Regularized elastic full-waveform inversion using deep learning

GEOPHYSICS®

Regularized elastic full waveform inversion using deep learning

Journal:	*Geophysics*
Manuscript ID	GEO-2018-0685.R3
Manuscript Type:	Technical Paper
Keywords:	artificial intelligence, full-waveform inversion, elastic, log analysis
Area of Expertise:	Seismic Inversion, Seismic Velocities and Traveltimes

Regularized elastic full waveform inversion using deep learning

Zhen-dong Zhang[1] and Tariq Alkhalifah[1]

1 Department of Physical Science and Engineering,

King Abdullah University of Science and Technology,

Thuwal 23955-6900, Saudi Arabia.

Email: zhendong.zhang@kaust.edu.sa

(May 28, 2019)

Running head: **DPFWI**

ABSTRACT

Obtaining high resolution models of the Earth, especially around the reservoir, is crucial to properly imaging and interpreting the subsurface. We present a regularized elastic full waveform inversion method that uses facies as prior information. Deep neural networks are trained to estimate the distribution of facies in the subsurface. Here, we use facies extracted from wells as the prior information. Seismic data, well logs, and interpreted facies have different resolution and illumination to the subsurface. Besides, a physical process, such as anelasticity in the subsurface, is often too complicated to be fully considered. Therefore, there are often no explicit formulas to connect the data coming from different geophysical surveys. A deep learning method can find the statistically-correct connection without the need to know the complex physics. In our proposed deep learning scheme, we specifically use it to assist the inverse problem instead of the widely used labeling task. We first conduct an adaptive data-selection elastic full waveform inversion using the observed seismic data and obtain estimates of the subsurface, which do not need to be perfect. Then we use extracted facies information from the wells and force the estimated model to fit the facies by training deep neural networks. In this way, a list of facies is mapped to a 2D or 3D inverted model guided mainly by the structure features of the model. The multidimensional distribution of facies is used either as a regularization term or as an initial model for the next waveform inversion. The proposed method has two main features: 1) it applies to any kind of distributions of data samples and 2) it interpolates facies between wells guided by the structure of the estimated models. Results with synthetic and field data illustrate the benefits and limitations of this method.

Keywords: Deep learning, Facies, Elastic, Waveform inversion.

INTRODUCTION

Facies constrained seismic inversions, including amplitude versus offset (AVO) analysis and elastic full waveform inversion (FWI), have shown their capability in improving the resolution of estimates (Kemper and Gunning, 2014; Naeini and Exley, 2016; Zhang et al., 2017, 2018b). Seismic facies are groups of seismic properties and conformity layers that will have a certain relationship with geological and lithological properties. Such rock physics relations can be used as physical constraints in inversion. It is known that not all the medium parameters, such as anisotropy parameters, can be estimated from surface collected seismic data (Alkhalifah and Plessix, 2014; Zhang and Alkhalifah, 2017). Facies information extracted from geological analysis (e.g., sedimentology, stratigraphy and core analysis) or other geophysical prospecting methods (e.g., well logs) can complement seismic surveys (Asnaashari et al., 2013). Not only improving resolutions of multiple parameters, but facies constraints can also help to avoid cycle-skipping faced by conventional FWI. Combined with the facies constraints and an improved objective function, we can somehow reduce the risks of converging into local minima in the elastic FWI.

Recently, lots of efforts have been made to find better objective functions that are immune from cycle skipping (Liu et al., 2018; Zhang et al., 2018a; Wu et al., 2019; Yi et al., 2019). The conventional wiggle-to-wiggle subtraction based measurement fails FWI when the predicted and observed data exceed the half-cycle limit (Virieux and Operto, 2009). Choi and Alkhalifah (2012) proposed a normalized global crosscorrelation based FWI, which reduces the dependency on seismic amplitudes. It is more sensitive to phase differences in the data, and thus, it is more immune to ambient noise in the field data (Chi et al., 2015). However, the global crosscorrelation objective function still suffers from high nonlinearity and the danger of converging to a local minimum when the initial model is far from the actual one. One intuitive remedy to the problem is to select parts of the data free from cycle-skipping in the inversion, which is referred to as multiscale inversion (Bunks et al., 1995; Martínez-Sansigre and Ratcliffe, 2014; Bi and Lin, 2014; Zhang and Alkhalifah, 2018). A selection of frequencies from low to high is a widely used strategy in multiscale inversion. However, it is not applicable to the data lacking low frequencies. An alternative choice of multiscale inversion is selecting data corresponding to offsets, which are free of cycle-skipping. The crucial step of these approaches is the scheme used in selecting data for inversion in each iteration. Martínez-Sansigre and Ratcliffe (2014) designed a probabilistic quality control to quantify the cycle-skipping. Bi and Lin (2014) used the traveltime difference as a criterion to select the proper data. Although such approaches show promising inversion results, the L_2 norm objective function is inconsistent with such inversion strategies since increasing the selection range and the reduction in the data difference are competing with each other, and thus, can not be handled using current optimization schemes. In our proposed approach, we use local similarity proposed by Fomel (2007) to measure the differences between the predicted and the observed data and utilize the calculated similarity to adaptively select the data to be included in the inversion. Thus, data that comply with the cycle-skipping criterion for the updated model is automatically incorporated into the inversion in each iteration (Zhang and Alkhalifah, 2018). We use a global crosscorrelation to maximize the similarity of the predicted and observed data. Both the selection range and the similarity should increase if the model is updated in the correct direction, and thus, the problem can be solved by existing optimization algorithms without modifications.

