Geophysical Journal International

Geophys. J. Int. (2015) **200,** 1134–1152
GJI Seismology

doi: 10.1093/gji/ggu445

Multicomponent, multi-azimuth pre-stack seismic waveform inversion for azimuthally anisotropic media using a parallel and computationally efficient non-dominated sorting genetic algorithm

Tao Li and Subhashis Mallick

Department of Geology and Geophysics and the School of Energy Resources, University of Wyoming, Laramie, WY 82071*, USA. E-mail: smallick@uwyo.edu*

Accepted 2014 November 11. Received 2014 November 10; in original form 2014 May 28

SUMMARY

Consideration of azimuthal anisotropy, at least to an orthorhombic symmetry is important in exploring the naturally fractured and unconventional hydrocarbon reservoirs. Full waveform inversion of multicomponent seismic data can, in principle, provide more robust estimates of subsurface elastic parameters and density than the inversion of single component (*P* wave) seismic data. In addition, azimuthally dependent anisotropy can only be resolved by carefully studying the multicomponent seismic displacement data acquired and processed along different azimuths. Such an analysis needs an inversion algorithm capable of simultaneously optimizing multiple objectives, one for each data component along each azimuth. These multicomponent and multi-azimuthal seismic inversions are non-linear with non-unique solutions; it is therefore appropriate to treat the objectives as a vector and simultaneously optimize each of its components such that the optimal set of solutions could be obtained. The fast non-dominated sorting genetic algorithm (NSGA II) is a robust stochastic global search method capable of handling multiple objectives, but its computational expense increases with increasing number of objectives and the number of model parameters to be inverted for. In addition, an accurate extraction of subsurface azimuthal anisotropy requires multicomponent seismic data acquired at a fine spatial resolution along many source-to-receiver azimuths. Because routine acquisition of such data is prohibitively expensive, they are typically available along two or at most three azimuthal orientations at a spatial resolution where such an inversion could be applied. This paper proposes a novel multi-objective methodology using a parallelized version of NSGA II for waveform inversion of multicomponent seismic displacement data along two azimuths. By scaling the objectives prior to ranking, redefining the crowding distance as functions of the scaled objective and the model spaces, and varying the crossover and mutation parameters over generations, the proposed methodology is also an improvement of the original NSGA II in overall computational efficiency, preservation of population diversity, and rapid sampling of the model space. By first inverting the near-offset pre-stack data for the background isotropic properties and obtaining constraints on the vertical velocities, followed by an inversion of the long-offset data, it is demonstrated that the proposed method can reliably estimate density and azimuthally anisotropic subsurface properties up to the complexity of an orthorhombic symmetry on noisy synthetic data computed from a model based on a real well log under an assumption of 1-D subsurface layers where the ambiguities between lateral heterogeneity and anisotropy could be ignored. In addition, a practical way to approximately compute the uncertainty values in the derived parameters using the method is also demonstrated.

Key words: Inverse theory; Seismic anisotropy; Computational seismology.

INTRODUCTION

Pre-stack waveform inversion (PWI) of seismic reflection data has been shown to be an effective tool for subsurface characterization of hydrocarbon reservoirs (Mallick 1999; Sen & Roy 2003). Because mode-converted (P–S_V) reflections are more sensitive to the variations in density compared to primary (P–P) reflections (Aki & Richards 2002), combining P–P with P–S_V reflections in a multicomponent seismic inversion has been shown to estimate density better than the single component inversion (Mallick 2000;

Padhi & Mallick 2013a,b). Under an isotropic assumption, such a multicomponent inversion requires estimation of three parameters, namely the *P*- and *S*-wave velocities and density. When additional anisotropic parameters are also required to be extracted from multicomponent seismic data, the inverse problem becomes complex. PWI using two- (*P–P*, *P–S*) and four component (*P–P*, *P–S*, *S–P*, *S–S*) surface seismic data for estimating subsurface elastic parameters and density under both isotropic and transversely isotropic with a vertical symmetry axis (VTI) assumptions has been demonstrated by Padhi & Mallick (2013a,b). With increasing interests in the naturally fractured and unconventional hydrocarbon reservoirs, it is however necessary to go past isotropy and VTI, and consider azimuthal anisotropy (Chaveste *et al.* 2013; Gaiser *et al.* 2013). Most unconventional reservoirs are inherently VTI and on top of this, *in situ* stress fields in conjunction with preferentially oriented natural fractures make them azimuthally anisotropic with orthorhombic (ORT) symmetry (Schoenberg & Douma 1988). Most naturally fractured reservoirs are also azimuthally anisotropic with transversely isotropic with a horizontal/tilted symmetry axis (HTI/TTI) or even ORT symmetry properties. Consequently, inverting surface seismic data at least for ORT subsurface properties is important for accurately characterizing these reservoirs.

Propagation of elastic waves in an ORT medium can be described by using nine independent elastic constants, density and the azimuthal direction of the vertical symmetry plane. These nine elastic constants could be given as the elastic (stiffness) coefficients namely C_{11}, C_{12}, C_{13}, C_{22}, C_{23}, C_{33}, C_{44}, C_{55} and C_{66} describing the constitutive law which linearly relates six components of the stress tensor with six components of the strain (deformation) tensor (Auld 1973). Alternatively, they could also be given as the Thomsen–Tsvankin parameters (Thomsen 1986; Tsvankin 1997) as the vertical *P*- and *S*-wave velocities (V_P, V_S), ε_1, ε_2, γ_1, γ_2, δ_1, δ_2 and δ_3 (see Appendix A for details). Inverting surface seismic data for ORT media properties thus requires extraction of 11 independent parameters—nine elastic constants, density and the direction of the symmetry plane. Using a multicomponent waveform inversion, Chang & McMechan (2009) demonstrated that such a method is capable of extracting subsurface anisotropic parameters and density for a simple three-layer case where the middle layer was anisotropic. Padhi & Mallick (2013a) have also successfully inverted multicomponent seismic waveform data for elastic model parameters and density under both isotropic and VTI assumptions for simple models. Padhi & Mallick (2013b) extended the VTI inversion problem to include complex models. To extend VTI inversion further to azimuthal anisotropy, two major challenges need to be overcome:

(1) Multicomponent seismic data requires simultaneous optimizations of multiple objectives, one for each data component. These optimizations are non-linear and non-unique with multiple optimal solutions known as the Pareto-optimal solutions defined as the set of solutions, none of which could be considered better than the other in terms of satisfying all the objectives (Deb *et al.* 2002; Padhi & Mallick 2013a,b). A traditional inversion that defines a single objective as a weighted sum of different objectives and then optimizes it provides only one out of the entire Pareto-optimal set and is generally biased by the choice of the individual weights. For multi-objective inverse problems, it is desirable to treat the entire set of objectives as a vector, simultaneously optimize all its components to estimate the Pareto-optimal set, and then interpret it to obtain the solution that is geologically meaningful. Padhi & Mallick (2013a,b) demonstrated that non-dominated sorting genetic algorithm (NSGA II) is able to estimate such a Pareto-optimal set and achieve a robust estimation of the subsurface properties up to a complexity of VTI anisotropy. For inversion to a complexity of ORT anisotropy is however a challenge because the computational expense of NSGA II increases exponentially with increasing number of objectives and the number of model parameters to be estimated for; requiring the need for the development of a computationally efficient NSGA II.

(2) If multicomponent and multi-azimuthal seismic data are recorded with a very fine grid spacing so that the subsurface reflection records from all possible azimuthal orientations and incidence (polar) angles between the source and receivers are available, they can be used in an inversion scheme and the exact anisotropic properties could be extracted (Dewangan & Grechka 2003). Acquisition of such multicomponent 3-D seismic data in such a fine grid spacing is however expensive, and in practice they are acquired such that data are predominantly oriented along two or at most three source-to-receiver azimuths. Therefore, inversion of multi-azimuth surface seismic data is restricted by the acquisition conditions, in reality applying inversion using the approach of Dewangan & Grechka (2003) is severely limited, and developing an inversion methodology keeping the practical limitations of surface seismic data acquisitions in mind is essential.

Considering the practical aspects outlined as above, here we formulated the inversion of multicomponent pre-stack seismic waveform data recorded along two azimuthal directions as a multi-objective optimization problem. We developed a parallel and computationally efficient version of NSGA II. We then generated multicomponent synthetic data (vertical and horizontal inline and crossline components, that is, the components parallel and perpendicular to the source–receiver directions) of particle displacement from an explosive point source using a multilayer model with isotropic, VTI, HTI and ORT properties based on a real well log from the Rock Springs Uplift, WY, USA. Finally, we jointly inverted these multicomponent data using our methodology.

METHODOLOGY

Forward modelling

Restricting to 1-D earth model, the forward computation of synthetic seismograms could be efficiently performed using the reflectivity method (Fuchs & Müller 1971; Kennett 1983; Fryer & Frazer 1984; Mallick & Frazer 1987, 1988, 1990, 1991; Sen & Roy 2003). Although the method is well documented in all these references, it is necessary to briefly outline its details in connection with the development of a computationally efficient NSGA II for anisotropic inversion.

If we assume that the elastic properties vary as a function of depth (z) only, the dependence of the displacement vector $\boldsymbol{u} = [u_x, u_y, u_z]^T$ and the components of the stress tensor $\boldsymbol{\tau}_{ij}(i, j = \{x, y, z\})$ that represent normal traction on the vertical (x–z and y–z) planes can be transformed out using the Fourier transforms to get a resulting system of equations in frequency-slowness domain in the form (Woodhouse 1974; Fryer & Frazer 1984)

$$\partial_z \boldsymbol{b} = \mathrm{i}\omega \boldsymbol{A} \cdot \boldsymbol{b} + \boldsymbol{F}, \tag{1}$$

where $\partial_z = \frac{\partial}{\partial z}$ is the derivative with respect to the vertical or z coordinate axis, and $\boldsymbol{b} = [u_x, u_y, u_z, \frac{1}{\mathrm{i}\omega}\tau_{zx}, \frac{1}{\mathrm{i}\omega}\tau_{zy}, \frac{1}{\mathrm{i}\omega}\tau_{zz}]^T$ is the displacement–stress vector with the elements being the scalar components of the particle displacement and the vertical traction. The

matrix $\boldsymbol{A}$ in eq. (1), known as the elastic system matrix, is a function of the elastic stiffness coefficients, the x and y components of horizontal slowness (p_x and p_y), angular frequency ω and density ρ. Finally, $\boldsymbol{F}$ in eq. (1) is the body force (source) term. The system of equations given by (1) could be solved using the reflection matrix approach (Kennett 1983), extended for anisotropic medium by Fryer & Frazer (1984) and Mallick & Frazer (1990, 1991). In this approach, seismic waves are decomposed into multiple plane waves, each of which is represented by constant angular frequency (ω) and horizontal slowness (p_x and p_y). First, the reflection and transmission coefficient matrices for the entire stack of layers are computed, which is called the reflection matrix and is denoted as

$$\boldsymbol{\mathcal{R}} = \begin{bmatrix} \boldsymbol{T}_u & \boldsymbol{R}_d \\ \boldsymbol{R}_u & \boldsymbol{T}_d \end{bmatrix}_{6\times 6}, \tag{2}$$

where $\boldsymbol{T}_u$ and $\boldsymbol{T}_d$ denote the 3×3 transmission coefficient matrices for upgoing and downgoing waves, and $\boldsymbol{R}_u$ and $\boldsymbol{R}_d$ denote the 3×3 reflection coefficient matrices for upgoing and downgoing waves. By setting the boundary condition for the reflection matrix at the top as:

$$\boldsymbol{\mathcal{R}} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{R}_u^f & \boldsymbol{I} \end{bmatrix}_{6\times 6}, \tag{3}$$

where $\boldsymbol{I}$ is the 3×3 identity matrix and $\boldsymbol{R}_u^f$ is the 3×3 reflection coefficient matrix at the free surface, and then calculating the reflection and transmission coefficient matrices of each layer interface, the entire reflection matrix of the stratified medium can be recursively computed from the shallowest to the deepest layer by the following set of equations:

$$\begin{cases} (\boldsymbol{T}_u)_{\text{new}} = \boldsymbol{T}_u(\boldsymbol{I} - \boldsymbol{r}_d\boldsymbol{R}_u)^{-1}\boldsymbol{t}_u, \\ (\boldsymbol{R}_d)_{\text{new}} = \boldsymbol{R}_d + \boldsymbol{T}_u\boldsymbol{r}_d(\boldsymbol{I} - \boldsymbol{R}_u\boldsymbol{r}_d)^{-1}\boldsymbol{T}_d, \\ (\boldsymbol{R}_u)_{\text{new}} = \boldsymbol{r}_u + \boldsymbol{t}_d\boldsymbol{R}_u(\boldsymbol{I} - \boldsymbol{r}_d\boldsymbol{R}_u)^{-1}\boldsymbol{t}_u, \\ (\boldsymbol{T}_d)_{\text{new}} = \boldsymbol{t}_d(\boldsymbol{I} - \boldsymbol{R}_u\boldsymbol{r}_d)^{-1}\boldsymbol{T}_d, \end{cases} \tag{4}$$

where $\boldsymbol{t}_u$ and $\boldsymbol{t}_d$ are the 3×3 transmission coefficient matrices for upgoing and downgoing waves, and $\boldsymbol{r}_u$ and $\boldsymbol{r}_d$ are the 3×3 reflection coefficient matrices for upgoing and downgoing waves at a given layer interface in the subsurface up to which the reflection matrix is calculated. Once the reflection matrices are obtained, the x, y and z components of the displacement response (u_x, u_y and u_z) could be computed by multiplying the reflection and transmission matrices with the source and the receiver terms (for details, see Kennett 1983; Fryer & Frazer 1984; Mallick & Frazer 1990). From the equation set (4), we can note down the following important facts:

(1) They provide a complete response, including all mode conversions and interbed reverberations.

(2) Because the reflection matrix $\boldsymbol{\mathcal{R}}$ is built from the shallowest to the deepest layer, it could be used to our advantage. For most inversion problems, the shallow part is the overburden and is given as prior information. Therefore, the reflection matrix for the overburden needs to be computed only once to which the reflection matrix computed over the inversion window could just be added using the equation set (4). As will be discussed later, a major improvement in the computational efficiency of our proposed method could be achieved by inverting the entire time window with several short and overlapping time windows, starting from the shallow to the deep part of the data. In such sliding window inversion, once the upper part is inverted, their reflection matrix response could just be computed only once and be added to the lower part that is being inverted for, leading to a fast and computationally efficient algorithm.

(3) The equation set (4) is for each plane wave and they are independent of one another. Consequently, they could all be simultaneously computed by different compute nodes in a parallel computing environment which can lead to a very fast computation of the synthetic responses.

