the posterior correlation matrices (right hand side of Figure 23). The marginal PPDs displayed in Figure 22 and the non-dimensional model resolution matrices and posterior model correlation matrices of Figure 23 are in good agreement with the geometrical characteristics of the misfit

function (Figure 21). This proves that both the implemented algorithms accurately estimate the uncertainties affecting the model parameters and their mutual correlations.

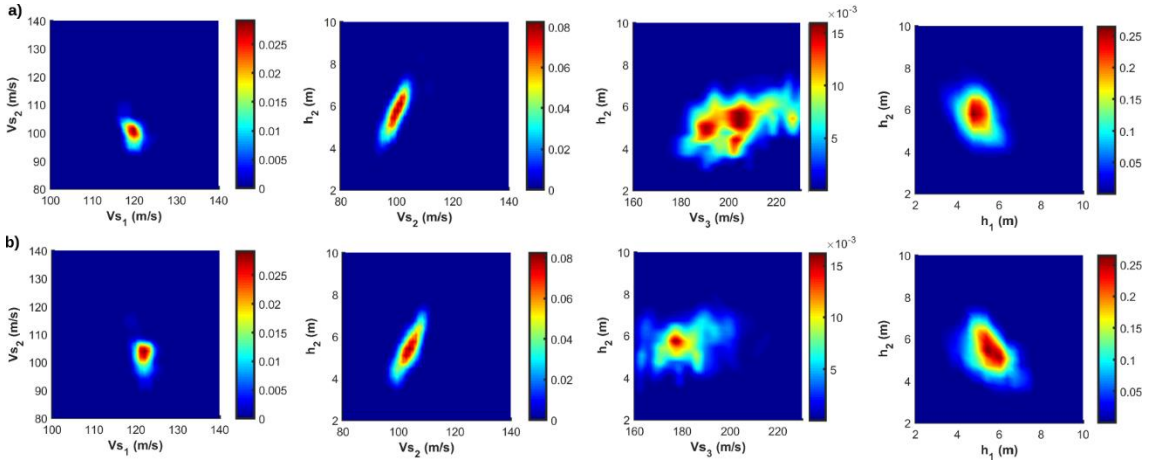


Figure 22: 2D Marginal PPDs estimated by (a) the rjMCMC and by (b) the HMC algorithms.

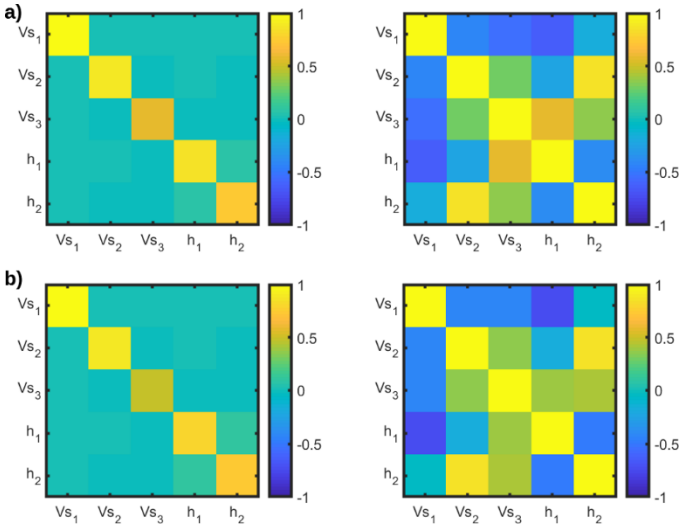


Figure 23: From left to right we represent: non-dimensional model resolution matrices, and posterior correlation matrices provided by the rjMCMC (a) and HMC(b) algorithms.

DISCUSSION

The main advantage of the rjMCMC algorithm is the possibility to provide stable and reliable predictions with a minimum of a-priori constraints infused into the inversion kernel. Indeed, it does not require any statistical test to choose the adequate model parameterization and do not include any regularization operator to force the model to honour some external constraints. Instead, it automatically adjusts the underlying model parametrization to produce solutions with appropriate