With facies information and reasonable estimates of the Earth model in hand, we need to find a way to connect them. Since physical processes that connect them in the real Earth are too complicated to be fully formulated, the explicit equations that can map one of them to the other are always based on some assumptions. Deep neural networks (DNNs) through training can develop the proper statistical connection that turns estimates from seismic data to facies from well logs or vice versa. Deep learning or its broader family, artificial intelligence (AI), has been mainly used as labeling tools for solving geophysical interpretation problems in the past few years (Van der Baan and Jutten, 2000; Araya-Polo et al., 2017; AlRegib et al., 2018; Di et al., 2018; Guitton, 2018). Other applications such as low-frequency components reconstruction (Ovcharenko et al., 2018) and waveform tomography (Araya-Polo et al., 2018) have also shown promising results of machine learning in solving geophysical problems. DNNs have also been used for facies clustering (West et al., 2002), which is more related to our proposed method. A significant difference between our proposed method and previous applications is that we adopt the deep learning scheme to solve an inverse problem. Instead of labeling particular facies, we register the probabilities for all the existing facies in the area. Those probabilities are used as weighting factors for a weighted summation over all the facies to obtain the distribution of facies (converted to v_p and v_s or other physical parameters) in the subsurface. This approach can avoid an inherent bias imposed by particular facies especially those corresponding to wild estimates at the beginning of FWI. The estimated distribution of facies can be utilized as a regularization term in elastic FWI (Zhang et al., 2018b).

This paper is divided into five sections. After the introduction, we share a novel objective function and train deep neural networks for facies-distribution estimation. In the examples, we first test the approach on the synthetic Marmousi model, then apply the method to the BigSky land field data to analyze the effectiveness and limitations of our method. Finally, we discuss the utilization of multiple geophysical data and deep neural networks to mitigate some of the problems we are facing in full waveform inversion.

THEORY

We start this section by developing the adaptive data-selection approach for elastic waveform inversion. We then describe the architecture of the deep neural networks used to utilize the facies information in the FWI.

Adaptive data-selection elastic FWI

Due to the oscillatory nature of seismic signals, the L_2 norm objective function suffers from cycle-skipping, when the mismatch between the predicted and observed data exceed a half-cycle. An intuitive remedy to this problem is to select parts of the data free of cycle-skipping, and the process can be done adaptively. The proposed objective function is written as

$$J(\mathbf{m}) = -\sum_{s}\sum_{r} \mathbf{A}\widehat{\mathbf{u}} \cdot \widehat{\mathbf{d}}, \tag{1}$$

where $\widehat{\mathbf{u}} = \frac{\mathbf{u}}{||\mathbf{u}||}$ and $\widehat{\mathbf{d}} = \frac{\mathbf{d}}{||\mathbf{d}||}$ are normalized predicted and observed data, respectively. The indexes s and r correspond to the source and receiver locations, respectively, and $\mathbf{A}$ is the selection matrix, which will be explained later. The objective here is to maximize the similarity of the predicted and the observed data.

The inverse problem is constrained by the elastic wave equation given by

$$\begin{pmatrix} \rho\mathbf{I}_3 & 0 \\ 0 & \mathbf{C}^{-1} \end{pmatrix} \frac{\partial \Psi}{\partial t} - \begin{pmatrix} 0 & E^T \\ E & 0 \end{pmatrix} \Psi - \mathbf{s} = 0, \tag{2}$$

where $\Psi = (v_1, v_2, v_3, \sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6)$ is the vector containing three particle velocities and six stresses, E denotes spatial-differentiation operators, $\mathbf{C}$ represents the stiffness matrix and $\mathbf{s}$ denotes the point source used for modeling.

As discussed in the introduction, not all the data can be predicted with accurate kinematic information in practice. The conventional L_2 norm based FWI requires that the predicted and observed data have a maximum difference of half wavelength for each event. Otherwise, the adjoint source would be cycle-skipped. In our proposed method, we do not need to calculate the mismatch of each event because it is impractical to isolate each event. Instead, we use the local similarity measure to generate a weighting matrix, which emphasizes the most coherent events in the predicted and observed data. As we update the model, more of the data will fit the cycle-skipping criterion.

The local similarity proposed by Fomel (2007) was initially applied to compare the similarity of two images (i.e., the PP and PS image registration). Different from the global correlation, the local analysis produce a local correlation as a variable function, that identifies local changes in the similarity of two signals. Each element of the similarity matrix is given by

$$a_s(t, r) = c_1 * c_2, \tag{3}$$

where t and r are indexes of time and receivers, respectively. $c_1 = \frac{\mathbf{M}\mathbf{U}^T\mathbf{d}}{\lambda^2\mathbf{I}+\mathbf{M}(\mathbf{U}^T\mathbf{U}-\lambda^2\mathbf{I})}$ and $c_2 = \frac{\mathbf{M}\mathbf{D}^T\mathbf{u}}{\lambda^2\mathbf{I}+\mathbf{M}(\mathbf{D}^T\mathbf{D}-\lambda^2\mathbf{I})}$. $\mathbf{U}$ and $\mathbf{D}$ are diagonal matrices composed from the elements of $\mathbf{d}$ and $\mathbf{u}$, respectively. $\mathbf{M}$ denotes a smooth filter. The dividing operation is achieved by solving two least-square inverses.