If the properties of the layers are physically meaningful, which, in mathematical terms, means that the elastic coefficient matrix $[C_{ij}]$ is positive definite, then the elastic system matrix $\boldsymbol{A}$ in eq. (1) will have three pairs of eigenvalues denoting the vertical slowness of upgoing and downgoing P, S_1 and S_2 waves, and the reflection matrix $\boldsymbol{\mathcal{R}}$ in eq. (2) is unconditionally stable (Fryer & Frazer 1984). The conditions for the positive definiteness for $[C_{ij}]$ for different anisotropic systems, is discussed in further detail in Appendix A. The 6×6 system of equations given by eq. (1) decouples into a 4×4 P–S_V system and a 2×2 S_H system for the isotropic and VTI case, and the P–S_V system could be considered independently of the S_H (Padhi & Mallick 2013b). For the azimuthally anisotropic case however, the P, S_1 and S_2 waves are always coupled, except along some specific directions (Auld 1973).

Multi-objective inversion

A multi-objective optimization problem involves finding solutions that simultaneously maximize or minimize more than one goal or objective while satisfying some constraints in the decision or model space. Without any loss of generality, all PWI problems could be written in the form

$$\boldsymbol{X}_k = f^{-1}(\boldsymbol{D}_{ij}). \tag{5}$$

In eq. (5), $\boldsymbol{D}_{ij} = [\boldsymbol{d}_{ij}^{(1)}, \boldsymbol{d}_{ij}^{(2)}, \ldots, \boldsymbol{d}_{ij}^{(m)}]$ is the data matrix, denoting the processed pre-stack seismic data of the jth component along the ith azimuth. In this study, $i = \{1, 2\}$ since we apply our inversion along two azimuths; $j = \{1, 2, 3\}$ since the data have three components (vertical, and the horizontal inline and crossline components of the displacement response); and m is the number of seismic data traces. Now consider $\boldsymbol{X}_k = [\boldsymbol{x}_k^{(1)}, \boldsymbol{x}_k^{(2)}, \ldots, \boldsymbol{x}_k^{(n)}]^T$ is the model or the decision vector in the model/decision space. In above, the parameter n is the number of subsurface layers, and $k = \{1, 2, \ldots, K\}$, with K being the number of possible models. For all isotropic layers where the wave propagation is described by two elastic constants and density per layer, each model vector $\boldsymbol{x}_k^{(i)}$ will consist of three components. For all VTI layers $\boldsymbol{x}_k^{(i)}$ will contain six components denoting the five elastic constants and density. For all HTI layers $\boldsymbol{x}_k^{(i)}$ will contain seven components representing the five elastic constants, density and azimuthal direction of vertical symmetry plane. Finally, for all ORT layers $\boldsymbol{x}_k^{(i)}$ will contain eleven components with the nine elastic constants, density and azimuthal direction of vertical symmetry plane. For our inverse problem, it is not possible to know beforehand the exact anisotropic symmetry for each layer. However, considering the fact that we apply our inversion along two azimuths, what we can find out is whether or not the reflection from a given layer is azimuthally independent, based on which we parametrize them either as a VTI or an ORT medium. We note here that it was possible to parametrize the entire subsurface as ORT for inversion. However, parametrizing them either as VTI or ORT based on azimuthally independent/dependent seismic responses improved overall computational efficiency. Additionally, note that we could parametrize the subsurface either as VTI

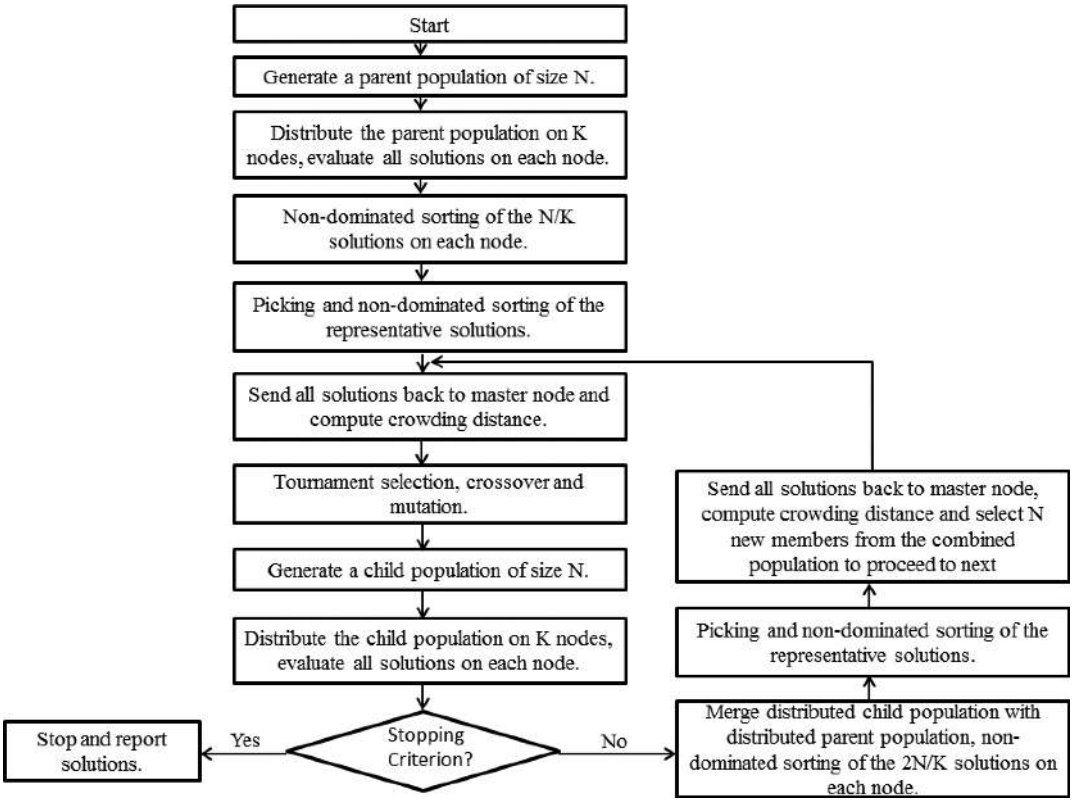

Figure 1. A parallel NSGA II workflow.

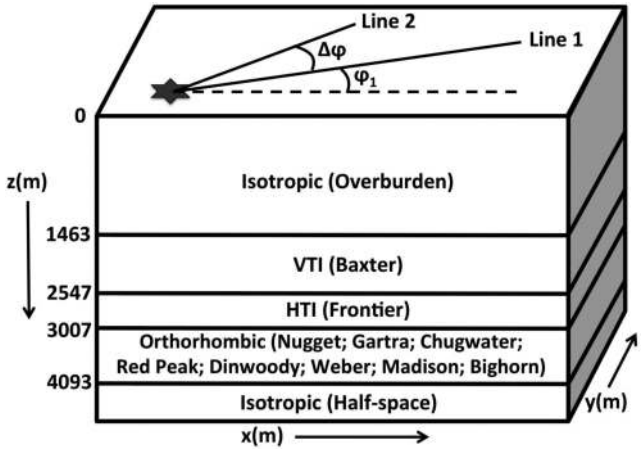

Figure 2. A schematic diagram of the synthetic model used to test the inversion methodology. Note that the surface seismic data along two lines, denoted as Line 1 and Line 2 with known azimuthal orientations are necessary to invert the subsurface for the azimuthally anisotropic elastic properties.

or ORT in our specific application because our model was isotropic and/or VTI in the shallow part and in the deep parts it is azimuthally anisotropic (HTI/ORT). This allowed us to compare the azimuthal seismic responses and parametrize our model as VTI up shallow and ORT down deep. However if there are isotropic/VTI layers deeper than the first azimuthally anisotropic layer in the model, they must be parametrized as ORT. The operator f in eq. (5) is the physical law or the forward computation operator described previously that links observed seismic data with subsurface elastic properties. Any model vector $\boldsymbol{X}$ could be used in this forward computation operator to generate the synthetic data matrix $\boldsymbol{S}_{ij} = [\boldsymbol{s}_{ij}^{(1)}, \boldsymbol{s}_{ij}^{(2)}, \ldots, \boldsymbol{s}_{ij}^{(m)}]$ exactly at the same points as for the data matrix $\boldsymbol{D}_{ij}$. We can then use the cross-correlation function to define a misfit or objective function in the objective space $\boldsymbol{Y}$ given as

$$\boldsymbol{Y}_{ij} = 1 - \frac{\boldsymbol{S}_{ij}.\boldsymbol{D}_{ij}}{\sqrt{\boldsymbol{S}_{ij}.\boldsymbol{S}_{ij}} \cdot \sqrt{\boldsymbol{D}_{ij}.\boldsymbol{D}_{ij}}}. \quad (6)$$

An inversion starts with a set of models in the model or decision space, and is iteratively modified until the objective defined by eq. (6) is minimized.

Parallel NSGA II

NSGA II is an elitist multi-objective evolutionary algorithm (MOEA) with a fast non-dominated sorting and diversity preservation mechanism. Key concepts of MOEA are explained in Appendix B. The purpose of MOEA is to find a set of solutions as close as possible to the true Pareto-optimal front while maintaining diversity in the obtained solutions, and NSGA II has been shown to be capable of achieving both (Deb *et al.* 2002). In any evolutionary algorithm (EA), elitism refers to the condition where both parent and child populations are considered in selecting members for the consecutive generation. This has been shown to improve convergence properties of these algorithms (Stoffa & Sen 1991; Sen & Stoffa 1992; Rudolph 1993; Mallick 1995, 1999).

The main challenge with any stochastic optimization algorithm is its computational complexity. The NSGA II for example has a computational complexity of $O(MN^2)$, where M is the number of objectives and N is the population size (Deb *et al.* 2002). This makes NSGA II computationally expensive for large population sizes. For an azimuthally anisotropic seismic inversion problem, maintaining a large population size is however necessary. One primary reason for this computational complexity is due to the non-dominated sorting procedure of NSGA II in defining ranks of the individual members

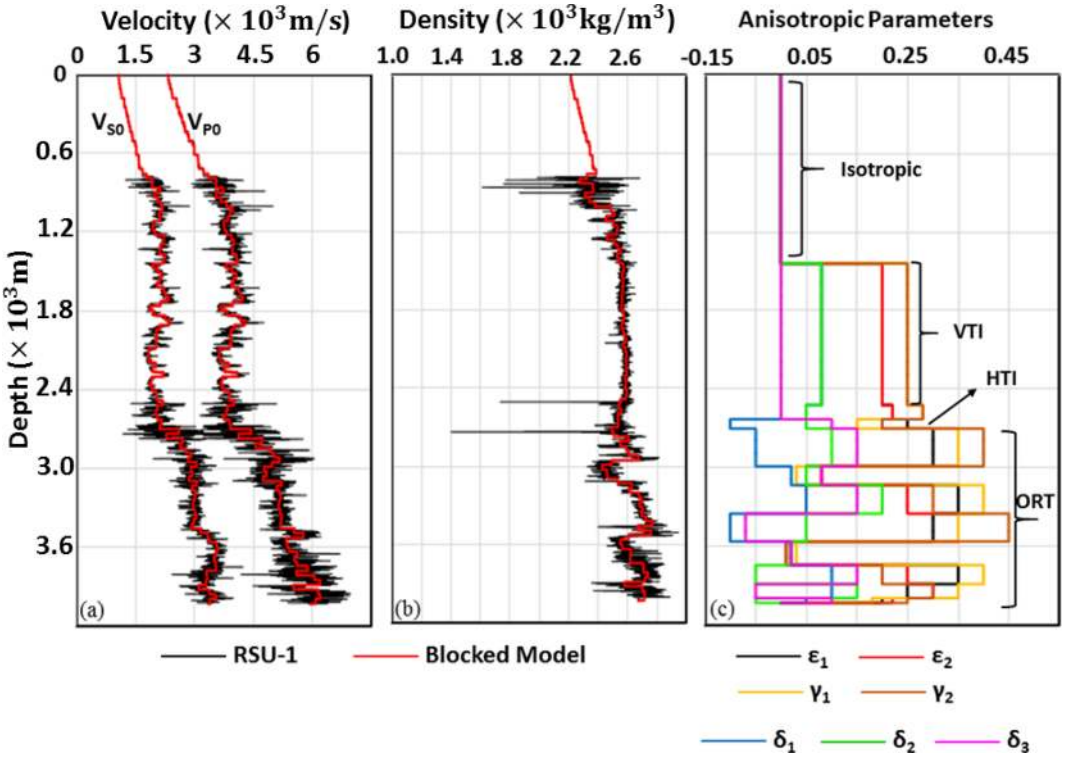

Figure 3. (a), (b) RSU-1 well data and the blocked model that were used to generate the synthetic data for inversion. (c) Thomsen–Tsvankin parameters: ε_1, ε_2, γ_1, γ_2, δ_1, δ_2 and δ_3 used to introduce VTI anisotropy in the Baxter shale region, HTI anisotropy in the Frontier region and ORT anisotropy for all the formations beneath the Frontier formation.

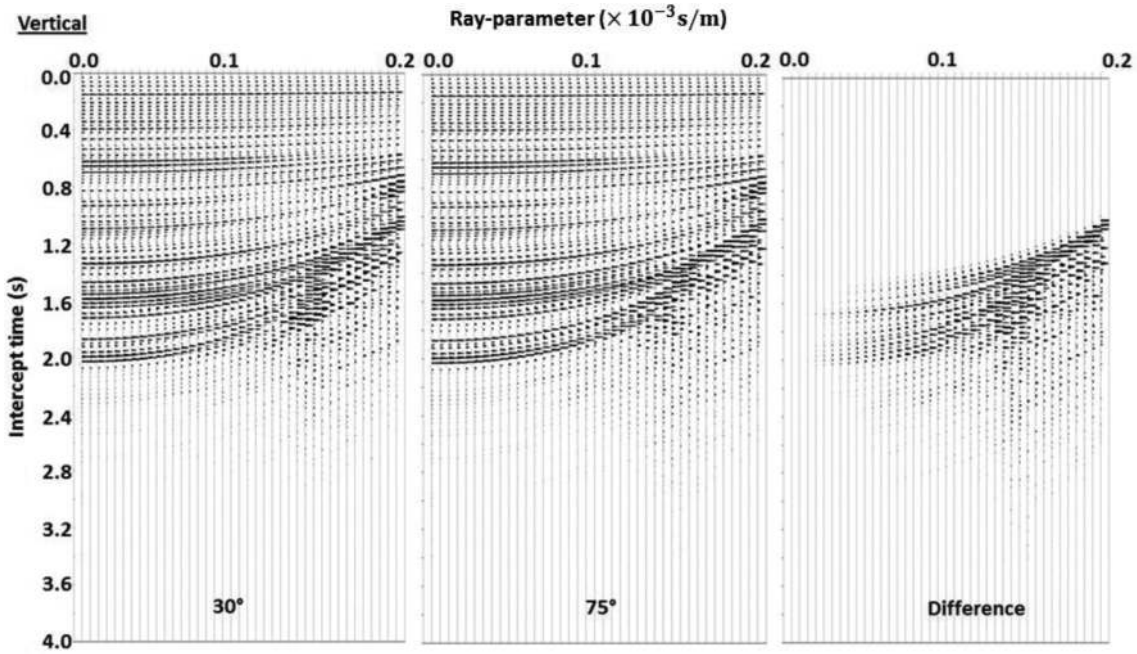

Figure 4. Synthetic displacement responses of the vertical component in $\tau - p$ domain. The angles indicated above are azimuthal angles clockwise from north. Each plot is normalized independently.