819 level of complexity to fit the data to statistically meaningful levels. The downside of rjMCMC is
820 that a specific recipe for the problem at hand is usually needed to speed up to convergence of the
821 sampling toward the stationary regime. To mitigate this issue, in this work we employed a parallel
822 tempering strategy and a delayed rejection perturbation scheme: the former ensures optimal
823 exploitation and exploration capabilities, while the latter locally adapts the proposal to the shape of
824 the posterior. In particular, the convergence speed of the algorithm critically depends on the choice
825 of the statistical characteristics of such proposal distribution: a suboptimal proposal distribution
826 causes a persistent rejection of models, thus significantly reducing the exploitation of the algorithm.
827 Note that the updating of the different chains in the rjMCMC inversion can be done in parallel with
828 the different chains distributed across different processors. Such parallelizability and the 20
829 available CPUs make it possible the application of the parallel tempering and the delayed rejection
830 strategies with a limited extra computational cost. Indeed, in the parallel tempering the models
831 sampled at $T>1$ must be disaggregated when computing the PPD and this fact implies that many
832 chains (each one involving a forward modelling run per iteration) are needed to converge to a stable
833 posterior. On the other hand, the delayed rejection implies the computation of an additional forward
834 solution in a single iteration when the first proposed model is rejected. For this reason, the
835 application of these two strategies must be carefully evaluated. In our inversion tests we found that
836 they are effective in speeding up the converge of the chains toward the stationary regime, in
837 promoting the exploration of the parameter space, and in increasing the acceptance rate of the
838 rjMCMC. Thanks to the limited number of unknown parameters (i.e. limited number of layers in the
839 inverted models) in our tests we found that an uncorrelated Gaussian proposal constitutes a good
840 compromise between acceptance rate and convergence of the sampling. If needed more
841 sophisticated strategies based on SVD decomposition and orthogonal projections could be used to
842 include a covariance model into the model perturbation scheme (Dosso et al. 2014). Another limit
843 of the rjMCMC is that the successively sampled models are usually highly correlated. For this
844 reason, to avoid biased PPD estimations (MacKay, 2003) not all the models sampled after the burn-

in phase have been used to numerically compute the PPD, but several iterations of the algorithm are allowed to elapse in between successive samples. In all the previous examples only 1 model every 5 is considered in the computation of the posterior model.

The main advantage of HMC is that it guarantees an efficient sampling with a rapid convergence toward the stationary regime. This ability rests on the exploitation of the derivative information that other Monte Carlo methods, such the well-known random walk Metropolis (that is the basis of the rjMCMC algorithm), do not consider. In particular, the HMC provides highly independent samples and for this reason a lower number of iterations are needed for a reliable estimation of the posterior model with respect to MCMC algorithms. Indeed, differently from the rjMCMC inversion, the high independence of successively sampled models granted by the HMC framework, allows us to exploit all the models collected after the burn-in phase in computing the PPD. In this work, the HMC algorithm has been implemented for a Gaussian prior model, but another outstanding benefit of HMC is the possibility to consider either parametric or non-parametric priors. However, making the HMC sampling still efficient in highly multimodal distributions is an active research field nowadays (Girolami and Calderhead, 2011; Nishimura and Dunson, 2016; Graham and Storkey, 2017). Differently from rjMCMC, one limit of HMC is that the model parameterization is an input to the inversion. In this study we propose to solve this issue by performing different inversions with different parameterizations and then using the sampled models and standard statistical tools to determine the most probable number of layers. Another limit is that the HMC algorithm can be applied to problems where the derivative of the posterior model can be rapidly computed and where this derivative is continuous. For example, such derivative becomes discontinuous if multiple dispersion modes are considered. For this reason, in this work we have limited the attention the fundamental mode only. Currently, we are investigating this issue in order to possibly extend the HMC to the inversion of multiple modes. For the sake of consistency, in this work also the rjMCMC inversion has been limited to the fundamental model. However, the rjMCMC algorithm does not exploit any derivative information and for this reason the current implementation can be