The original definition of local similarity is not suitable for the problem because it ignores the polarity information of the two signals. For example, two signals with opposite polarity (fully cycle-skipped) have a similarity of 1 in this case. To overcome a potential fully cycle-skipping, we add a polarity detection in the selection matrix which is given by

$$a_p(t,r) = \begin{cases} 0, & sign(u) * sign(d) < 0 \\ 1, & otherwise \end{cases}, \tag{4}$$

and each element of the selection matrix, $\mathbf{A}$, is given by $a(t,r) = a_s(t,r) * a_p(t,r)$.

To obtain the gradient function of the proposed objective function, we take its derivative with respect to the model parameters as follows

$$\frac{\partial J}{\partial \mathbf{m}} = \sum_s \sum_r \frac{\partial \mathbf{u}}{\partial \mathbf{m}} \cdot \left(\frac{\mathbf{A}}{||\mathbf{u}||} \left(\widehat{\mathbf{u}} \left(\widehat{\mathbf{u}} \cdot \widehat{\mathbf{d}} \right) - \widehat{\mathbf{d}} \right) + \frac{\partial \mathbf{A}}{\partial \mathbf{u}} \right), \tag{5}$$

where $\frac{\partial \mathbf{A}}{\partial \mathbf{u}}$ can be approximated by $\frac{2.0\mathbf{A}}{\mathbf{u}}$.

The inverse problem here is nonlinear since the selection matrix is updated in each iteration. We cannot use the L_2 norm based measurements in the inversion because the norm of $\mathbf{A}$ is increasing as we update the model, and the data mismatch is reduced if the updating is in the right direction. In the proposed objective function (equation 1), the norms of $\mathbf{A}$ and $\widehat{\mathbf{u}} \cdot \widehat{\mathbf{d}}$ are both increasing as we update the model, and thus, allows for a consistent procedure treatable by current optimization methods.

The model is updated iteratively using the L-BFGS method (Liu and Nocedal, 1989; Wu et al., 2015), which is given by

$$\mathbf{m} = \mathbf{m}_0 - \lambda \mathbf{H}^{-1} \mathbf{g}, \tag{6}$$

where λ is the step length calculated by the line-search method, which satisfies the Wolfe condition (Wolfe, 1969). $\mathbf{H}$ is the approximated Hessian matrix. $\mathbf{g}$ is the calculated gradient.

The idea of the proposed inversion strategy is straightforward: choose the data free of cycle-skipping for inversion. The selection criteria are based on the local-similarity of two traces. In most of the cases, the initial model can regenerate cycle-skipping free data in the near-offsets. Our proposed inversion algorithm can initially utilize such data and gradually include more data for inversion. However, in the extreme case when the predicted data is far from the observed one, our proposed method can fail to update the initial model.

Architecture of deep neural networks

The existing waveform inversion strategies can always face the risk of converging to one of the local minima. Fitting the surface collected seismic data cannot guarantee that the estimation is the true solution since the observed data are effectively fewer than unknowns, thus resulting in a Null space. It has been shown that regularization can be helpful in constraining the inverse problem (Guitton, 2012; Asnaashari et al., 2013; Zhang et al., 2018b). Here, we use the data obtained from other geophysical surveys such as well logs as the prior information in our proposed inversion. However, currently used explicit equations that can connect the different data are based on strong approximations to the subsurface. Recently blooming data science tells us that statistical principles behind large data samples can be effective tools in merging such different information. Deep neural networks are trained here to seek such principles using inverted velocities from seismic data, and facies extracted from wells.

A deep neural network is nothing but a nonlinear system of equations that turns the input into the output (Van der Baan and Jutten, 2000). It has multiple hidden layers between the input and output layers. With the input layer denoted as $\mathbf{x}$, the kth hidden layer can be expressed as $\mathbf{a}_k = \phi_k\{\mathbf{W}_k(\ldots \phi_1[\mathbf{W}_1\mathbf{x} + \mathbf{b}_1]) + \mathbf{b}_k\}$, and the output layer is written as $\mathbf{y} = \mathbf{W}\mathbf{a} + \mathbf{b}$. The input, $\mathbf{x}$, can be raw data or features (e.g., v_s/v_p) extracted from the data. The output, $\mathbf{y}$, depends on the problem. For example, it can be 0 or 1 for labeling applications. The forward-propagation process utilizes the output of the previous layer as the input for the next layer. ϕ denotes the activation function which defines the output of that node with fed input. It can be the sigmoid, rectified linear unit (ReLu) and some other functions. The training process updates $\mathbf{W}$ and $\mathbf{b}$ for each layer to seek a more accurate mathematical manipulation capable of mapping the input to the output using a loss function of sparse softmax cross entropy (Glorot et al., 2011). We use three features, v_p, v_s and v_s/v_p, as inputs. Four hidden layers with 64 nodes in each layer are deployed as shown in Figure 1. A ReLu activation function is used (Nair and Hinton, 2010). For each layer, we use a random dropout of 10% to avoid overfitting (Srivastava et al., 2014). Besides, a random data augmentation technique is applied to balance the proportion of different facies in training the data (Krizhevsky et al., 2012). The Adam gradient is used to update the weighting matrix of neural networks. In our application, we output the probabilities for all facies instead of one specific kind. After obtaining the percentages of being a certain facies, we can calculate the distribution of facies (converted to v_p and v_s) by a weighted summation over n_f facies, $\bar{v} = \sum_{i=1}^{n_f} p_i v_i$. $\bar{v}$ denotes averaged P- or S-wave velocity, which is equivalent to the posterior expectation in Zhang et al. (2018b). p_i and v_i are probabilities estimated by the trained DNNs and the known facies. Such a weighted summation avoids potential biasing by a particular kind of facies when the DNNs fail. Besides, it can interpolate between different facies. In practice, we can never know all the facies in the subsurface and we do not need to know all of them in our proposed method. The probabilities act as interpolation weights for the known facies. If the corresponding facies for certain pairs of v_p and v_s is not available as prior knowledge, the converted v_p and v_s still have a chance of being (or close to) the correct ones through interpolation. The converted v_p and v_s can vary more continuously than the discrete facies. The variance of the estimated probabilities (unsupervised) provides an approach to evaluate the inverted velocity models. A small variance indicates that the trained neural networks fail to classify the input v_p and v_s pair to a specific facies with high confidence. The inverted values, in this case, might not be as good as those with high confidence (large variance). However, this evaluation heavily depends on the assumption that the prior information (facies) is reliable.