(see Appendix B). In this work, the original NSGA II was parallelized for computational efficiency. Padhi & Mallick (2013a,b) have outlined NSGA II in detail and here we only outline its parallel implementation. Fig. 1 is a schematic flow diagram of our parallel NSGA II. First a random parent population of size N is generated honouring the constraints in the decision/model space. This population is then distributed to K nodes, evaluated on each node, and sorted into different ranks based on suitably scaled values of all objective functions. Once the ranks are assigned, one solution is picked from each Pareto-optimal front (each rank) on each node as the representative solution. After non-dominated sorting of all representative solutions, they are sent back to the master node to

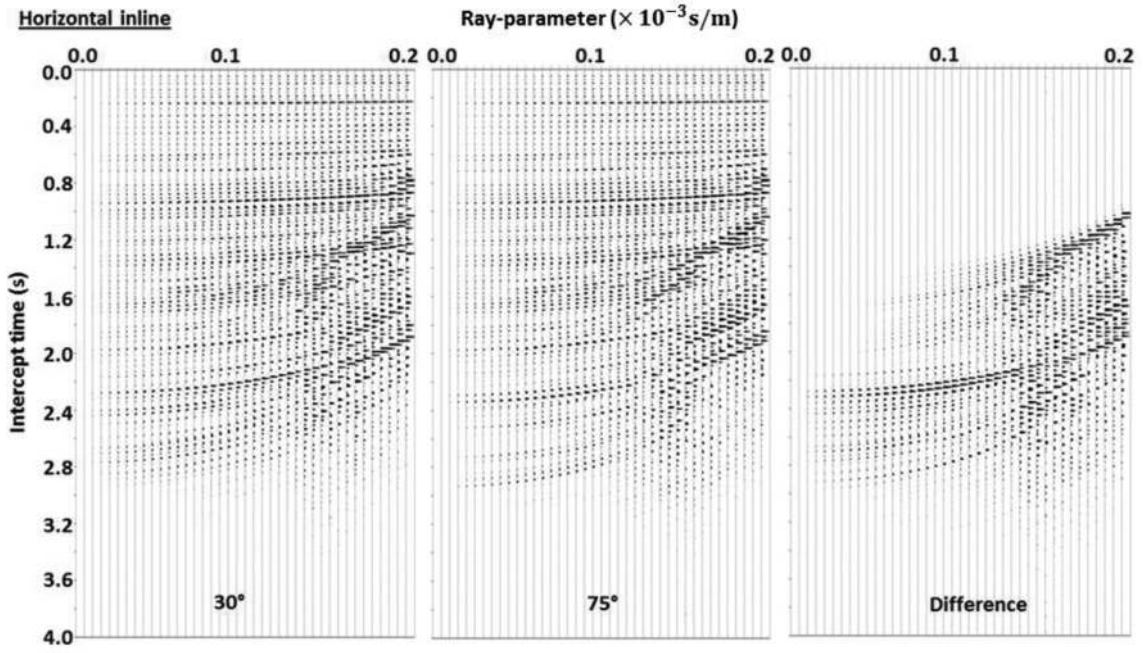

Figure 5. Same as Fig. 4, but for the horizontal inline component.

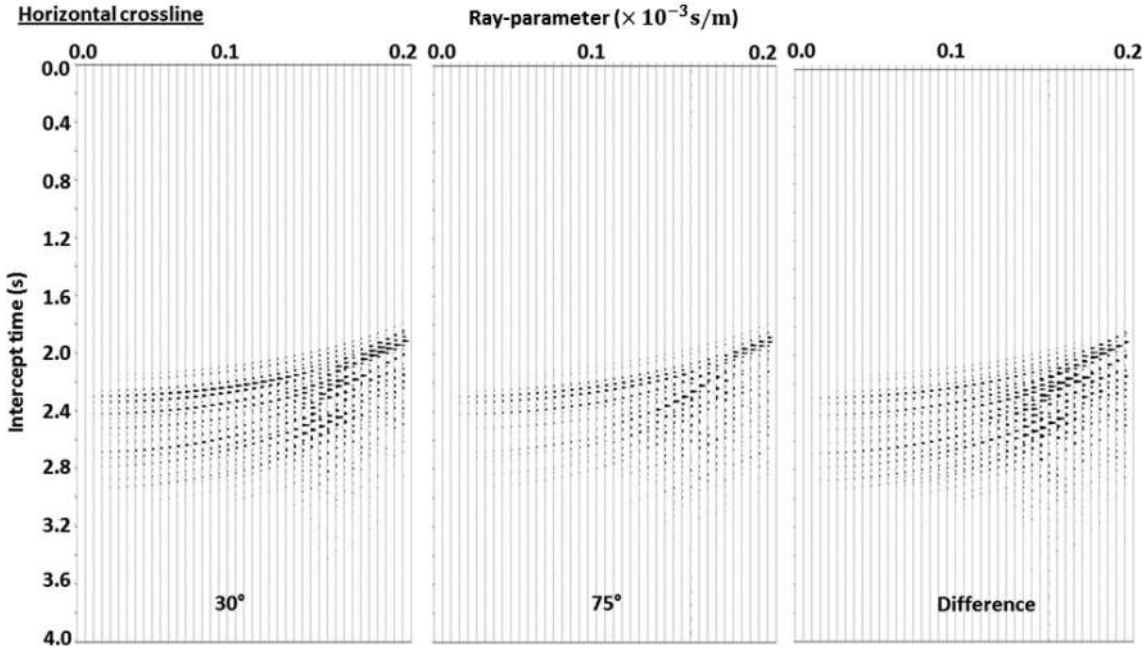

Figure 6. Same as Fig. 4, but for the horizontal crossline component.

construct the parent population. Ranks of each solution in the parent population were assigned according to the non-dominated sorting results of representative solutions. Since the dominances were pre-checked on each node, they are guaranteed in the parent population, and the computational complexity is reduced since it is no longer necessary to check all N solutions. After ranking, the crowding distances for the members belonging to each rank are then calculated. The parent population then undergoes the genetic algorithm (GA) processes of tournament selection (reproduction), crossover and mutation to generate the child population. We must emphasize here that the original NSGA II of Deb *et al.* (2002) proposed the tournament selection based on ranks computed using only the raw fitness values (objectives) and the crowding distances measured in the objective space alone. In our parallel implementation, we however found that better results are obtained if the raw fitness values are first linearly scaled (see Appendix B) before ranking and the members belonging to a given rank are assigned new measure of the crowding distance that are computed by combining the normalized distance measured in this scaled objective space with that measured in the model space (Appendix B). As can be seen in Fig. 1, following rank assignments based on the scaled objectives and computing the new crowding distances, the procedures for tournament selection and advancement of the population over generations via crossover, mutation, and elitism in our parallel NSGA II is exactly the same

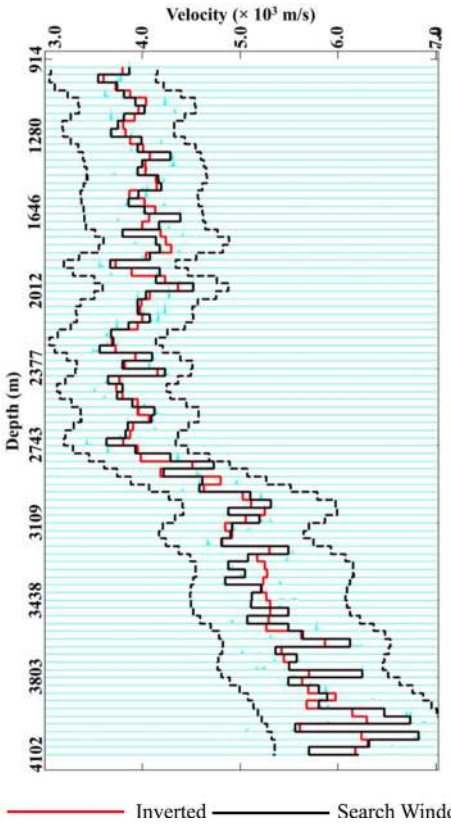

Figure 7. Comparison between the true *P*-wave velocities with the inversion results and the search window used. Here, the near-offset pre-stack data were inverted using an isotropic assumption. Normalized PPD of the inverted results is also plotted in light blue (cyan). The width of each PPD curve helps to quantify the uncertainties associated with the estimates of each model parameter for each layer. Solutions with the highest likelihood were picked as best inversion results from these approximate PPD plots.

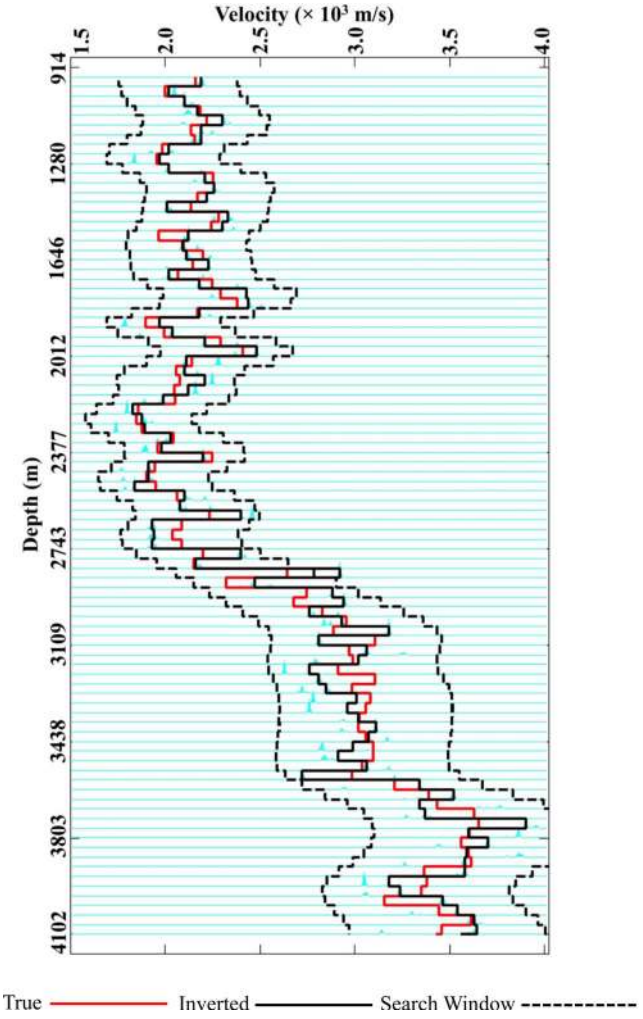

Figure 8. Same as Fig. 7, but for the *S*-wave velocity.

as what was explained in detail by Padhi & Mallick (2013a,b). We would however like to reemphasize here that apart from being a parallel algorithm, our implementation of NSGA II differs from that originally proposed by Deb *et al.* (2002). Below is a list of how our implementation is different along with justifications:

(1) We sort our population at different dominance levels or ranks not based on the raw objectives, but the linearly scaled objectives.

(i) From our previous experience on seismic waveform inversions with single objective problems, we found that if the objectives aren't scaled they have a tendency to quickly converge to a local optimum (Mallick 1995, 1999). In a previous work, Stoffa & Sen (1991) also reported the same issue. Scaling the raw objectives in seismic inverse problems, irrespective of whether they are single or multi-objective is of crucial importance. Our preliminary experiments on anisotropic waveform inversion suggest that linearly scaling each objective prior to ranking and computing the crowding distances help avoid premature convergence.

(2) Crowding distances are computed both along the scaled objective and the model space.

(i) This is also of crucial importance for the anisotropic seismic inversion problems. As will be shown later in our examples, while our total number of objectives are only six (three component seismic data along two azimuths), our total number of model parameters is 689. For problems where the number of model parameters to be solved for is much higher than the number of objectives such as ours, basing crowding distance only in the objective space do not necessarily guarantee that the actual models in the model space are diverse, and we found that combining the normalized distance measured in the scaled objective space with that measured in the model space is a better way to maintain diversity within the generations than using the objective space alone.

Classical implementations of GA use binary coding in which each model parameter is coded as a string of binary digits (bits). For crossover, two members of the population are randomly chosen as parents. A crossover site is then randomly selected within each parameter space and the bit contents between the parents on the right hand side of each selected crossover site for each model parameter are swapped with a given probability of crossover P_c to produce two children (see Stoffa & Sen 1991; Mallick 1995 for details). Following crossover, the bits in each member of the child population are then sequentially visited and modified with a given probability of mutation P_m (Goldberg 1989; Stoffa & Sen 1991; Mallick 1995). Although simple and straightforward, note that the binary coding can only discretely sample the model space. Considering the fact that the model space is continuous, not discrete, it is desirable to

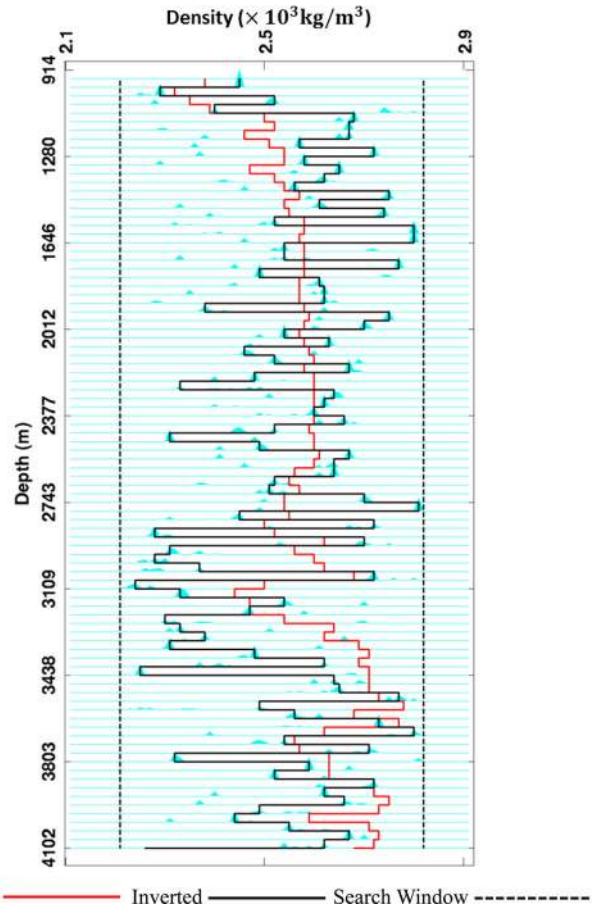

Figure 9. Same as Fig. 7, but for the density. Note that applying an isotropic inversion failed to capture the density trend.