871 directly used for inverting higher modes. Another critical element of HMC is the choice of a
872 suitable mass matrix. In the examples discussed here we found that setting this matrix equal to the
873 inverse of the locally approximated posterior covariance guaranteed stable posterior estimations.
874 Other choices of the mass matrix, for examples a diagonal matrix (e.g. a scalar multiple of the
875 identity matrix) resulted in very slow convergence and unreliable posterior assessments (see
876 Fichtner et al. 2019). Another critical aspect of HMC is the choice of the length of the trajectory in
877 the phase space. We overcome this issue by drawing the *L* parameter (which controls the number of
878 time integration steps and the trajectory length) from a previously defined uniform distribution in
879 each iteration. However, the trajectory length can be adaptively set, for example by adopting the so
880 called no U-turns sampling method (Gelman et al. 2013; Hoffman and Gelman et al. 2014) that
881 terminates the integration when the trajectory begins to return towards its starting point. In any case,
882 in the examples discussed here we found that 3–6 time steps per trajectory are usually optimal.
883 Using fewer time steps, leads to very high acceptance rates but at the expense of a limited model
884 space exploration and slow convergence. Otherwise, a larger number of time steps resulted in low
885 acceptance rates and in a decreased accuracy of the numerical integration.
886 To get the benefits of both the rjMCMC and HMC algorithms, we are also investigating the
887 possibility to hybridize these two methods and to implement a transdimensional Hamiltonian
888 algorithm for surface wave inversions, that is an inversion approach that infers the most appropriate
889 model parameterizations from the data and that is also characterized by an extremely efficient
890 sampling of the parameter space. An example of this approach applied to reflection seismic data can
891 be found in Sen and Biswas (2017).
892 We ran all our experiments on two deca-core intel E5-2630 @2.2 GHz (128 Gb RAM) and
893 employing parallel Matlab inversion codes. Table 4 lists some more quantitative details about the
894 HMC and rjMCMC seismic inversions on the second subsurface model (the *Vs* model with the
895 velocity inversion). The total computational time was 11 and 19 minutes for rjMCMC and HMC,
896 respectively. We point out that in our current implementations the rjMCMC algorithm is more

scalable than HMC, although the HMC requires a lower number of sampled models to provide a stable posterior. In particular, note that the total number of forward modelling runs in the HMC is higher than that associated to the rjMCMC. Indeed, for a single chain and for a single iteration, the rjMCMC algorithm requires just one forward modelling run, while in the HMC several forward evaluations are needed to compute the Jacobian matrix associated to each considered model. This number of forward evaluations depends on the number of unknown parameters, on the length of the trajectory in the phase space (value of the L parameter), and on the adopted FD scheme. For a Q-dimensional model space, the number of forward evaluations per iteration is given by $(Q\times2+1)\times L$ if a central FD scheme is adopted. This number reduces to $(Q+1)\times L$ if a forward FD is employed. For this reason, a computationally efficient derivation of the Jacobian matrix would be crucial to drastically reduce the computational time of each iteration. Although we employ a computationally intensive numerical approach to derive the Jacobian, we reduce the computational effort of the HMC by computing each column of this matrix in parallel. Indeed, the columns of the Jacobian are independent to each other and their computation can be distributed across different processors. This fact, together with the limited computational cost of the forward solver, did not hamper the application of this method to the considered examples. Possible strategies to drastically reduce the number of forward evaluations could be the adoption of a less computationally expensive FD scheme (e.g. forward finite difference) and/or the adoption of a more parsimonious parameterization of the model space. From the one hand, additional experiments we carried out (not shown here for brevity) demonstrated that the application of the forward difference approach does not sensibly affect the sampling efficiency of the method. On the other hand, since the number of columns on the Jacobian is equal to the number of unknowns, we can reduce the number of forward modelling runs by reducing the number of model parameters, for example by inverting for the *Vs* and layer thickness and not for the *Vp*/*Vs* ratio that is known to be not constrained by the data. To put this in perspective, if we consider a forward FD scheme and we remove the *Vp*/*Vs* ratio from the model vector, the number of forward evaluations on the seismic inversion on the second subsurface model

923 is reduced from 229500 to 81000 (see Table 4). Note that this value is lower than the number of
924 forward evaluations ran by the rjMCMC inversion on the same example (103200). Machine
925 learning approaches could also be useful to reduce the computational cost of MCMC inversion
926 (Hansen and Cordua 2017). In our specific case, machine learning algorithm can be trained to predict
927 a local approximation of the Jacobian matrix around a considered model.
928

	rjMCMC	**HMC**
Average computational time for a single forward modelling evaluation	≈ 0.063 s	
Number of iterations (per chain)	3000	3000
Number of chains	20	1
Total number of forward modelling evaluations	103200 (this number takes into account also the forward modellings related to the delayed rejection)	229500 (this number takes into account the forward modelling runs needed to compute the Jacobian with the central finite difference approach)
Total number of sampled models	60000	3000
Number of models used to compute the posterior	5600	2970
Average computational time of a single iteration	≈ 0.21 s (considering all the 20 chains	≈ 0.38 s (This number varies for

	running in parallel)	different iterations because is dependent on the drawn *L* value)
Average computational time for the Jacobian computation in a single iteration	*Not applicable*	≈ 0.32 s (This number varies for different iterations because is dependent on the drawn *L* value)

Table 4: Some details of the rjMCMC and HMC inversions for the last seismic inversion.