The proposed inversion algorithm is illustrated by Figure 2 and summarized in the following steps:

1. Conduct the proposed elastic FWI (equation 1) using the shot gathers.

2. Extract facies from well logs or other sources.

3. Select several vertical profiles near the well from the estimated model (from step 1) and build the connections between these estimates and the interpreted facies (from step 2) by training neural networks.

4. Use the trained neural networks (from step 3) to predict the distribution of certain facies on the whole model (from step 1), and then use a weighted summation ($\bar{v} = \sum_{i=1}^{n_f} p_i v_i$) to generate v_p and v_s models, denoted as m^c.

5. Use the converted velocities (from step 4) as the initial model (for high-quality seismic data, e.g., synthetic data) or a regularization term ($m^{inv} - m^c$, if the extracted facies are more reliable) for another cycle of elastic FWI.

6. Repeat steps 1, 3, 4, 5 if there are apparent classification errors in the estimated distribution of facies.

NUMERICAL EXAMPLES

We first use a modified elastic Marmousi model to verify the effectiveness and robustness of the proposed algorithm. We follow that by an application on Land data.

Synthetic Marmousi model

The actual S-wave velocity is generated for this Marmousi model by setting $v_s = v_p/\sqrt{3} + 0.1(v_p - 2.4)$. The actual and initial velocities are shown in Figures 3. Initial models are 1D linear gradient models, which are far from the actual ones ($v_s = v_p/\sqrt{3}$). 220 sources and 330 receivers are evenly deployed on the surface of the model and the recorded data are two-component particle velocities. The maximum offset is 6.6 km. A staggered finite-difference scheme is implemented to solve the elastic wave equation (Virieux, 1986). The source wavelet is a Ricker wavelet ($f_p = 5$ Hz) without frequencies below 5 Hz, in which case our proposed inversion approach fails to converge to the global minimum without prior information as shown in Figure 4. We extract ten facies from pseudo wells at $x = 1\,km, 3\,km, 5\,km$ as shown in Table 1. There is no need to extract all the existing facies from the well and these ten facies are the dominant ones. The weighted summation using probabilities as weights can interpolate between the facies when converted to v_p and v_s. Then we use the estimated v_p, v_s (as shown in Figure 4) and their ratio, v_s/v_p, at the same location as data features. The interpreted facies (Table 1) from the pseudo wells are labels of the training data set. After the DNNs are well trained, the full dimension of inverted velocities and their ratios are used as input data to generate a possible distribution of facies. Figure 5 shows the normalized data loss versus iteration at every 100 steps. A total of 70% data loss for the training data set and a 55.6% test accuracy are achieved. K-fold cross-validation can be used to aid the design of neural networks (Kohavi et al., 1995). We did not apply the K-fold cross-validation in this example since the neural networks used can generate acceptable initial models for elastic FWI. Although the test accuracy is relatively low, the converted parameters still have a chance to be close to the actual values thanks to the weighted summation. Besides, the following elastic FWI can improve the accuracy by matching the observed seismic data. The distribution of facies is converted to v_p using a weighted summation ($\bar{v} = \sum_{i=1}^{n_f} p_i v_i$, v_i are given in Table 1) as shown in Figure 6a. It has a similar structure as the actual v_p but with some loss in detail. The largest probabilities of falling into one particular facies for the whole model are shown in Figure 6b. The large values in the shallow area indicate that the trained DNNs can classify the inverted velocities to a particular facies with high confidence (one large p_i and the rest are smaller ones). However, the smaller values in the deep part indicate that the trained DNNs are slightly puzzled in the classification and they give similar probabilities to nearby facies (a list of small p_i). The variances of the probabilities ($var = mean(abs(input - input.mean)^2)$) as shown in Figure 6c indicate a similar conclusion. In this case, the variance can provide an indicator of the uncertainties in the inverted velocities from this elastic FWI. In the definition above, a large variance indicates that the estimation matches the known facies well while a small one indicates a mismatch to a particular facies (i.e., we cannot pick a model from the output probabilities). It is also possible that the classification is biased by a particular facies, and thus, have a big variance. However, from the data loss (Figure 5) and the converted v_p (Figure 6a), this does not happen in this example. A smoothed version of the estimated distribution of facies (e.g., Figure 6a) shown in Figures 7a and 7b is used as the initial model for a L_2 norm based elastic FWI. The final inverted velocities after adding prior information are shown in Figure 7c and 7d. The inverted model is close to the actual one except for the areas near boundaries. For a better comparison, we also compare profiles of velocities of actual, initial, inverted without regularizations and inverted with regularizations in Figure 8. Estimated S-wave velocities without constraints are trapped in one of the local minima as the arrow indicates. Data comparison in Figure 9 indicate a similar conclusion. The predicted data using the proposed algorithm (Figure 9d) is much closer to the observed one (Figure 9a) than the one without regularization (Figure 9c). A similarity measurement ($\frac{\mathbf{u}\cdot\mathbf{d}}{\sqrt{\mathbf{u}\cdot\mathbf{u}}\sqrt{\mathbf{d}\cdot\mathbf{d}}}$) is shown in Figure 10. The measured value should be equal to 1 when the predicted data is the same as the observed one. It shows that the adaptive data-selection objective function alone fails in the far-offset and the proposed approach that utilizes facies is able to match the observed data in the far-offset.