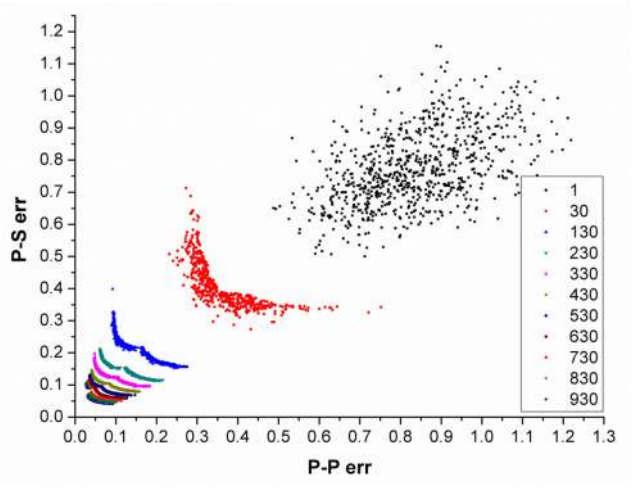

Figure 10. Evolution of all solutions through generations under the defined error functions or objectives given as the misfit between synthetic data and real data measured as a cross-correlation. As shown in the legend, the dots in different colours represent different generations. Note how the solutions, widely distributed in the objective space, slowly converge to the most optimal set of solutions over generations. Also note that although the objectives are defined as the *P–P* error and *P–S* error, they are, in fact the scaled misfits between the observed and synthetic vertical and horizontal inline data components.

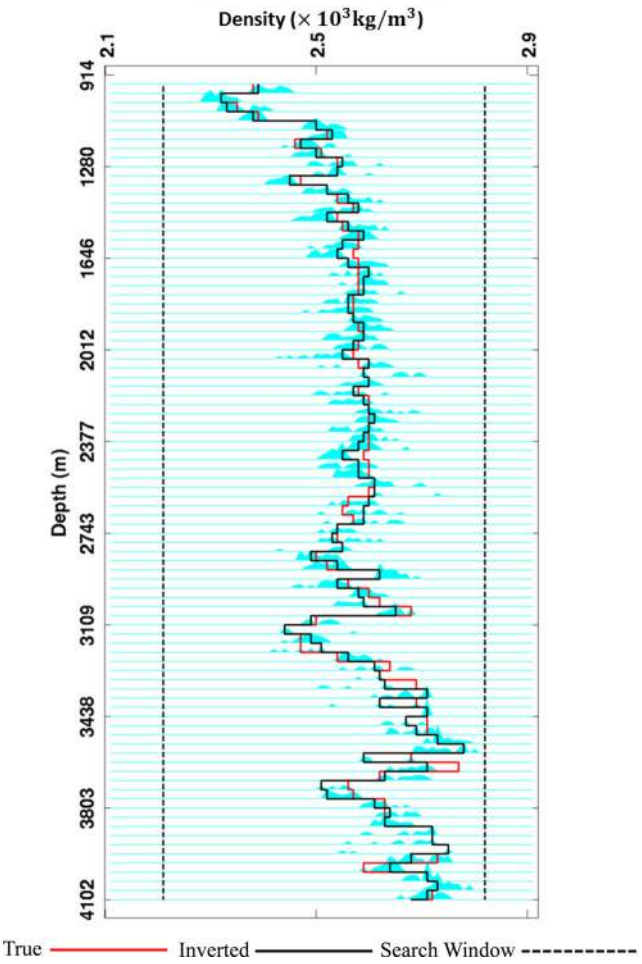

Figure 11. Comparison between the true densities with the inversion results after 5000 generations. Here, the entire pre-stack data were inverted under an azimuthally anisotropic assumption. Normalized PPD of the inverted results is also plotted in light blue (cyan).

implement a methodology capable of continuous sampling of this space. This is achieved by using a real coded GA with simulated binary crossover (SBX) and real parameter mutation (RPM) as outlined by Deb & Agrawal (1995, 1999). Exact mathematical details of SBX and RPM have been given by Padhi & Mallick (2013b) and here we just briefly outline them. In addition to the probability of crossover (P_c), SBX uses an additional parameter—the crossover distribution index η. A high value of η tends to produce children close to their parents while a low value of η tends to produce them away. In addition to the probability of mutation (P_m) RPM is controlled by another parameter—the mutation distribution index κ, a small value of which produces the parameter of the mutated solution far away from the original solution and a large value produces a solution close to it. In inverting multicomponent seismic data for isotropic and VTI subsurface properties, Padhi & Mallick (2013b) found that a moderate to high value of P_c between 0.7 and 0.9 and low value of P_m between 0.03 and 0.1 are reasonable choices. They also found that using $\eta = 20$ and $\kappa = 10$ provides reasonable inversion results. However note that instead of keeping these parameters constant, they should ideally be varied over generations. At the beginning when none of the solutions are near the global optimum, one should ideally use a high value of P_c and a low value of η so that the model space is exhaustively sampled by performing crossover often and producing child solutions away from their parents.

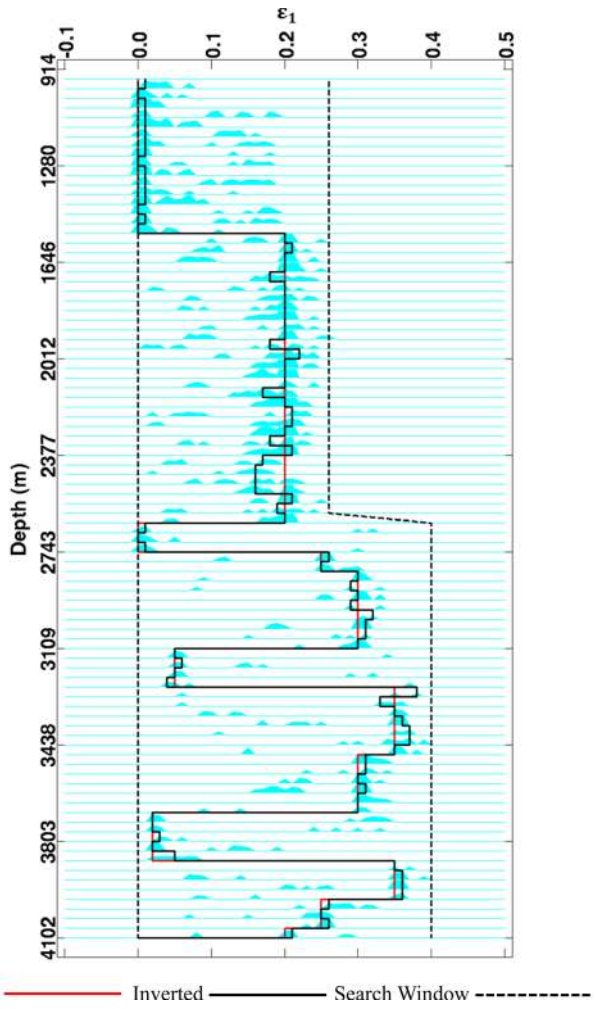

Figure 12. Same as Fig. 11, but for the ε_1.

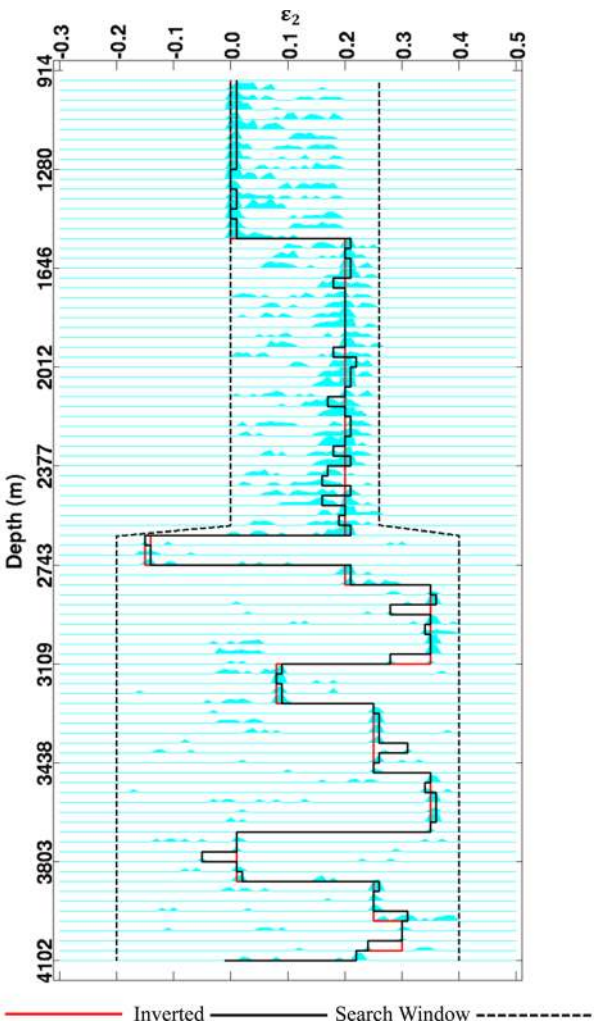

Figure 13. Same as Fig. 11, but for the ε_2.

However as the generation progresses, P_c should be slowly reduced and η should be slowly increased so that the good solutions aren't lost. Based on the same argument, P_m should be high at the beginning and be reduced over the generations and κ should be low at the beginning and increased over the generations. Besides scaling of the objectives, non-dominated sorting, and crowding distance computation based on both model and scaled objective spaces, varying P_c, P_m, η and κ over generations would provide NSGA II additional robustness not only in preserving diversity and propagating good solutions over generations, but it would also allow an exhaustive sampling of the model space. It will be shown later that by varying these SBX and RPM parameters over generations produced good results in our anisotropic inversion.

Parallel forward modelling algorithm

The reflectivity method, used as the forward modelling engine in our application is computed in the plane-wave domain. As mentioned before, computing one plane-wave seismogram is independent of the other. Thus they could be simultaneously computed using different compute-nodes in a parallel computing environment. Being a stochastic optimization method, NSGA II requires computing many (of the order of millions) plane-wave seismograms during its process. Consequently, parallelizing the forward computation can substantially improve the computational efficiency.

MULTIPHASE INVERSION OF THE DATA

For any inverse problem it is always desirable that the number of parameters being solved for is reduced, because of the fact that the computational cost of stochastic search algorithms increases with the increasing number of unknown parameters to be evaluated. By comparing the near-offset synthetic data generated by using full anisotropic model and the near-offset data generated by using isotropic models with same vertical *P*- and *S*-wave velocities, and densities, we observed that the differences between them are not large. It is thus possible that the near-offset data could be first inverted under an isotropic assumption to obtain an estimate of the velocities and densities, which could then be followed by an anisotropic inversion of the full offset data with the near offset inversion results being used as constraints. By doing this, the computational cost could be significantly reduced. This is because of the fact that for the anisotropic inversion the background isotropic model does not need to be re-estimated, which will reduce the number of model parameters, and therefore the overall computation cost. For our application, we therefore implemented such a multiphase inversion, that is, we first inverted the near offset traces for the background isotropic properties. The vertical *P*- and *S*-wave velocities obtained from this isotropic inversion of near offset traces were then

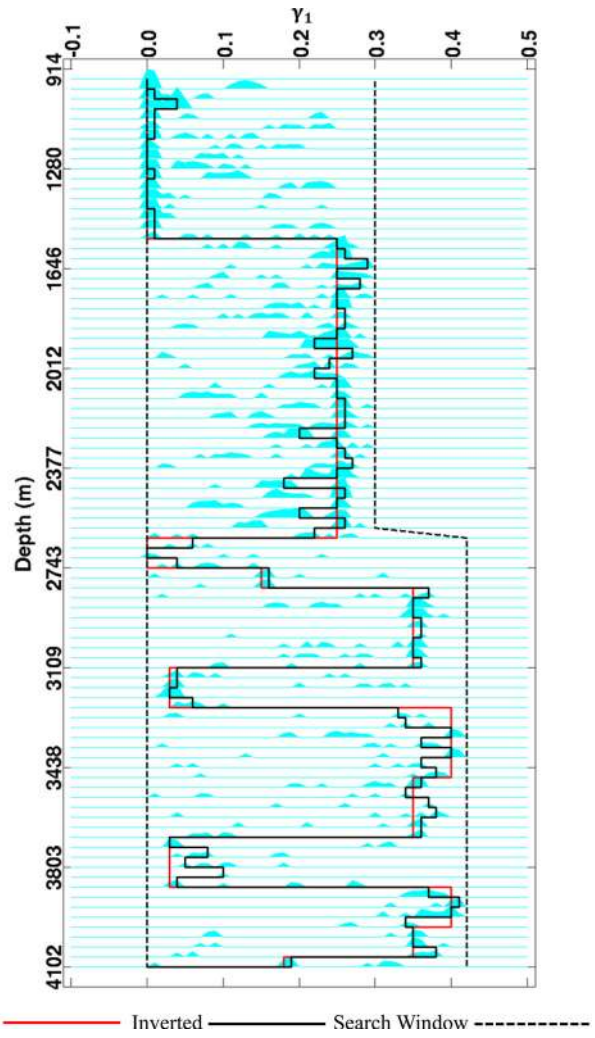

Figure 14. Same as Fig. 11, but for the γ_1.