CONCLUSIONS

We implemented a Hamiltonian Monte Carlo (HMC) algorithm and a reversible jump Marko chain Monte Carlo (rjMCMC) algorithm for Rayleigh wave dispersion curve inversion. This inverse problem is highly non-linear and highly ill-conditioned. For this reason, the ultimate goal of this study was to compare two inversion approaches that guarantee reliable uncertainty quantifications, that estimate the most probable model parameterization (i.e. number of layers in the subsurface), and that yield stable predictions not affected by the choice of the starting model. We limited the attention to synthetic tests to maintain the discussion at a didactic level and to draw essential conclusions about the main benefits and drawbacks of the two implemented approaches in the context of dispersion curve inversion. Our experiments demonstrated that HMC and rjMCMC are very promising approaches for dispersion curve inversion as they provide reliable assessment of the posterior uncertainties also for this highly non-linear and severely ill-conditioned inverse problem. In particular, the HMC and rjMCMC algorithms yielded uncertainty quantifications and model predictions in accordance with the expected model parameter illuminations and model parameter correlations. For example, the uncertainty increases passing from *Vs* to *Vp*/*Vs* ratio and increases as

946 the depth of investigation increases or when the noise in the observed data increases.
947 Notwithstanding the two algorithms use different model parameterizations and prior assumptions,
948 they estimate congruent and comparable posterior uncertainties and model parameter correlations
949 especially for the parameters better illuminated by the data, that is for those parameters for which
950 the PPD is mainly influenced by the likelihood function instead of by the a-priori constraints. The
951 transdimensional inversion approach and the combined used of HMC and statistical tools (such as
952 BIC and $p(\chi^2)$) demonstrated to be two valid strategies to determine the model parameterization (i.e.
953 number of layers), at least for subsurface models with a limited number of layers (less than 4 if only
954 the fundamental model is considered).
955 For both approaches an efficient parallel code is essential to decrease the computational time of the
956 inversion procedure related to the many forward modelling runs needed by the HMC to numerically
957 compute the Jacobian matrix, and by the rjMCMC to make evolve the different chains in a single
958 iteration.
959
960 **ACKNOWLEDGMENTS**
961 The authors wish to thank the associate editor and the anonymous reviewers for their constructive
962 criticism and comments that helped to significantly improve the quality of the original manuscript.
963
964 **Conflict of interest**
965 The authors declare no conflict of interest
966
967 **Data Availability Statement**
968 Data available on request from the authors

References

Aleardi, M. (2015). The importance of the Vp/Vs ratio in determining the error propagation, the stability and the resolution of linear AVA inversion: a theoretical demonstration. Bollettino di Geofisica Teorica ed Applicata, 56(3), 357-366.

Ando, T. (2010). Bayesian model selection and statistical modeling. Chapman and Hall/CRC.

Aster, R. C., Borchers, B., and Thurber, C. H. (2018). Parameter estimation and inverse problems. Elsevier.

Betancourt, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo. arXiv preprint arXiv:1701.02434.

Bishop, C.M. (2006). Pattern Recognition and Machine Learning, Springer.

Blanes, S., Casas, F., and Sanz-Serna, J. M. (2014). Numerical integrators for the Hybrid Monte Carlo method. SIAM Journal on Scientific Computing, 36(4), A1556-A1580.

Bodin, T., and Sambridge, M. (2009). Seismic tomography with the reversible jump algorithm. Geophysical Journal International, 178(3), 1411-1436.

Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K., and Rawlinson, N. (2012). Transdimensional inversion of receiver functions and surface wave dispersion. Journal of Geophysical Research: Solid Earth, 117, B02301.

Boiero, D., Bergamo, P., Bruno Rege, R., and Socco, L. V. (2011). Estimating surface-wave dispersion curves from 3D seismic acquisition schemes: Part 1—1D models. Geophysics, 76(6), G85-G93.

Cercato, M. (2011). Global surface wave inversion with model constraints. Geophysical Prospecting, 59(2), 210-226.