BigSky field data

The field data used to verify the effectiveness of our proposed method come from the BigSky Carbon Sequestration Partnership (BSCSP), which is a US CO_2 storage project. The land data were collected using a 3D multi-component seismic survey (3D-9C) with a minimum frequency of 15 Hz, which is a challenge for conventional FWI. The 3D survey geometry is plotted in Figure 11, and a 2D inline across the central area is selected for the test to allow us to practically use high frequencies. There is a well at the edge of the survey area, which we use to extract the initial velocities by smoothing, and use them as references for the inverted velocities. The raw data set corresponding to the vertical component is plotted in Figure 12a. Following the synthetic example, we first manually interpret the existing facies from the well logs as shown in Figure 12b. There are 11 facies marked in the well logging profile and these interpreted facies are used as labels for training the DNNs. Initial P- and S-wave velocities calculated from Backus averaging as shown in Figures 13a and 14a are used as initials for elastic FWI. Estimated P- and S-wave velocities without facies constraints are shown in Figures 13b and 14b, respectively. A strong lateral smoothing filter is applied to the gradient because of the overall weak lateral variation in this area. Besides, a total variation (TV) regularization is also applied to the estimated model so as to add high-wavenumber components to the estimates (Guitton, 2012; Alkhalifah et al., 2018). After training the neural network and applying facies constraints, we conduct elastic FWI again and obtain inverted P- and S-wave velocities as shown in Figures 13c and 14c, respectively. Lateral variations indicated by the arrows demonstrate that the proposed regularization is not necessarily 1D and it produces lateral variations guided by the structures of estimates. Figure 15 shows the data comparison. From the observed data and well logs, we know there exists high-contrast velocity layers. However, the strong reflections are hidden in the noisy data (Figure 15a), and thus, cannot be recovered by fitting the seismic data (Figure 15c). Facies extracted from wells provide complimentary illumination as shown in Figure 15d. The strong reflections in the predicted data using the model estimated from the proposed method can generally match those in the observed data. The vertical profile comparison in Figure 16 shows that inverted P- and S-wave velocities are close to the well logs. We assign a large weight to the regularization term since we think that the well logs are more reliable as compared to the inverted models from matching the land seismic data.

DISCUSSIONS

Different geophysical surveys have their advantages in imaging the subsurface. For example, seismic surveys have a larger illumination area than well logs, but with lower resolution. Often, not all the parameters, such as when using the anisotropy assumption, are resolvable from surface collected seismic data. However, well logs can provide such anisotropy parameters corresponding usually for only a limited area around the well. Full waveform inversion in most cases aims to fit the observed seismic data. Despite its elegant theorem, FWI faces many problems in practice. The real Earth has too complex physics to be fully represented by numerical simulations, and thus, a wiggle-to-wiggle matching is impractical especially for land data. In this case, perfectly fitting the land seismic data results in an overfitted estimates which are often incorrect. Incorporating well logs as regularization to the inverse problem can add physical constraints and hopefully reduce the risk of overfitting (Asnaashari et al., 2013). The utilization of predicted distribution of facies is case dependent. In the synthetic example, we use the predicted model from DNNs as initial models and improve them by matching the noise-free seismic data. In our land data example, we think that the well log is more reliable than our estimates from surface seismic data at least for the region near the well. Thus, we also use the predicted model from DNNs as a strong regularization to remove the negative influence of matching the noisy part of the seismic data. Meanwhile, our seismic inversion can produce generally coherent structures (Shen et al., 2018). With detailed 1D velocities from the well and a reliable trend from FWI, we can image the 2D or 3D subsurface better. A robust FWI algorithm is needed to generate such general coherent structures. Our proposed adaptive data-selection objective function can reduce the local minima by gradually including the data in inversion, which is also considered as a multiscale approach. However, the proposed objective function is not fully cycle-skipping free as many other methods do. It fails when the predicted and observed data are quite different from each other. Thus, utilizing other geophysical surveys is another approach to stabilize FWI along with modifying the objective function.