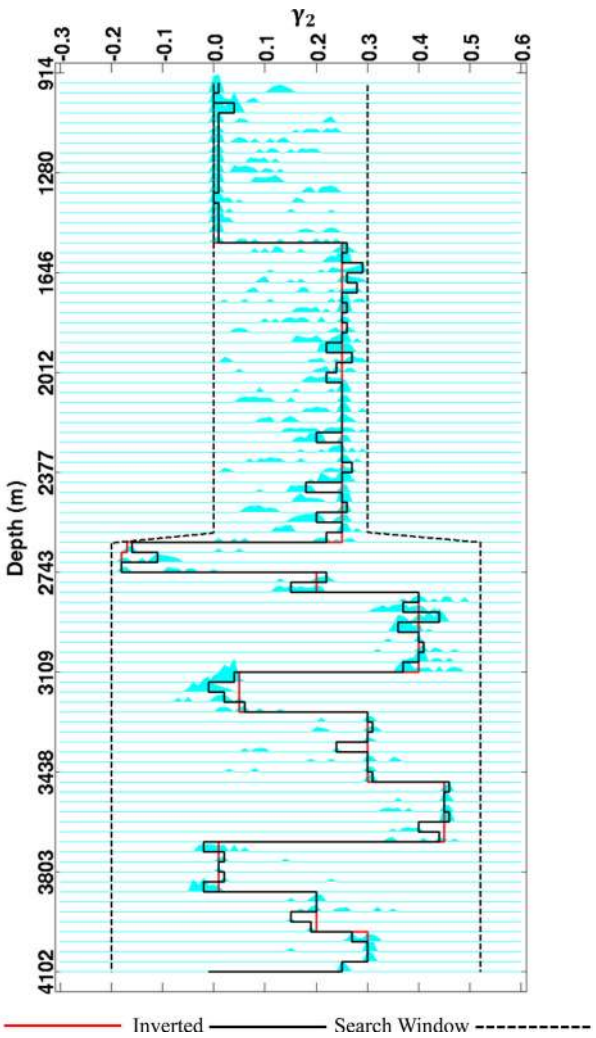

Figure 15. Same as Fig. 11, but for the γ_2.

used for the subsequent step of the anisotropic inversion of the full offset data.

NUMERICAL EXAMPLE

We generated a stratified model of *P*- and *S*-wave velocity (V_p and V_S) and density (ρ), based on an actual well log data from the Rock Springs uplift (Well RSU-1), WY, USA. Fig. 2 shows the schematic diagram of the synthetic model that was used to test our inversion method. The model had different types of anisotropic layers with different directions of the symmetry planes. The original well log data were blocked applying the Sequential Backus Averaging (Lindsay & Koughnet 2001) and the thickness of each layer was about 36.5 m. The blocked model had 113 layers, specifically, 39 isotropic layers, 30 VTI layers, 3 HTI layers, 40 ORT layers and an isotropic half-space. Assuming that the first 23 isotropic layers in the overburden are known, our model vector consisted of a total number of 689 parameters. Figs 3(a) and (b) show the actual RSU-1 logs and the blocked well log that was used in our inversion. Fig. 3(c) shows the Thomsen–Tsvankin anisotropy parameters that were used to introduce anisotropy in the RSU-1 model. Additionally, azimuthal directions of the vertical symmetry planes in HTI and ORT layers were assigned based on the *in situ* stress field analysis of the real RSU seismic data by Mukherjee *et al.* (2012). The entire model shown in Fig. 3 was used to compute synthetic three component (vertical and horizontal inline and crossline components of particle displacement from an explosive point source) seismograms in the intercept time–ray parameter ($\tau-p$) domain using the reflectivity method along two azimuths (N30°E and N75°E). These computed synthetic data in the $\tau-p$ domain are shown in Figs 4–6. Note that these synthetics were generated using a zero-phase Ricker wavelet with a central frequency of 35 Hz and for our inversion tests, we added 3 per cent random noise to these data.

We first used our parallel NSGA II to invert the near-offset two component (vertical and horizontal inline) data under isotropic assumption for the vertical *P*- and *S*-wave velocities (V_P, V_S) and density (ρ). The near-offset data consisted of 20 traces spanning 0–0.02 s km^{-1} on the ray parameter axis. We chose a low range for ray parameter because of the fact that we wanted to restrict to the near offsets only for this isotropic inversion. In this numerical experiment, we used a population size of 800, $P_m = 0.04$, $P_c = 0.90$, and the number of generations $= 930$. We also used 20 for the crossover index η and 10 for the mutation index κ. Note that these choices were based on the isotropic and VTI inversion results using NSGA II given by Padhi & Mallick (2013b). Next, to resolve anisotropic parameters, we inverted the three-component full-offset

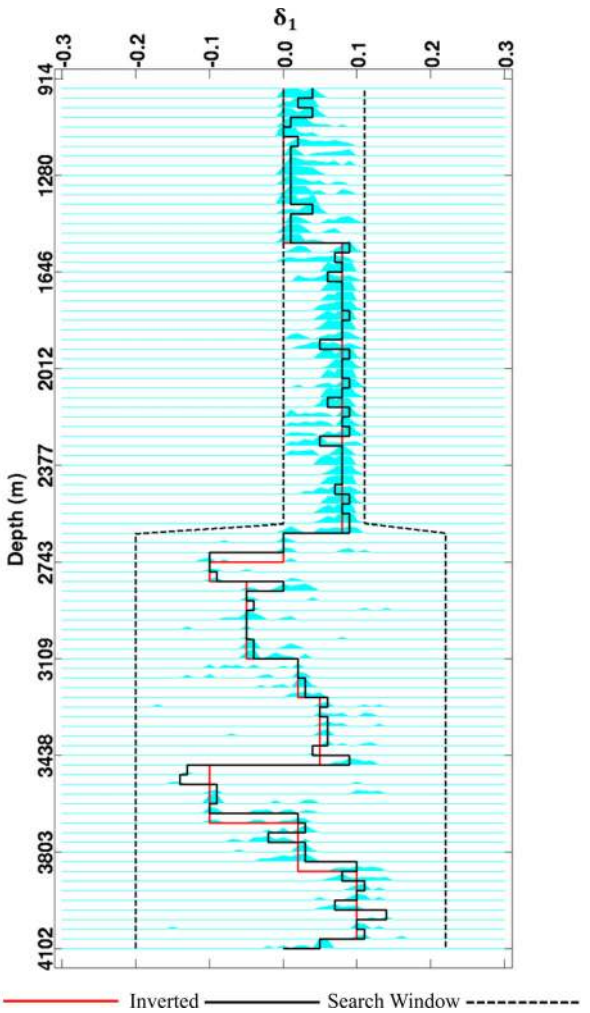

Figure 16. Same as Fig. 11, but for the δ_1.

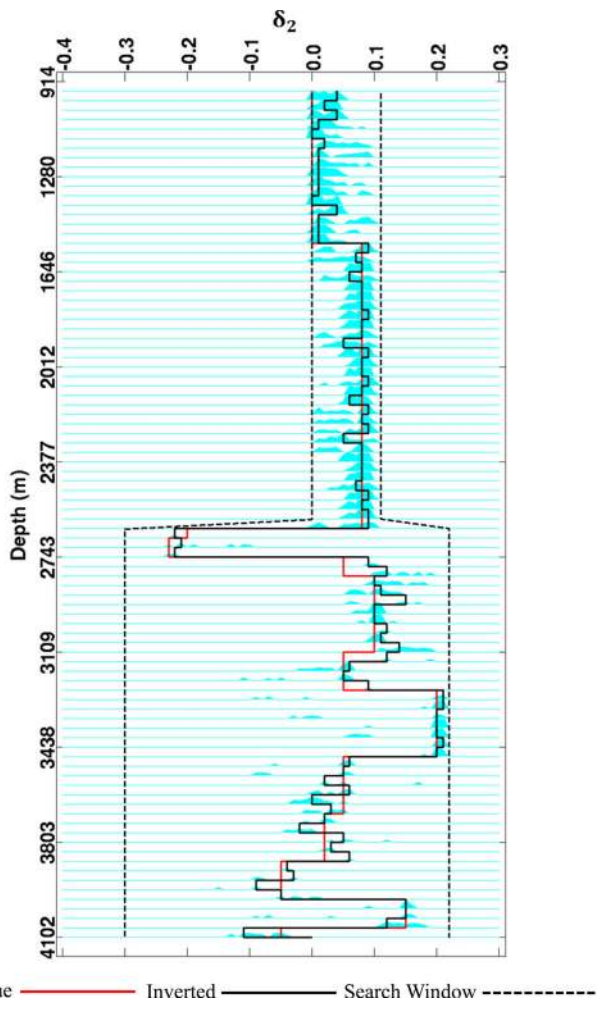

Figure 17. Same as Fig. 11, but for the δ_2.

data with 36 traces with ray parameter ranging from 0–0.18 s km^{-1} along two different azimuths in the $\tau - p$ domain with an anisotropic assumption. For this anisotropic inversion, we used a population size of 4000 and a number of generations of 5000. Additionally, we used the estimated V_P and V_S from the near-offset isotropic inversion as hard constraints and only estimated ρ and other anisotropic parameters. For this inversion, we also found that varying P_c, P_m, η and κ over generation provides better convergence instead of keeping them constant. For the of probability of crossover P_c we varied it linearly with generation as $P_c = 0.7 - 0.1 \times \frac{t}{t_{max}}$, and for the crossover distribution index η we varied it as $\eta = 1.0 + 19.0 \times \frac{t}{t_{max}}$. In these linear functions, t and t_{max}, respectively denote the current generation number and the maximum number of generations. Note that based on our previous arguments, in our choice of generation dependent P_c and η, we deliberately chose P_c to decrease and η to increase with generation such that the model space is widely sampled at the beginning and is slowly reduced over generations as we approach the true solution. Following Deb & Agrawal (1999), to get a mutation effect of 1 per cent perturbance in solutions out of the entire set of population, we varied the mutation index as $\kappa = 100 + t$ and the probability of mutation as $P_m = \frac{1}{n} + \frac{t}{t_{max}} \times (1 - \frac{1}{n})$, where n is the total number of variables (model parameters). Note here that similar to the choice of the generation-dependent P_c and η, the choice of the generation-dependent P_m and κ as above were also motivated by maintaining diversity in the population at early stage and slowly reduce it over time as the solutions near the global minimum.

RESULTS

All inverted models shown in the following figures were obtained by first estimating the Pareto-optimal set of solutions, and then choosing a solution with the highest likelihood of the *a posteriori* probability density (PPD) of the estimated Pareto front. Because the true PPD for multi-objective inverse problems is difficult to estimate (Padhi & Mallick 2013b), we plotted all solutions of the rank-1 members in the model (decision) space in a normalized histogram display as an estimate of the approximate PPD (for details, see Mosegaard & Tarantola 1995; Padhi & Mallick 2013b), and the highest likelihood solutions were estimated from these approximate PPD plots.

Figs 7–10 show our inversion result of near-offset pre-stack data under an isotropic assumption. In this study, we first filtered the *P*- and *S*-wave models by applying a 10 Hz high-cut filter, and then used a reasonably wide (±20 per cent of the filtered velocity profiles) search window for vertical *P*- and *S*-wave velocities, and the search window for density was set constant in the range of 2.2 and 2.8 g cm^{-3}. Even though our search windows were somewhat arbitrary, in a practical scenario it is expected to be similar. For example, the initial guess of the low-frequency vertical *P*- and *S*-wave

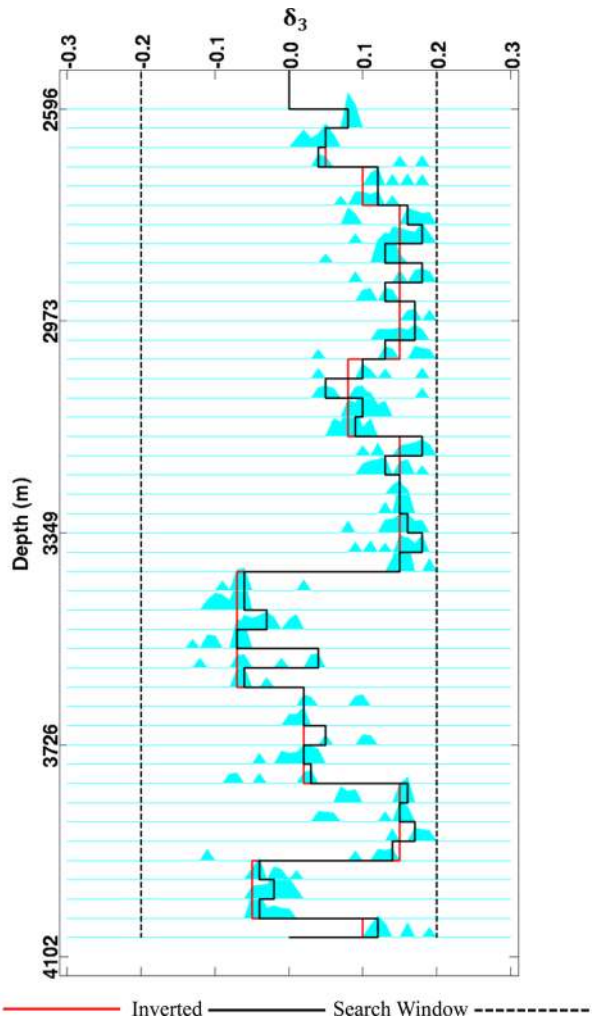

Figure 18. Same as Fig. 11, but for the δ_3.

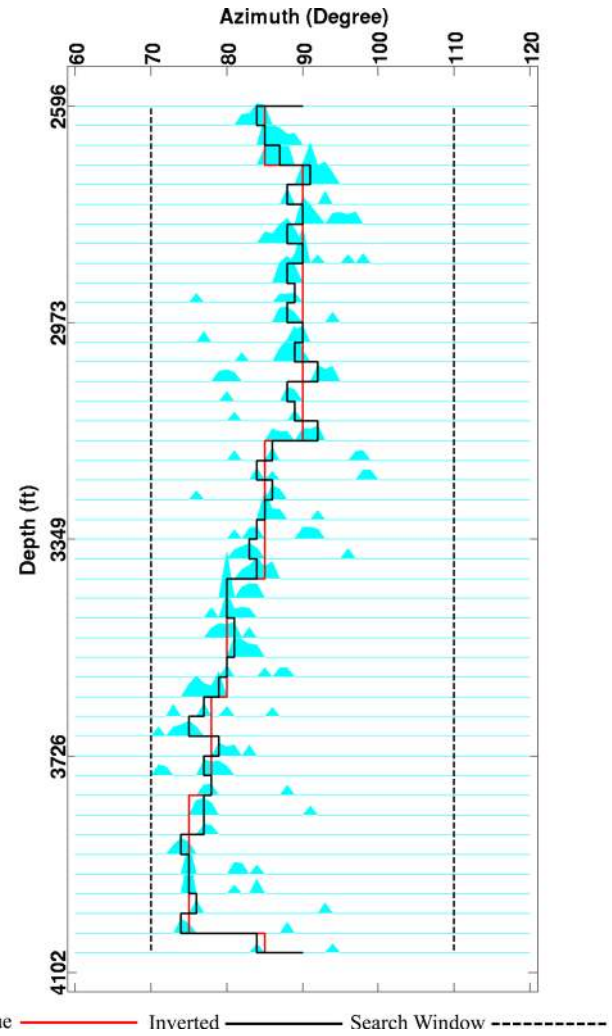

Figure 19. Same as Fig. 11, but for the azimuthal angle of vertical symmetry planes.

velocities can be estimated by correcting the near-offset pre-stack data for normal moveout (NMO), and using a search window similar to ours around the NMO estimated interval velocities is therefore a practical choice. Although velocity–density relations are available in the literature (for example, Gardner *et al.* 1974; Castagna 1991, among others), or it could be estimated directly from the wireline logs, in general establishing a background density trend is much more difficult than establishing the background velocity trends. Consequently, we chose to use a flat and wide search window for density to find out how well our inversion method is capable of extracting the true density values. Fig. 7 compares the inverted V_P with true V_P along with the search window used for the isotropic near-offset inversion. Figs 8 and 9 are same as Fig. 7 but they are respectively for V_S and ρ. Fig. 10 shows the evolution of solutions through generations under the defined objective functions—the *P*–*P* error on the horizontal, and the *P*–*S* error on the vertical axis. Note from Fig. 10, how the models that are randomly and widely distributed in the objective space at the beginning converge to the optimal solution set over generations. Note here that although we denote the objectives as *P*–*P* and *P*–*S* errors in Fig. 10, they actually are the estimates of the scaled vertical and horizontal inline components of the misfit functions as defined by eq. (6). It can be noted from Figs 7–9 that near-offset vertical and horizontal inline waveform data could estimate the vertical *P*- and *S*-wave velocities quite well, but failed to obtain a good estimate of the density. Including wide angle reflections in waveform inversions allow better estimation of the subsurface properties than using just the low angle reflections (Virieux & Operto 2009). Therefore, poor estimation of density using near offset data only is not surprising. However note here that the reason for restricting to near offsets only is due to the fact that here we applied inversion under the assumption of isotropy while the subsurface is azimuthally anisotropic. Consequently, inverting the near offset data only is the most viable choice where the azimuthal anisotropy effects are not large and the background isotropic model could be extracted.