Cho, Y., Gibson Jr, R. L., and Zhu, D. (2018). Quasi 3D transdimensional Markov-chain Monte Carlo for seismic impedance inversion and uncertainty analysis. Interpretation, 6(3), T613-T624

994 Dadi, S., Gibson, R., and Wang, K. (2016). Velocity log upscaling based on reversible jump
995 Markov chain Monte Carlo simulated annealing RJMCMC-based sonic log upscaling. Geophysics,
996 81(5), R293-R305.
997 Dal Moro G., Pipan M. and Gabrielli P. (2007). Rayleigh wave dispersion curve inversion via
998 genetic algorithms and marginal posterior probability density estimation. Journal of Applied
999 Geophysics 61, 39–55.
1000 Dettmer, J., Molnar, S., Steininger, G., Dosso, S. E., and Cassidy, J. F. (2012). Trans-
1001 dimensional inversion of microtremor array dispersion data with hierarchical autoregressive error
1002 models. Geophysical Journal International, 188(2), 719-734.
1003 Di Giulio, G., Punzo, M., Bruno, P. P., Cara, F. and Rovelli, A., (2019). Using a vibratory source
1004 at Mt. Etna (Italy) to investigate the wavefield polarization at Pernicana Fault, Near Surface
1005 Geophysics, 17(4), 313-329.
1006 Dosso, S. E., Dettmer, J., Steininger, G., and Holland, C. W. (2014). Efficient trans-dimensional
1007 Bayesian inversion for geoacoustic profile estimation. Inverse Problems, 30(11), 114018.
1008 Dosso, S. E., Holland, C. W., and Sambridge, M. (2012). Parallel tempering for strongly
1009 nonlinear geoacoustic inversion. The Journal of the Acoustical Society of America, 132(5), 3030-
1010 3040.
1011 Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid Monte Carlo.
1012 Physics Letters B, 195, 216–222.
1013 Farrugia, D., Paolucci, E., D'Amico, S., and Galea, P. (2016). Inversion of surface wave data for
1014 subsurface shear wave velocity profiles characterized by a thick buried low-velocity layer.
1015 Geophysical Journal International, 206(2), 1221-1231.
1016 Feng S., Sugiyama T. and Yamanaka H. (2005). Effectiveness of multimode surface wave
1017 inversion in shallow engineering site investigations. Exploration Geophysics 36, 26–33.

Fernandez Martinez, J. L., Fernandez Muniz, M. Z., and Tompkins, M. J. (2012). On the topography of the cost functional in linear and nonlinear inverse problems. Geophysics, 77(1), W1-W15.

Fichtner, A., Zunino, A., and Gebraad, L. (2019). Hamiltonian Monte Carlo solution of tomographic inverse problems. Geophysical Journal International, 216(2), 1344-1363.

Fichtner, A., and Simutė, S. (2018). Hamiltonian Monte Carlo inversion of seismic sources in complex media. Journal of Geophysical Research: Solid Earth, 123(4), 2984-2999.

Foti, S., Hollender, F., Garofalo, F., Albarello, D., Asten, M., Bard, P. Y., Comina, C., Cornou, C., Cox, B., Di Giulio, G., Forbriger, T., Hayashi, K., Lunedei, E., Martin, A., Mercerat, D. (2018). Guidelines for the good practice of surface wave analysis: A product of the InterPACIFIC project. Bulletin of Earthquake Engineering, 16(6), 2367-2420.

Galetti, E., and Curtis, A. (2018). Transdimensional electrical resistivity tomography. Journal of Geophysical Research: Solid Earth, 123(8), 6347-6377.

Galetti, E., Curtis, A., Baptie, B., Jenkins, D., and Nicolson, H. (2016). Transdimensional Love-wave tomography of the British Isles and shear-velocity structure of the East Irish Sea Basin from ambient-noise interferometry. Geophysical Journal International, 208(1), 36-58.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. and Rubin, D.B. (2013). Bayesian Data Analysis, 3rd edition, CRC Press.

Girolami, M. and Calderhead, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. Journal of the Royal Statistical Society: Series B, 73(2):123–214.

Graham, M. M., and Storkey, A. J. (2017). Continuously tempered hamiltonian monte carlo. arXiv preprint arXiv:1704.03338.