There are often no effective procedure to connect the models obtained from different geophysical surveys. Physical processes in the real Earth are too complicated to be described by a set of equations. In practice, either statistical principles or simplified physics are used in solving geophysical problems. Deep learning techniques can learn from field data samples and thus avoid the need to define explicit formulas, which should be helpful to geophysical applications. There are many deep learning frameworks available and one of them is TensorFlow developed by Google (Abadi et al., 2016). These open-source frameworks are well developed and the users are only responsible for preparing the data. In geophysical applications, we might not have enough data samples for an effective learning. In this case, artificial data augmentation is needed. Also, data overfitting might be an issue for training the networks and a random dropout or random data augmentation can solve this problem (Krizhevsky et al., 2012; Srivastava et al., 2014). In conclusion, problems faced in geophysical applications can have solutions in other methodologies such as natural language processing and object detection. Our proposed algorithm utilizes seismic data and other geophysical data (i.e., well logs) in velocity estimation. The training of our examples can be done within five minutes using one GPU card (Tesla K40). The examples indicate that the needed number of wells (or the number of facies as prior) depend on the heterogeneity of the area. However, the limited number of wells should not be problematic since the proposed method also utilizes the seismic data. The known facies information is not limited to v_p and v_s, it can include other information such as anisotropy parameters. Thus, the proposed method also has the potential of resolving more parameters such as the anisotropy parameters, which we are currently investigating.

CONCLUSIONS

Elastic FWI with facies constraints can mitigate the cycle-skipping caused by bad initial models. Facies are usually interpreted from different geophysical observations such as well logs. The proposed inversion algorithm aims to fit not only seismic data but also the well logs. The well logs are not used as a direct-constraint in inversion. Delicately designed deep neural networks are trained to find the correct mathematical manipulation that can turn the estimated velocities from seismic data into the measured velocities from the well logs. Training the networks is a data-driven inverse process, and thus, avoids considering complex physical processes. Although the well logs have limited lateral illumination, the trained DNNs can map them to 2D or 3D models with a structure-guided interpolation. The estimated distribution of facies can be used as a physical constraint for conventional elastic FWI. It can be imperfect at the beginning and can be updated iteratively in elastic FWI. In our synthetic example, both seismic data and well logs are reliable, and thus, the final model can fit both of them. In the field land data, considering the difficulties we often face with land FWI, we assume the well logs are more reliable and therefore assign a larger weight to the regularization term. Both examples verify the effectiveness of the proposed inversion algorithm. One weakness of the problem is that a successful training needs many data samples, which might not be available. However, with data augmentation technique, artificially created data samples can help to train the neural networks.

ACKNOWLEDGMENTS

We thank Yike Liu, Zedong Wu and Yunseok Choi for their helpful discussions. We want to thank Erika Gasperikova, Antoine Guitton and four anonymous reviewers, for the effort put into the review of this paper. We thank KAUST for its support and specifically the seismic wave analysis group members for their valuable insights. For computer time, this research used the resources of the Supercomputing Laboratory at King Abdullah University of Science & Technology (KAUST) in Thuwal, Saudi Arabia. We thank Vecta Oil and Gas and especially Brian Devault for the BigSky data and the helpful discussions.

REFERENCES

Abadi, M., P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al., 2016, Tensorflow: a system for large-scale machine learning.: OSDI, 265–283.

Alkhalifah, T., and R.-É. Plessix, 2014, A recipe for practical full-waveform inversion in anisotropic media: An analytical parameter resolution study: Geophysics, **79**, R91–R101.

Alkhalifah, T., B. B. Sun, and Z. Wu, 2018, Full model wavenumber inversion: Identifying sources of information for the elusive middle model wavenumbers: Geophysics, **83**, R597–R610.

AlRegib, G., M. Deriche, Z. Long, H. Di, Z. Wang, Y. Alaudah, M. A. Shafiq, and M. Alfarraj, 2018, Subsurface structure analysis using computational interpretation and learning: A visual signal processing perspective: IEEE Signal Processing Magazine, **35**, 82–98.

Araya-Polo, M., T. Dahlke, C. Frogner, C. Zhang, T. Poggio, and D. Hohl, 2017, Automated fault detection without seismic processing: The Leading Edge, **36**, 208–214.

Araya-Polo, M., J. Jennings, A. Adler, and T. Dahlke, 2018, Deep-learning tomography: The Leading Edge, **37**, 58–66.

Asnaashari, A., R. Brossier, S. Garambois, F. Audebert, P. Thore, and J. Virieux, 2013, Regularized seismic full waveform inversion with prior model information: Geophysics, **78**, R25–R36.

Bi, H., and T. Lin, 2014, Impact of adaptive data selection on full waveform inversion, *in* SEG Technical Program Expanded Abstracts 2014: Society of Exploration Geophysicists, 1094–1098.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Chi, B., L. Dong, and Y. Liu, 2015, Correlation-based reflection full-waveform inversion: Geophysics, **80**, R189–R202.

Choi, Y., and T. Alkhalifah, 2012, Application of multi-source waveform inversion to marine streamer data using the global correlation norm: Geophysical Prospecting, **60**, 748–758.

Di, H., Z. Wang, and G. AlRegib, 2018, Seismic fault detection from post-stack amplitude by convolutional neural networks: Presented at the 80th EAGE Conference and Exhibition 2018.

Fomel, S., 2007, Local seismic attributes: Geophysics, **72**, A29–A33.

Glorot, X., A. Bordes, and Y. Bengio, 2011, Deep sparse rectifier neural networks: Proceedings of the fourteenth international conference on artificial intelligence and statistics, 315–323.

Guitton, A., 2012, Blocky regularization schemes for full-waveform inversion: Geophysical Prospecting, **60**, 870–884.