Figs 11–20 show our anisotropic inversion result of the full-offset pre-stack data, that used the estimated V_P and V_S from the near-offset isotropic inversion as constraints. Figs 11–19 respectively show ρ, Thomsen–Tsvankin anisotropic parameters, and the azimuthal directions for vertical symmetry planes in the HTI and ORT layers. Note from Figs 12–18, the search windows for anisotropic parameters for the isotropic, VTI and HTI layers were set slightly narrower than those for the ORT layers. There was no specific reason for using different search windows, but note that even though the search windows used for the isotropic, VTI and HTI layers were a bit narrower than those used for the ORT layers, they still sample a large variation of models in the model space and the inversion results won't be different if they were kept same for all layers. Finally,

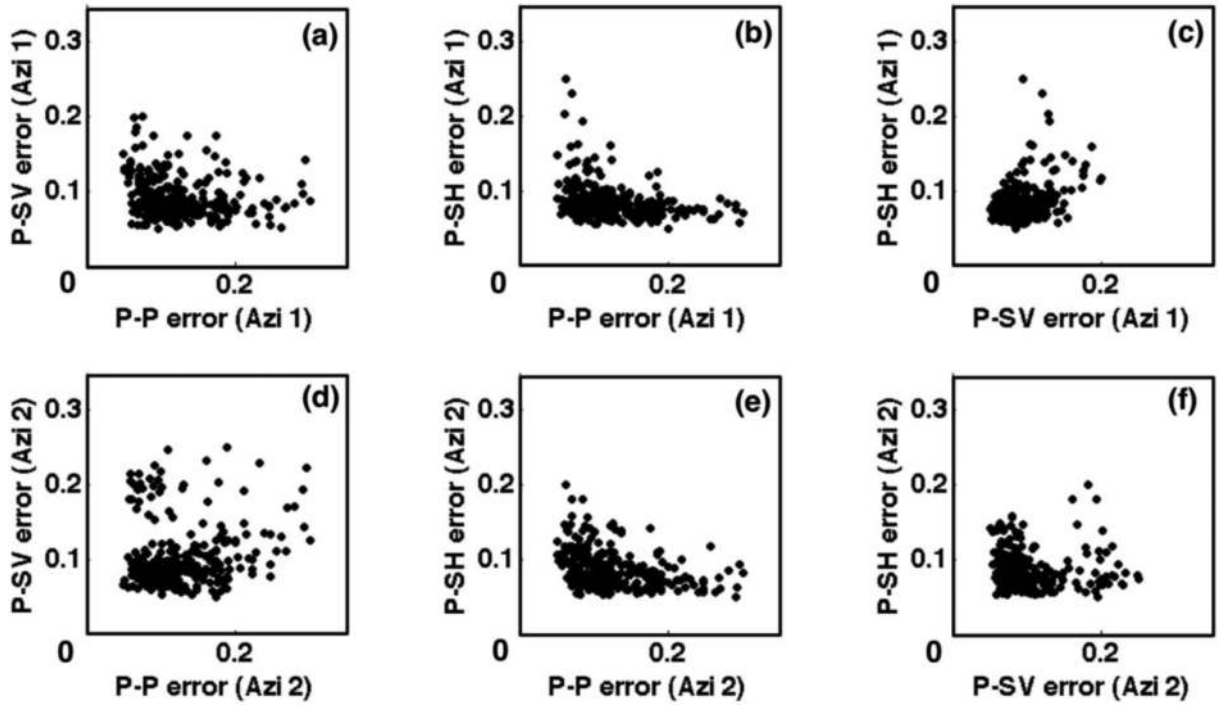

Figure 20. Cross-plots of different data errors for the three-component data inversion under full anisotropic assumption. In all these plots, the *P–P* error is the scaled misfit in the vertical component response, $P–S_V$ error is the scaled misfit in the horizontal inline response, and $P–S_H$ error is the scaled misfit in the horizontal crossline response. (a) *P–P* error versus $P–S_V$ error along azimuth 1, (b) *P–P* error versus $P–S_H$ error along azimuth 1, (c) $P–S_V$ error versus $P–S_H$ error along azimuth 1, (d) *P–P* error versus $P–S_V$ error along azimuth 2, (e) *P–P* error versus $P–S_H$ error along azimuth 2 and (f) $P–S_V$ error versus $P–S_H$ error along azimuth 2. Horizontal and vertical axes in all these displays represent the errors for the corresponding data components they represent. Each of these cross-plots represents different 2-D sections of the estimated 6-D Pareto-optimal front.

Fig. 20 shows the error analysis plot, or different 2-D cross-sections of the estimated 6-D Pareto-optimal front for our final anisotropic inversion result. As for the isotropic inversion, 'P', 'SV' and 'SH' used in the axes labels of Fig. 20 respectively denote the 'vertical', 'horizontal inline' and 'horizontal crossline' components of the particle displacement.

DISCUSSION

We tested two-step inversion of noisy synthetic data generated from a realistic earth model derived from a real well-log with isotropic, VTI, HTI and ORT subsurface layers:

(1) The near-offset two-component data along one azimuth was first inverted for the background isotropic properties.
(2) The entire-offset three-component data along two different azimuths was then inverted under full anisotropic assumptions with inverted results from (1) used as constraints.

For step-1 (Figs 7–10), we found that while the vertical *P*- and *S*-wave velocities could be recovered from the near-offset data to a reasonable accuracy, the estimation of density is rather poor. We therefore believe that

(1) Near-offset data could be inverted for the vertical *P*- and *S*-wave velocities.
(2) These velocities could then be used to constrain the vertical velocities for an anisotropic inversion of the entire-offset data, which, in turn, can significantly reduce the computation expense of the anisotropic inversion.
(3) Density could not be accurately estimated from the near-offset isotropic inversion. When the subsurface is anisotropic, an accurate estimation of density requires full offset data.

As demonstrated in Figs 11–21, using the background velocities from near-offset isotropic inversion as constraints, all anisotropic parameters including density could be recovered quite well from the anisotropic inversion. Not only the extracted model is close to the true model, the PPD plots, shown as light blue (cyan) curves in Figs 11–19, show well-constrained model estimates. In addition, the widths of these PPD plots provide an estimate of the uncertainty associated with the estimation of each parameter. Note that the vertical resolution of in our inversion results is approximately quarter wavelength of the dominant seismic frequency. Because we used a 35 Hz Ricker wavelet and the maximum vertical P-wave velocity in our model was approximately 6000 m s^{-1} (see Fig. 3), our vertical resolution was 43 m or lower.

In the examples presented here we used tight constraints for all the model parameters of the first 23 layers which represent the isotropic overburden and they constitute the top 900 m of the model. This is similar to a situation when inversion is applied over a certain data window assuming the overburden structure to be known. It must however be noted that assumptions about the overburden is likely to have an effect on the inversion of field data as the error will propagate into the deeper layers.

In this work, we have focused only on horizontally stratified (1-D) earth models. However in a practical situation, our algorithm can be applied to field data in a number of scenarios after the seismic data are migrated. Although 1-D assumptions exclusively made in our algorithms could be applicable to real data after migration for relatively simple structures, it is likely to fail for geologically

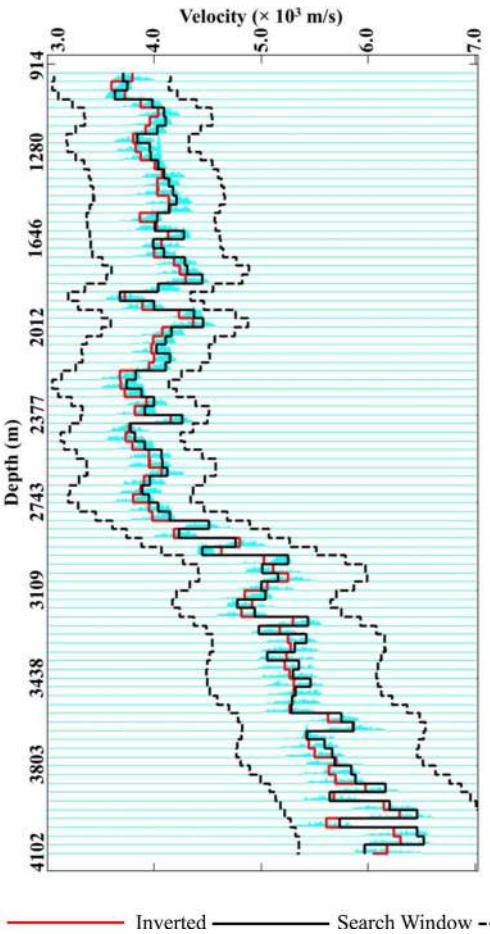

Figure 21. Comparison between the true V_P with the inverted V_P. Here, the full-offset pre-stack data were inverted using a full anisotropic assumption, and the inversion results from near-offset data were used to narrow down the search window. Normalized PPD of the inverted results is also plotted in light blue (cyan).

complex structures. While a 3-D anisotropic PWI is ideal for handling a geologically complex subsurface, such a method is still in early stage of its development (Warner *et al.* 2013). At the same time, a methodology capable of handling at least the local 3-D structures for every source–receiver offset/ray-parameter/angle pair is required so that the inversion could be applied to real data where the subsurface is relatively complex. Jiao *et al.* (2000) developed a residual migration velocity analysis in τ–p domain and showed how local dips could be incorporated and the same methodology could be easily implemented into PWI. In addition, ray-based offset-to-angle transform outlined by Mukhopadhyay & Mallick (2011) could also allow incorporating local dips into the PWI. These issues are currently being investigated and will be discussed in a separate paper in connection with inversion of real data.

Although here we applied our inversion to surface seismic data, the flexibility of the reflectivity method for the forward synthetic computation allows extending the method to other data types such as the ocean–bottom cable (OBC) or vertical seismic profile (VSP) data (see Mallick & Frazer 1988 for details). In addition, the inversion of three-component data demonstrated here could be easily extended to invert more data components such as nine component (9C, i.e. three component sources and receivers) data if available. While 9C surface seismic acquisitions are not common, they are frequently recorded in walkaway VSP geometries and extending our inversion methodology for such data could be of practical interest.

Being population-based global search methods, all MOEAs are capable of finding solutions as close as possible to the true global Pareto front. Since a large number of solutions are evaluated during run, the solutions can be stored and the approximate PPD plots can be generated to get an idea of the uncertainty of the evaluated model parameters. While Kozlovskaya *et al.* (2007) claim that the quality and the confidence interval for multi-objective inverse problems could be estimated from the Pareto-optimal fronts alone, it was pointed out by Padhi & Mallick (2013b) that as the number of objectives increase, combining the Pareto-optimal fronts with the estimates of the PPD is more appropriate. Not only do they allow interpreting how well behaved the inverted solutions are, but the width of each PPD curve allows us to quantify the uncertainties associated with the estimates of each model parameter for each layer. In our displays of the estimated models from inversion (Figs 7–9 and 11–19), we have also displayed the corresponding PPD estimates in the background. The quality of these estimated PPD values provide useful information on the inversion. For example, poor estimation of density from the near-offset isotropic inversion (Fig. 9) is reflected in the corresponding PPD plots as well. Similarly, reasonably good estimates of velocity from near-offset isotropic inversion (Figs 7 and 8) and those of density and other anisotropic parameters from the full-offset anisotropic inversion (Figs 11–19) are visible in tight and well-constrained nature of the corresponding PPD plots.

One important point we must note here in connection with how we estimate the PPD. Although our estimates are perfectly coherent with the current state-of-the-art, we must note that the PPD we used in this study is not a true posterior distribution in the Bayesian sense. GA is a stochastic optimization method and tends to underestimate the posterior uncertainty (Sen & Stoffa 1996). This is in fact true for all stochastic optimization algorithms such as GA, simulated annealing (SA), particle swarm optimization (PSO), etc. The main advantage of these stochastic methods is that they can avoid local minima by exhaustively sampling the model space, however models sampled 'away' from the global minima do not contribute to the estimation of the posterior uncertainty which in the Bayesian sense they should, if they have non-zero likelihood in the prior. Therefore what we get by modelling a probability density function (PDF) from the optimized ensemble is not a posterior PDF in the Bayesian sense but simply a normalized likelihood function. To estimate the true posterior PDF, it is necessary to employ a more exhaustive sampling method such as a Gibb's sampler (Sen & Stoffa 1996). However, using such, especially for anisotropic seismic inverse problem is computationally prohibitive. Therefore for our application here we have regarded the normalized likelihood estimates as approximation of the posterior PDF even though earlier studies have demonstrated that our approach would underestimate the uncertainties.

Our proposed methodology has some computational advantages:

(1) By inverting near-offset pre-stack data under isotropic assumption first, the values of vertical *P*- and *S*-wave velocities could be constrained within an acceptable range. These isotropic results could be utilized to reduce the number of parameters to be estimated for in the anisotropic inversion using the full range of offsets. This substantially reduces the overall computational cost.

(2) The parallel implementation of NSGA II further increases the computational efficiency of the algorithm. As all MOEA methods require large population size while dealing with complex nonlinear problems, the parallelization suggested here is of crucial importance.