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

1042 Groos, L., Schäfer, M., Forbriger, T., and Bohlen, T. (2017). Application of a complete
1043 workflow for 2D elastic full-waveform inversion to recorded shallow-seismic Rayleigh waves.
1044 Geophysics, 82(2), R109-R117.
1045 Fernández-Martínez, J. L. (2015). Model reduction and uncertainty analysis in inverse problems.
1046 The Leading Edge, 34(9), 1006-1016.
1047 Hansen, T. M., and Cordua, K. S. (2017). Efficient Monte Carlo sampling of inverse problems
1048 using a neural network-based forward—applied to GPR crosshole traveltime inversion. Geophysical
1049 Journal International, 211(3), 1524-1533.
1050 Haskell N.A. 1953. The dispersion of surface waves on multi-layered media. Bulletin of the
1051 Seismological Society of America 43, 17–34.
1052 Hoffman, M. D., and Gelman, A. (2014). The No-U-Turn sampler: adaptively setting path
1053 lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research, 15(1), 1593-1623.
1054 Kritikakis, G., Vafidis, A., Papakonstantinou, K., and O'Neill, A. (2014). Comparative study of
1055 different inversion techniques applied on Rayleigh surface wave dispersion curves. Near Surface
1056 Geophysics, 12(3), 361-371.
1057 Luke B., Calder ´on C., Stone R.C. and Huynh M. (2003). Nonuniqueness in inversion of seismic
1058 surface-wave data. Proceedings of Symposium on the Application of Geophysics to Engineering
1059 and Environmental Problems (SAGEEP), Denver, pp. 1342–1347. Environmental and Engineering
1060 Geophysical Society.
1061 Luo, Y., Xia, J., Miller, R. D., Xu, Y., Liu, J., and Liu, Q. (2009). Rayleigh-wave mode
1062 separation by high-resolution linear Radon transform. Geophysical Journal International, 179(1),
1063 254-264.
1064 Mackenze, P. B. (1989). An improved hybrid Monte Carlo method. Physics Letters B, 226(3-4),
1065 369-371.
1066 MacKay, D.J. (2003). Information Theory, Inference and Learning Algorithms, Cambridge
1067 University Press.

Malinverno, A. (2000). A Bayesian criterion for simplicity in inverse problem parametrization. Geophysical Journal International, 140(2), 267-285.

Malinverno, A., and Briggs, V. A. (2004). Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes. Geophysics, 69(4), 1005-1016.

Mallick, S., and Frazer, L.N. (1987). Practical aspects of reflectivity modeling. Geophysics, **52**, 1355–1364.

Mandolesi, E., Ogaya, X., Campanyà, J., and Agostinetti, N. P. (2018). A reversible-jump Markov chain Monte Carlo algorithm for 1D inversion of magnetotelluric data. Computers & Geosciences, 113, 94-105.

Maraschini, M., and Foti, S. (2010). A Monte Carlo multimodal inversion of surface waves. Geophysical Journal International, 182(3), 1557-1566.

Menke, W. (2018). Geophysical data analysis: Discrete inverse theory. Academic press.

Mosegaard, K., and Tarantola, A. (1995). Monte Carlo sampling of solutions to inverse problems. Journal of Geophysical Research: Solid Earth, 100(B7), 12431-12447.

Mosegaard, K., and Sambridge, M. (2002). Monte Carlo analysis of inverse problems. Inverse problems, 18(3), R29.

Muir, J.B. and Tkalcic, H. (2015). Probabilistic joint inversion of lowermost mantle P-wave velocities and core mantle boundary topography using differential travel times and hierarchical Hamiltonian Monte-Carlo sampling, in AGU 2015 Fall meeting, San Francisco.

Neal, R.M. (1996). Bayesian Learning for Neural Networks. Springer.

Neal, R.M. (2011). MCMC using Hamiltonian dynamics, in Handbook of Markov Chain Monte Carlo, Brooks, S., Gelman, A., Jones, G. & Meng, X.-Li., Chapman and Hall, 113–162.

Nishimura, A., and Dunson, D. (2016). Geometrically tempered hamiltonian monte carlo. arXiv preprint arXiv:1604.00872.

Picozzi, M., and Albarello, D. (2007). Combining genetic and linearized algorithms for a two-step joint inversion of Rayleigh wave dispersion and H/V spectral ratio curves. Geophysical Journal International, 169(1), 189-200.

Qiu, X., Wang, Y., and Wang, C. (2019). Rayleigh-wave dispersion analysis using complex-vector seismic data. Near Surface Geophysics.

Ray, A. K., and Chopra, S. (2016). Building more robust low-frequency models for seismic impedance inversion. First Break, 34(5), 47-52.