———, 2018, 3D convolutional neural networks for fault interpretation: Presented at the 80th EAGE Conference and Exhibition 2018.

Kemper, M., and J. Gunning, 2014, Joint impedance and facies inversion–seismic inversion redefined: First Break, **32**, 89–95.

Kohavi, R., et al., 1995, A study of cross-validation and bootstrap for accuracy estimation and model selection: Ijcai, Montreal, Canada, 1137–1145.

Krizhevsky, A., I. Sutskever, and G. E. Hinton, 2012, Imagenet classification with deep convolutional neural networks: Advances in neural information processing systems, 1097–1105.

Liu, D. C., and J. Nocedal, 1989, On the limited memory BFGS method for large scale optimization: Mathematical programming, **45**, 503–528.

Liu, Y., B. He, H. Lu, Z. Zhang, X.-B. Xie, and Y. Zheng, 2018, Full-intensity waveform inversion: Geophysics, **83**, R649–R658.

Martínez-Sansigre, A., and A. Ratcliffe, 2014, A probabilistic QC for cycle-skipping in full waveform inversion, *in* SEG Technical Program Expanded Abstracts 2014: Society of Exploration Geophysicists, 1105–1109.

Naeini, E. Z., and R. Exley, 2016, Quantitative interpretation using facies-based seismic inversion, *in* SEG Technical Program Expanded Abstracts 2016: Society of Exploration Geophysicists, 2906–2910.

Nair, V., and G. E. Hinton, 2010, Rectified linear units improve restricted boltzmann machines: Proceedings of the 27th international conference on machine learning (ICML-10), 807–814.

Ovcharenko, O., V. Kazei, D. Peter, X. Zhang, and T. Alkhalifah, 2018, Low-frequency data extrapolation using a feed-forward ANN: Presented at the 80th EAGE Conference and Exhibition 2018.

Shen, X., L. Jiang, J. Dellinger, A. Brenders, C. Kumar, M. James, J. Etgen, D. Meaux, R. Walters, and N. Abdullayev, 2018, High resolution full waveform inversion for structural imaging in exploration, *in* SEG Technical Program Expanded Abstracts 2018: Society of Exploration Geophysicists, 1098–1102.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, 2014, Dropout: a simple way to prevent neural networks from overfitting: The Journal of Machine Learning Research, **15**, 1929–1958.

Van der Baan, M., and C. Jutten, 2000, Neural networks in geophysical applications: Geophysics, **65**, 1032–1047.

Virieux, J., 1986, P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method: Geophysics, **51**, 889–901.

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

West, B. P., S. R. May, J. E. Eastwood, and C. Rossen, 2002, Interactive seismic facies classification using textural attributes and neural networks: The Leading Edge, **21**, 1042–1049.

Wolfe, P., 1969, Convergence conditions for ascent methods: SIAM review, **11**, 226–235.

Wu, S., Y. Wang, Y. Zheng, and X. Chang, 2015, Limited-memory BFGS based least-squares pre-stack kirchhoff depth migration: Geophysical Journal International, **202**, 738–747.

Wu, Z., T. Alkhalifah, Z. Zhang, F. Alonaizi, and M. Almalki, 2019, A new full waveform inversion method based on shifted correlation of the envelope and its implementation based on opencl: Computers & Geosciences, **129**, 1–11.

Yi, J., Y. Liu, Z. Yang, H. Lu, B. He, and Z. Zhang, 2019, A least-squares correlation-based full traveltime inversion for shallow subsurface velocity reconstruction: Geophysics, **84**, 1–54.

Zhang, Z., T. Alkhalifah, Z. Wu, Y. Liu, B. He, and J. Oh, 2018a, Normalized nonzero-lag crosscorrelation elastic full-waveform inversion: Geophysics, **84**, R15–R24.

Zhang, Z., E. Z. Naeini, and T. Alkhalifah, 2017, Facies constrained elastic full waveform inversion: Presented at the 79th EAGE Conference and Exhibition 2017.

Zhang, Z.-D., and T. Alkhalifah, 2017, Full waveform inversion using oriented time-domain imaging method for vertical transverse isotropic media: Geophysical Prospecting, **65**, 166–180.

———, 2018, Adaptive data-selection elastic full-waveform inversion, *in* SEG Technical Program Expanded Abstracts 2018: Society of Exploration Geophysicists, 5163–5167.

Zhang, Z.-d., T. Alkhalifah, E. Z. Naeini, and B. Sun, 2018b, Multiparameter elastic full waveform inversion with facies-based constraints: Geophysical Journal International, **213**, 2112–2127.

LIST OF TABLES

LIST OF FIGURES

8 Vertical profiles across the low-velocity zones. The inverted velocities are far from the actual ones without using facies as constraints and v_s suffers from a severe cycle-skipping problem as the pink arrow indicates in d). Facies constraints can eliminate artifacts caused by cycle skips.

9 Data comparison (vertical component). a) Observed, b) initial, c) inverted without facies constraints and d) inverted with facies constraints.

10 Correlation of the predicted and observed data ($\frac{\mathbf{u}\cdot\mathbf{d}}{\sqrt{\mathbf{u}\cdot\mathbf{u}}\sqrt{\mathbf{d}\cdot\mathbf{d}}}$). The initial model cannot provide accurate prediction in the far-offsets. The adaptive-selection objective function fails when the predicted and observed data are far from each other. The inverted model of the proposed approach can provide accurate prediction at the far-offsets.