(3) By scaling the objectives prior to ranking and computing crowding distance, redefining the crowding distance in terms of the normalized distances measured both along the scaled objective space and the model space, and varying the crossover and mutation parameters over generations, we believe that not only we presented here a parallel implementation, but we also provided an improvement of the original NSGA II. This is especially true for problems such as the multicomponent seismic waveform inversion where the number of model parameters to be estimated is much higher than the number of objectives.

Using 10 nodes with 4 cores per node Intel Xeon W5580 @ 3.2G Hz processors with 48 Mb of memory, our current implementation of isotropic inversion takes approximately 10 hr. The full anisotropic inversion on the other hand takes approximately 24 hr using 20 similar nodes. Although these runtimes could be significantly reduced using more nodes, we still think that further investigations are necessary to improve the computational efficiency of our methodology along the following lines:

(1) In the reflection matrix method used for forward modelling, the calculation is actually done in the frequency-slowness ($\omega-p$) domain. Because the reflection matrix is built from the top down (see eq. 4), the behaviour of propagating seismic waves in the lower layers is independent of the layers above them. Therefore a sliding inversion window in $\tau-p$ domain for each component of the data could be applied from top to the bottom of the whole dataset. Inversion the pre-stack seismic data in $\tau-p$ domain part by part could significantly reduce the number of parameters that are being solved for in the model space, and thus could significantly improve its computational efficiency.

(2) The search algorithm currently implemented in our method samples the model space via crossover and mutation. This sampling is somewhat random, and in many instances, the process of crossover and mutation results in a model that is not physically meaningful such that the elastic stiffness matrix $[C_{ij}]$ is not positive definite. As NSGA II generates new members, it is mandatory to check whether the generated models satisfy the positive definite conditions (Appendix A) and accept them only when they do. We believe that a robust and advanced sampling strategy of the model space by incorporating rock physics into NSGA II could potentially increase the sampling efficiency and lead to a better convergence of the methodology compared to our current implementation.

While our proposed two phase inversion is computationally efficient, it has a fundamental problem with the estimation of the other model parameters. Once the V_P and V_S are fixed from the near-offset isotropic inversion, the other parameters being inverted for are biased by the choice of their values. Additionally, if the methodology is extended to include sliding windows, the model parameters fixed for the overburden are likely to bias the estimates for the current window that is being inverted for. In our near-offset isotropic inversion (Figs 7–9), note that although our estimates of V_P and V_S were reasonable, they are not exact. Consequently, fixing their values may impact on the final inversion result. So, we also ran a full anisotropic inversion where V_P and V_S, in addition to ρ and other anisotropic parameters were inverted for. For this inversion we used the estimated V_P and V_S from the near-offset inversion as constraint and used a narrow search window ($\pm$5 per cent) to estimate them. These inverted values of V_P and V_S, shown in Figs 21 and 22 demonstrate that this leads us to achieve a better estimation of these two parameters (compare Figs 21 and 22, respectively with Figs 7 and 8). Fig. 23 compares the waveform fit between the observed and predicted data

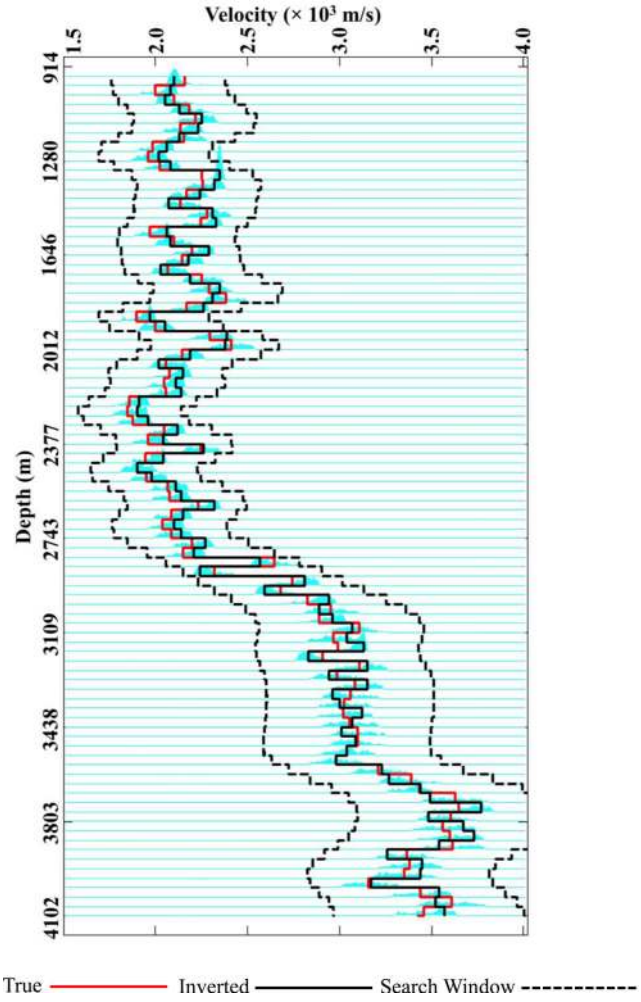

Figure 22. Same as Fig. 21, but for V_S.

for all three components along a single azimuth (N75°E) for this anisotropic inversion demonstrating the fact that all major features on the observed data were predicted quite well by the final inversion model. From these results we think that for the two-step inversion, it may be necessary to invert for V_P and V_S in addition to ρ and other anisotropic parameters during the full anisotropic inversion. While the estimates of V_P and V_S from the near-offset inversion help reducing their search windows during the anisotropic inversion, please note this is likely to take additional computation time than the case where they are not inverted for. Consequently, further investigations are required to improve the computational efficiency of our proposed multiphase inversion.

As noted above, by making the probabilities of crossover and mutation (P_c and P_m) and the crossover and mutation distribution indices (η and κ) to vary with generation helped the anisotropic inversion. Our choice of the specific linear functions that were used to vary each of these parameters, although were based on some preliminary experiments, we must admit that we did not make rigorous investigations on how each of them must be optimally varied. Therefore we think that investigations are also required to find out how each of these parameters should be changed to obtain the best results.

Although we applied our inversion to noisy synthetic data our noise is random. In most field data, the noise could however be coherent. For example, although the interbed multiple reflections

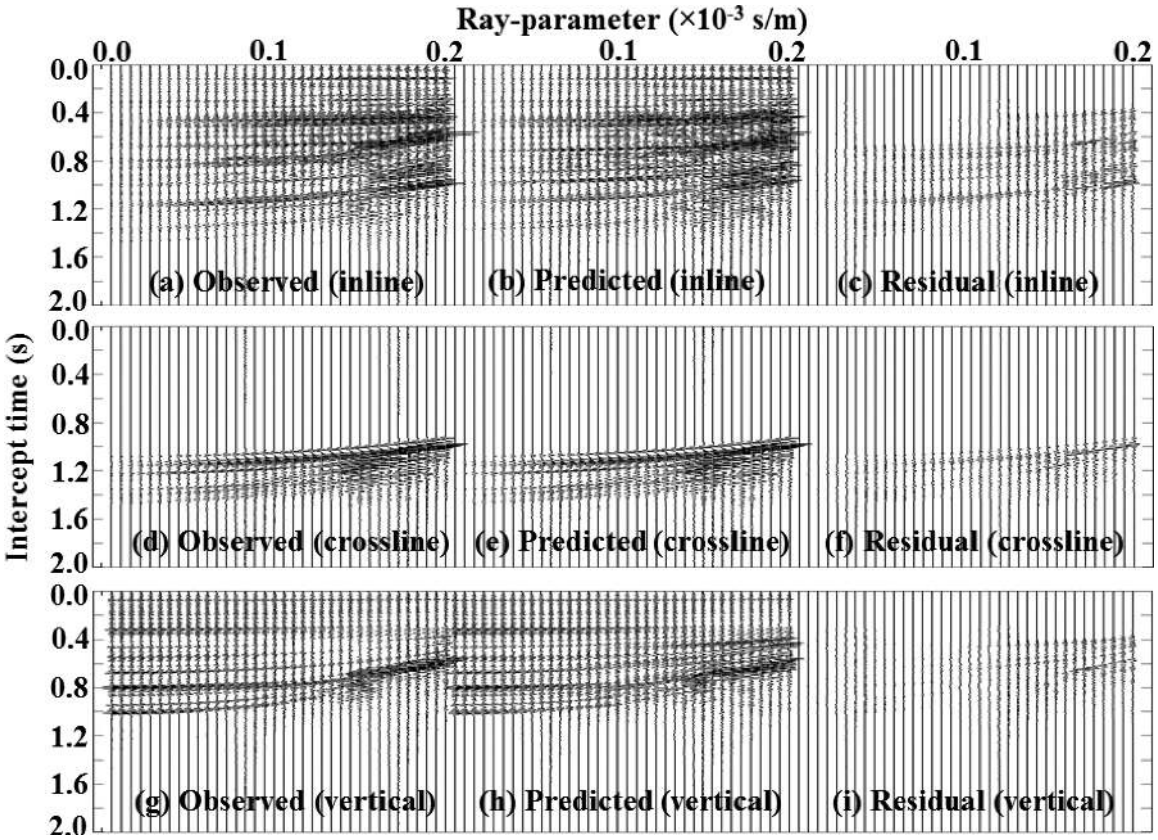

Figure 23. Waveform fitting for the anisotropic waveform inversion along a single azimuth (N75°E). The result along the other azimuth is similar and therefore not shown. (a) Observed data, horizontal inline component, (b) Predicted (synthetic) data, horizontal inline component, (c) Data residual, horizontal inline component. (d)–(f) Same as (a)–(c), but for the horizontal crossline component. (g)–(i) Same as (a)–(c), but for the vertical component.

are included in our algorithm, the multiples that our method will honour are only those generated within the inversion window. In real data, multiples from the overburden layers that are not being inverted for may contaminate data within the inversion window. Although advanced processing methods are available to attenuate such coherent noise, it is often difficult to assess their effects on the amplitudes and therefore the validity of applying any waveform based method. We are in the process of applying our inversion to real multicomponent seismic data where these practical issues will be addressed.

CONCLUSION

Estimating elastic parameters and density from seismic waveform data is a non-linear problem with non-unique solutions within acceptable error limits of fitting the data. Such non-uniqueness can be tackled by using multicomponent data to constrain the various model parameters being inverted for. Casting the problem of inverting such multiple data sets as a multi-objective optimization problem, we have developed and used a global search method for finding compromise solutions that can simultaneously fit all the data components to reasonable accuracy. We applied the method to invert noisy synthetic pre-stack seismic waveform data generated using a multilayer model obtained from a real well log with isotropic, azimuthally independent (VTI) and azimuthally dependent (HTI, ORT) anisotropic layers. Our results show that inverting two-component near-offset data under isotropic assumptions followed by a three-component anisotropic inversion of the full offset data along two azimuths can successfully retrieve the subsurface elastic anisotropic parameters. Even though our proposed methodology is valid for 1-D subsurface, it could be applicable to real data after pre-stack migration for many simple geological settings of practical interest. In addition, we also outline methodologies how our proposed method could be extended to 3-D structures such that more complex geology could be handled. We therefore believe that our method could potentially be applicable to invert real multicomponent seismic data. Further investigations are however necessary to improve the computational efficiency of the method.

ACKNOWLEDGEMENTS

We thank the GJI Editor, Dr Rene-Edouard Plessix and two anonymous reviewers for their valuable comments that improved the quality of the manuscript. We also thank Amit Padhi, Dario Grana, Sven Treitel and Jim Gaiser for critically reading an early version of the manuscript and suggesting improvements. The work was supported by the Energy GA fund from the University of Wyoming, and a research grant on unconventional resources from the School of Energy Resources, University of Wyoming.

REFERENCES

Aki, K. & Richards, P.G., 2002. *Quantitative Seismology,* University Science Books.

Auld, B.A., 1973. *Acoustic Fields and Waves in Solids,* Vol. 1, John Wiley and Sons.

Castagna, J.P., 1991. Petrophysical imaging using AVO, *Leading Edge,* **12**(3), 172–178.

Chang, H. & McMechan, G., 2009. 3D 3-C full-wavefield elastic inversion for estimating anisotropic parameters: a feasibility study with synthetic data, *Geophysics,* **74,** WCC159–WCC175.

Chaveste, A., Zhao, Z., Altan, S. & Gaiser, J., 2013. Robust rock properties through PP-PS processing and interpretation-Marcellus Shale, *Leading Edge,* **32,** 86–92.

Deb, K. & Agrawal, R.B., 1995. Simulated binary crossover for continuous search space, *Complex Syst.,* **9,** 115–148.

Deb, K. & Agrawal, S., 1999. A niched-penalty approach for constraint handling in genetic algorithms, in *Proceedings of the ICANNGA-99,* Portoroz, Slovenia.

Deb, K., Pratap, A., Agarwal, S. & Meyarivan, T., 2002. A fast and elitist multi-objective genetic algorithm: NSGA-II, *IEEE Trans. Evol. Comput.,* **6**(2), 181–197.

Dewangan, P. & Grechka, V., 2003. Inversion of multicomponent, multi-azimuth, walkaway VSP data for the stiffness tensor, *Geophysics,* **68,** 1022–1031.

Fryer, G.J. & Frazer, L.N., 1984. Seismic waves in stratified anisotropic media, *Geophys. J. R. astr. Soc.,* **78,** 691–710.

Fuchs, K. & Müller, G., 1971. Computation of synthetic seismograms with the reflectivity method and comparison with observations, *Geophys. J. R. astr. Soc,* **23,** 417–433.

Gaiser, J., Verm, R. & Chaveste, A. 2013. Pseudo S-wave broadband response of C-waves after domain change, *Leading Edge,* **32,** 50–62.

Gardner, G.H.F, Gardner, L.W. & Gregory, A.R., 1974. Formation velocity and density—the diagnostic basis for stratigraphic traps, *Geophysics,* **39,** 770–780.

Goldberg, D.E., 1989. *Genetic Algorithms in Search, Optimization & Machine Learning,* Addison-Wesley.

Jiao, J., Stoffa, P.L., Sen, M.K. & Seifoullaev, R.K., 2000. Residual migration velocity analysis in the plane-wave domain, *SEG Tech. Prog., Expanded Abstracts,* **19,** 942–945.

Kennett, B.L.N., 1983. *Seismic Wave Propagation in Stratified Media,* Cambridge Univ. Press.

Kozlovskaya, E., Vecsey, L., Plomerova, J. & Raita, T., 2007. Joint inversion of multiple data types with the use of multiobjective optimization: problem formulation and application to the seismic anisotropy investigations, *Geophys. J. Int.,* **171**(2), 761–779.