Sambridge, M., and Mosegaard, K. (2002). Monte Carlo methods in geophysical inverse problems. Reviews of Geophysics, 40(3), 3-1.

Sambridge, M. (2014). A parallel tempering algorithm for probabilistic sampling and multimodal optimization. Geophysical Journal International, 196(1), 357-374.

Sambridge, M., Gallagher, K., Jackson, A., and Rickwood, P. (2006). Trans-dimensional inverse problems, model comparison and the evidence. Geophysical Journal International, 167(2), 528-542.

Sajeva, A., Aleardi, M., Mazzotti, A., Stucchi, E., and Galuzzi, B. (2014). Comparison of stochastic optimization methods on two analytic objective functions and on a 1D elastic FWI. In 76th EAGE Conference and Exhibition 2014. Doi: 10.3997/2214-4609.20140857.

Sajeva, A., and Menanno, G. (2017). Characterisation and extraction of a Rayleigh-wave mode in vertically heterogeneous media using quaternion SVD. Signal Processing, 136, 43-58.

Sajeva, A., Aleardi M., Galuzzi, B., Stucchi, E., Spadavecchia, E., and Mazzotti, A. (2017). Comparing the performances of four stochastic optimisation methods using analytic objective functions, 1D elastic full-waveform inversion, and residual static computation. Geophysical Prospecting, **65**, 322-346.

Scalzo, R., Kohn, D., Olierook, H., Houseman, G., Chandra, R., Girolami, M., and Cripps, S. (2019). Efficiency and robustness in Monte Carlo sampling for 3-D geophysical inversions with Obsidian v0. 1.2: setting up for success. Geoscientific Model Development, 12(7), 2941-2960.

Schwartz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6, 461–464.

Sen, M. K., and Biswas, R. (2017). Transdimensional seismic inversion using the reversible jump Hamiltonian Monte Carlo algorithm. Geophysics, 82(3), R119-R134.

Sen, M. K., and Stoffa P.L, (2013). Global optimization methods in geophysical inversion. Cambridge University Press.

Sen, M. K., and Roy, I. G. (2003). Computation of differential seismograms and iteration adaptive regularization in prestack waveform inversion. Geophysics, 68(6), 2026-2039.

Sen, M. K., and Stoffa, P. L. (1996). Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion. Geophysical Prospecting, 44(2), 313-350.

Sisson, S. A. (2005). Transdimensional Markov chains: A decade of progress and future perspectives. Journal of the American Statistical Association, 100(471), 1077-1089.

Socco, L. V., and Boiero, D. (2008). Improved Monte Carlo inversion of surface wave data. Geophysical Prospecting, 56(3), 357-371.

Socco L.V., and Strobbia C. (2004). Surface-wave method for nearsurface characterization: a tutorial. Near Surface Geophysics 2, 165–185.

Steininger, G., Dettmer, J., Dosso, S.E. and Holland, C.W., (2013). Transdimensional joint inversion of seabed scattering and reflection data. Journal acoustic Society of America, 133, 1347–1357.

Tarantola, A. (2005). Inverse problem theory and methods for model parameter estimation. siam.

Theune, U., Jensås, I. Ø., and Eidsvik, J. (2010). Analysis of prior models for a blocky inversion of seismic AVA data. Geophysics, 75(3), C25-C35.

Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. Environmental Modelling & Software, 75, 273-316.

Wathelet, M. (2005). Array recordings of ambient vibrations: Surfacewave inversion, Phd thesis, Universit´e de Li`ege, France.

Xia, J., Miller, R. D., Park, C. B., and Tian, G. (2003). Inversion of high frequency surface waves with fundamental and higher modes. Journal of Applied Geophysics, 52(1), 45-57.

Xiang, E., Guo, R., Dosso, S. E., Liu, J., Dong, H., and Ren, Z. (2018). Efficient hierarchical trans-dimensional Bayesian inversion of magnetotelluric data. Geophysical Journal International, 213(3), 1751-1767.

Xing, Z., and Mazzotti, A. (2018). Two-grid genetic algorithm full waveform inversion of surface waves: two actual data examples. In 80th EAGE Conference and Exhibition 2018.

Zhu, D., and Gibson, R. (2018). Seismic inversion and uncertainty quantification using transdimensional Markov chain Monte Carlo method. Geophysics, 83(4), R321-R334.