Lindsay, R. & Koughnet, R.V., 2001. Sequential Backus Averaging: upscaling well logs to seismic wavelength, *Leading Edge,* **20,** 188–191.

Mallick, S., 1995. Model-based inversion of amplitude-variations-withoffset data using a genetic algorithm, *Geophysics,* **60,** 939–954.

Mallick, S., 1999. Some practical aspects of prestack waveform inversion using a genetic algorithm: an example from the east Texas Woodbine gas sand, *Geophysics,* **64,** 326–336.

Mallick, S., 2000. Prestack waveform inversion of multicomponent seismic data, *SEG Tech. Prog., Expanded Abstracts,* **19,** 2273–2276.

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

Mallick, S. & Frazer, L.N., 1988. Rapid computation of multioffset vertical seismic profile synthetic seismograms for layered media, *Geophysics,* **53,** 479–491.

Mallick, S. & Frazer, L.N., 1990. Computation of synthetic seismograms for stratified azimuthally anisotropic media, *J. geophys. Res.,* **95,** 8513–8526.

Mallick, S. & Frazer, L.N., 1991. Reflection/transmission coefficients and azimuthal anisotropy in marine seismic studies, *Geophys. J. Int.,* **105,** 241–252.

Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *J. Geophys. Res.*, **100(B7),** 12 431–12 447.

Mukherjee, D., Mallick, S., Shafer, L. & Campbell-Stone, E., 2012. Estimation of in-situ stress fields from P-wave seismic data, *SEG Tech. Prog., Expanded Abstracts,* doi:10.1190/segam2012-0375.1.

Mukhopadhyay, P.K. & Mallick, S., 2011. An accurate ray-based offset-to-angle transform from normal moveout uncorrected multicomponent data in a transversely isotropic medium with vertical symmetry axis, *Geophysics,* **76,** C41–C51.

Padhi, A. & Mallick, S., 2013a. Accurate estimation of density from the inversion of multicomponent prestack seismic waveform data using a nondominated sorting genetic algorithm, *Leading Edge,* **32,** 94–98.

Padhi, A. & Mallick, S., 2013b. Multicomponent prestack seismic waveform inversion in transversely isotropic media using a non-dominated sorting genetic algorithm, *Geophys. J. Int.,* **196,** 1600–1618.

Rudolph, G., 1993. Convergence of evolutionary algorithms in general search spaces, in *Proceedings of the Third IEEE Conference on Evolutionary Computation,* Piscataway, pp. 50–54.

Schoenberg, M. & Douma, J., 1988. Elastic-wave propagation in media with parallel fractures and aligned cracks, *Geophys. Prospect.,* **36,** 571–590.

Sen, M.K. & Roy, I.G., 2003. Computation of differential seismograms and iteration adaptive regularization in prestack waveform inversion, *Geophysics,* **68,** 2026–2039.

Sen, M.K. & Stoffa, P.L., 1992. Rapid sampling of model space using genetic algorithms: examples from seismic waveform inversions, *Geophys. J. Int.,* **108,** 281–292.

Sen, M.K. & Stoffa, P.L., 1996. Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion, *Geophys. Prospect.,* **44,** 313–350.

Stoffa, P.L. & Sen, M.K., 1991. Nonlinear multiparameter optimizations using genetic algorithms: inversion of plane-wave seismograms, *Geophysics,* **56,** 1794–1810.

Thomsen, L., 1986. Weak elastic anisotropy, *Geophysics,* **51,** 1954–1966.

Tsvankin, I., 1997. Anisotropic parameters and P-wave velocity for orthorhombic media, *Geophysics,* **62,** 1292–1309.

Virieux, J. & Operto, S., 2009. An overview of full waveform inversion in exploration geophysics, *Geophysics,* **74,** WCC1–WCC26.

Warner, M. *et al.*, 2013. Anisotropic 3D full-waveform inversion, *Geophysics,* **78,** R59–R80.

Woodhouse, J.H., 1974. Surface waves in a laterally varying layered structure, *Geophys. J. R. astr. Soc.,* **37,** 461–490.

APPENDIX A: PHYSICALLY MEANINGFUL THOMSEN–TSVANKIN PARAMETERS FOR VTI/HTI/ORTHORHOMBIC MEDIA

The physical limits for the Thomsen–Tsvankin parameters for VTI/HTI and orthorhombic medium can be readily obtained from the equations that convert them into elastic stiffness coefficients. Below, we outline only the VTI and orthorhombic case. Extending the VTI results to HTI is straightforward because HTI is simply a 90° rotation of the VTI medium about the y coordinate axis.

VTI case

The equations that convert the Thomsen parameters into the elastic coefficients are

$$C_{33} = \rho V_P^2, \tag{A1}$$

$$C_{44} = \rho V_S^2, \tag{A2}$$

$$C_{11} = C_{33}(1 + 2\varepsilon), \tag{A3}$$

$$C_{66} = C_{44}(1 + 2\gamma), \tag{A4}$$

and

$$C_{13} = [(C_{33} - C_{44})\{2\delta C_{33} + (C_{33} - C_{44})\}]^{\frac{1}{2}} - C_{44}. \tag{A5}$$

Note that C_{33}, C_{44}, C_{11} and C_{66} must be non-negative. If not, it will mean that the wave propagates with negative velocity along certain

directions which is physically impossible. In addition, while the *S*-wave velocity can be zero for the acoustic case, *P*-wave velocity is never zero. Additionally, density cannot be a negative quantity or zero. Using these arguments, eqs (A1) and (A2) gives

$$\rho > 0; \qquad V_P > 0; \qquad V_S \geq 0. \tag{A6}$$

Because C_{11} also cannot be zero or negative, from eq. (A3) we get

$$\varepsilon > -\frac{1}{2}. \tag{A7}$$

On similar arguments, eq. (A4) gives

$$\gamma \geq -\frac{1}{2}. \tag{A8}$$

Finally, from eq. (A5), we get

$$C_{33} > C_{44}, \tag{A9}$$

and

$$C_{33} - C_{44} \geq 2\delta C_{33}. \tag{A10}$$

Note that the inequality shown in eq. (A9) is based on the fact that the *S*-wave velocity must always be less than the *P*-wave velocity, not even equal.

All constraints shown by eqs (A6)–(A10) must be satisfied in choosing random models in any inversion method. Although the eqs (A6)–(A9) are quite obvious and is not likely to be encountered in choosing random models, eq. (A10) is important in constraining the negative δ values.

Orthorhombic case

For orthorhombic medium, the equations for computing the elastic stiffness coefficients from the Thomsen–Tsvankin parameters are

$$C_{33} = \rho V_P^2, \tag{A11}$$

$$C_{55} = \rho V_S^2, \tag{A12}$$

$$C_{22} = C_{33}\,(1 + 2\varepsilon_1)\,, \tag{A13}$$

$$C_{66} = C_{44}\,(1 + 2\gamma_1)\,, \tag{A14}$$

$$C_{11} = C_{33}\,(1 + 2\varepsilon_2)\,, \tag{A15}$$

$$C_{44} = C_{55}\,(1 + 2\gamma_2)\,, \tag{A16}$$

$$C_{23} = [(C_{33} - C_{44})\,\{2\delta_1 C_{33} + (C_{33} - C_{44})\}]^{\frac{1}{2}} - C_{44}, \tag{A17}$$

$$C_{13} = [(C_{33} - C_{55})\,\{2\delta_2 C_{33} + (C_{33} - C_{55})\}]^{\frac{1}{2}} - C_{55}, \tag{A18}$$

and

$$C_{12} = [(C_{11} - C_{66})\,\{2\delta_3 C_{11} + (C_{11} - C_{66})\}]^{\frac{1}{2}} - C_{66}. \tag{A19}$$

Based on eqs (A11)–(A19) and the arguments for the VTI case, the following are the constraints on the Thomsen–Tsvankin parameters for the models to be physically meaningful

$$\rho > 0; \qquad V_P > 0; \qquad V_S \geq 0, \tag{A20}$$

$$\varepsilon_1 > -\frac{1}{2}; \qquad \varepsilon_2 > -\frac{1}{2}; \tag{A21}$$

$$\gamma_1 \geq -\frac{1}{2}; \qquad \gamma_2 \geq -\frac{1}{2}, \tag{A22}$$

$$C_{33} > C_{44}; \qquad C_{33} > C_{55}; \qquad C_{11} > C_{66}; \tag{A23}$$

$$C_{33} - C_{44} \geq 2\delta_1 C_{33}, \tag{A24}$$

$$C_{33} - C_{55} \geq 2\delta_2 C_{33}, \tag{A25}$$

and

$$C_{11} - C_{66} \geq 2\delta_3 C_{11}. \tag{A26}$$

Again, the constraint on δ_1, δ_2 and δ_3 shown in eqs (A24)–(A26) are especially important.

APPENDIX B: KEY CONCEPTS OF MULTI-OBJECTIVE EVOLUTIONARY ALGORITHM

NSGA II, like all MOEAs, attempts to estimate the Pareto-optimal set of solutions using two fundamental concepts (1) non-dominance and (2) population diversity. Details of these concepts were discussed by Padhi & Mallick (2013b). Here, we review some of its fundamental concepts in the context of our parallel and computationally efficient implementation.

Dominance, non-dominance and Pareto-optimal front

In multi-objective optimization with a total number of m objectives, if a set of solutions $\mathbf{x}^1$ dominates over another set, say $\mathbf{x}^2$, it is then denoted as

$$\boldsymbol{x}^1 \succ \boldsymbol{x}^2. \tag{B1}$$

Considering that the multi-objective optimization is a minimization problem, $\boldsymbol{x}^1 \succ \boldsymbol{x}^2$ under the following conditions:

(1) If $\boldsymbol{x}^1 = [x_1^1, x_2^1, \ldots\ldots x_n^1]^T$ and $\boldsymbol{x}^2 = [x_1^2, x_2^2, \ldots\ldots x_n^2]^T$ in the decision space $\boldsymbol{X}$ maps respectively onto $\boldsymbol{y}^1 = [y_1^1, y_2^1, \ldots\ldots y_m^1]^T$ and $\boldsymbol{y}^2 = [y_1^2, y_2^2, \ldots\ldots y_m^2]^T$ in the objective space $\boldsymbol{Y}$ then

(i) $y_i^1 \leq y_i^2, \forall i = 1, 2, \ldots\ldots m.$
(ii) $y_j^1 < y_j^2$ for at least one value of j where $1 \leq j \leq m$.

In above, the symbol $\forall$ denotes 'for all'. The Pareto-optimal solutions in the decision space are those solutions that do not dominate one another. Due to their non-dominance, in the objective space, they form a front, known as the Pareto-optimal front. The first essential step in NSGA II is sorting the models in descending order of their respective dominance levels, or in other words, sorting them into different Pareto-optimal fronts.

Scaling

The concept of scaling has been discussed by Goldberg (1989) and has been used in the past for single component seismic inversions

(Stoffa & Sen 1991; Sen & Stoffa 1992; Mallick 1995, 1999). At early generations where most solutions are far from global optimum, ranking members and computing their crowding distances purely based on their raw objectives is improper because relatively better solutions in the population tend to drive algorithm towards a local optimum (Goldberg 1989). So, it is advisable to scale down the relatively better solutions and scale up the relatively poor ones during early generations and slowly relax this scaling process as the generation progresses and convergence to the global optimum is reached. In seismic inversion problems, two types of scaling of the objective function have been used (1) linear scaling and (2) exponential scaling.

In linear scaling, the poor solutions are scaled up and good solutions are scaled down using a linear function in such a way that the average value of the objective stays the same. For example, if f is the raw objective and f' is the corresponding scaled objective, the linear scaling computes the scaled objective as

$$f' = af + b. \tag{B2}$$

In eq. (B2), the constants a and b are chosen such that the average values of both raw and scaled objectives are equal, that is, $f'_{\rm avg} = f_{\rm avg}$ and the maximum value of the scaled objective is given as $f'_{\rm max} = C_{\rm m} f_{\rm avg}$, where $C_{\rm m}$ is a user-defined parameter. Using these conditions for the average and maximum values of the scaled fitness f', the constants a and b could be readily derived as

$$a = \frac{f_{\rm avg}\,(C_{\rm m} - 1)}{f_{\rm max} - f_{\rm avg}}, \tag{B3}$$

and

$$b = f_{\rm avg}\,(1 - a)\,. \tag{B4}$$

In our application, varying $C_{\rm m}$ from 1.2 at generation 1 to 2.0 for the last generation for each objective produced very good results.

In an exponential scaling, a parameter T which is similar to the temperature in simulated annealing (SA) is used to scale the objective values using an exponential function (Stoffa & Sen 1991). Similar to SA, T is initially chosen to be very high and slowly reduced over the generations. Details of this exponential scaling including how to vary T optimally over generations can be found in Stoffa & Sen (1991). Because we did not use exponential scaling in our applications, we do not repeat them here.

Rank, diversity, and crowding distance

In NSGA II, the best set of solutions out of the entire population is chosen such that they do not dominate one another but dominate the rest of the solutions, and are assigned a rank 1 (the first Pareto-optimal front). Another set of solutions are chosen such that they do not dominate one another, but are dominated by the members belonging to rank 1, and they dominate the others that are not yet ranked. This set is assigned rank 2 (the second Pareto-optimal front). This process of ranking in descending order of their dominance is continued until the entire population is exhausted. Ranking the population according to their dominance levels is one criterion to ensure that the best set of models is propagated over generations. Although it provides the mechanism for selecting the best members, it may still fail when most models do not dominate one another, that is, they all belong to the same rank or Pareto-optimal front. In theory, all stochastic searches including MOEA are based on an infinite population size. In practice however, the size of the population of the models must be finite, and any stochastic search using a finite size of population tend to cluster near a single solution over time. Such a clustering, known as the genetic drift (Goldberg 1989) is desirable when the solutions are near the global optimum, but they must be avoided at all other times to avoid premature convergence. For single objective problems, Sen & Stoffa (1992) proposed that running several independent GA inversions and combining them together can potentially overcome genetic drift.

Since for multi-objective problems running several GA inversions is expensive, the concept of 'population diversity', introduced by Deb *et al.* (2002) is a useful and practical alternative. Thus, in addition to sorting according to their dominance levels, to avoid clustering, the MOEAs assign a measure of population diversity to individual members. In standard MOEAs, the diversity is typically given as a normalized distance of a member from its neighbours, measured along each objective axis, and is called the 'crowing distance' (Deb *et al.* 2002). Out of two members belonging to the same rank, the one that is least crowded is given a higher chance for being selected than the other. This ensures that while choosing the good solutions, diversity in the population is also maintained.

While in standard MOEA, raw objectives are used in assigning ranks; in our implementations we use the scaled objectives for such. Additionally, we compute the crowding distance both in the scaled objective and model space and combine them together to define a new crowding distance for each model.