11 3D survey geometry. We choose one 2D line for inversion as indicated by the arrow. There is a well at the edge of the survey area.

12 a) An example from the raw data set (vertical component) and b) extracted facies from the P-wave well log. The shot gather is noisy and lacks low frequencies, which is challenging for FWI. Line segments in different colors indicate the interpreted facies.

13 P-wave velocities. a) Initial v_p from the well logs, b) estimated v_p without facies constraints and c) estimated v_p with facies constraints. Arrows point to lateral variations in the estimates.

14 S-wave velocities. a) Initial v_s from the well logs, b) estimated v_s without facies constraints, and c) estimated v_s with facies constraints.

15 Data comparison (vertical component). a) Observed, b) initial, c) inverted without facies constraints and d) inverted with facies constraints. The proposed facies constraints

can predict some high-contrast layers and generate strong reflections hidden in the noisy seismic data. The arrows point to some of these reflections. The shot gathers are plotted at the same scale so that a direct comparison can be made.

16 Vertical profiles of the smoothed well log, initial model, inverted model without facies constraints and inverted model with facies constraints. a) v_p and b) v_s.

Table 1: Ten facies in the model.

Facies number	1	2	3	4	5	6	7	8	9	10
P-wave velocity (km/s)	1.5	1.7	2.2	2.5	2.65	3.2	3.5	3.8	4.0	4.5
S-wave velocity (km/s)	0.78	0.92	1.24	1.45	1.55	1.94	2.15	2.32	2.46	2.78

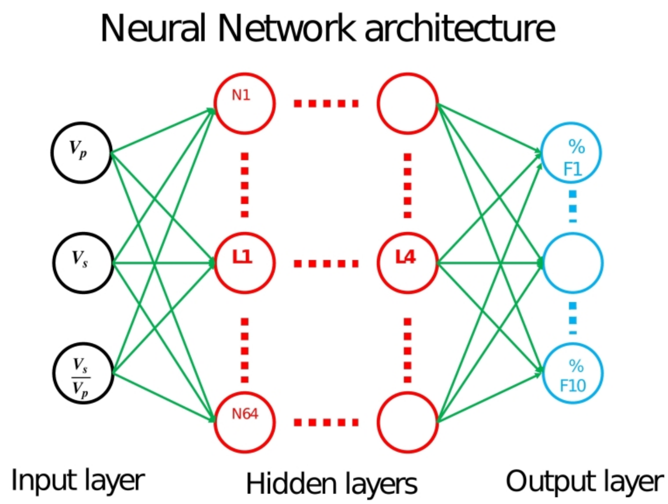

Figure 1. The Neural Network architecture. Three features are used in the input layer. Four hidden layers with 64 nodes are fully connected neural networks with a dropout rate of 10%. The output layer provides probabilities of being certain facies for the current input.

338x190mm (300 x 300 DPI)

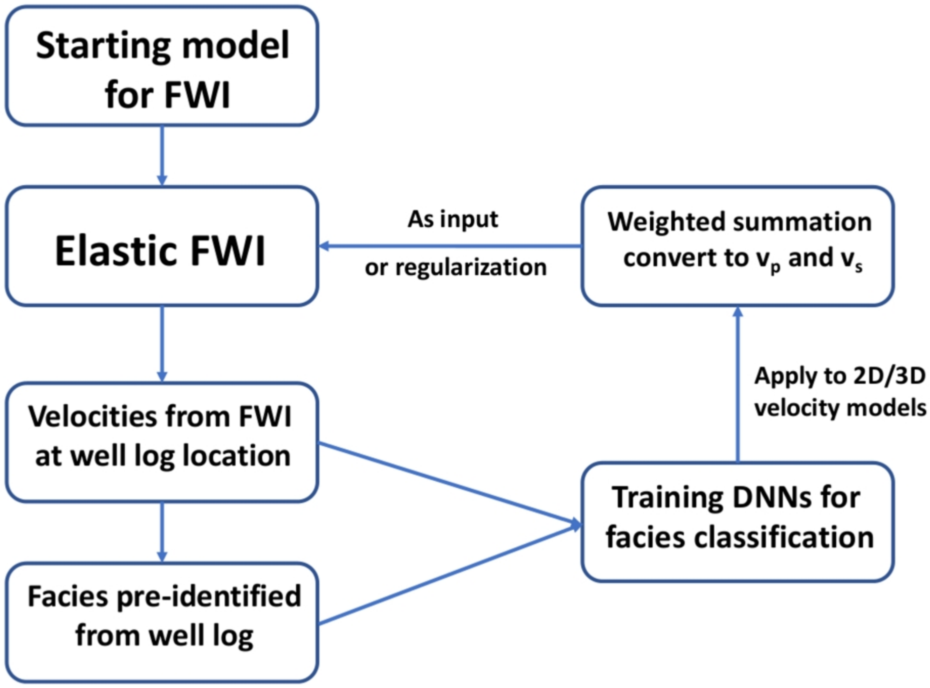

Figure 2. Workflow for the proposed method. Notice that we need to convert the estimated distribution of facies to vp and vs using a weighted summation. These converted velocities can be used as input (fully matching the seismic data) or as a conventional regularization term (matching the data and well logs) for the next stage of elastic FWI.

316x228mm (300 x 300 DPI)

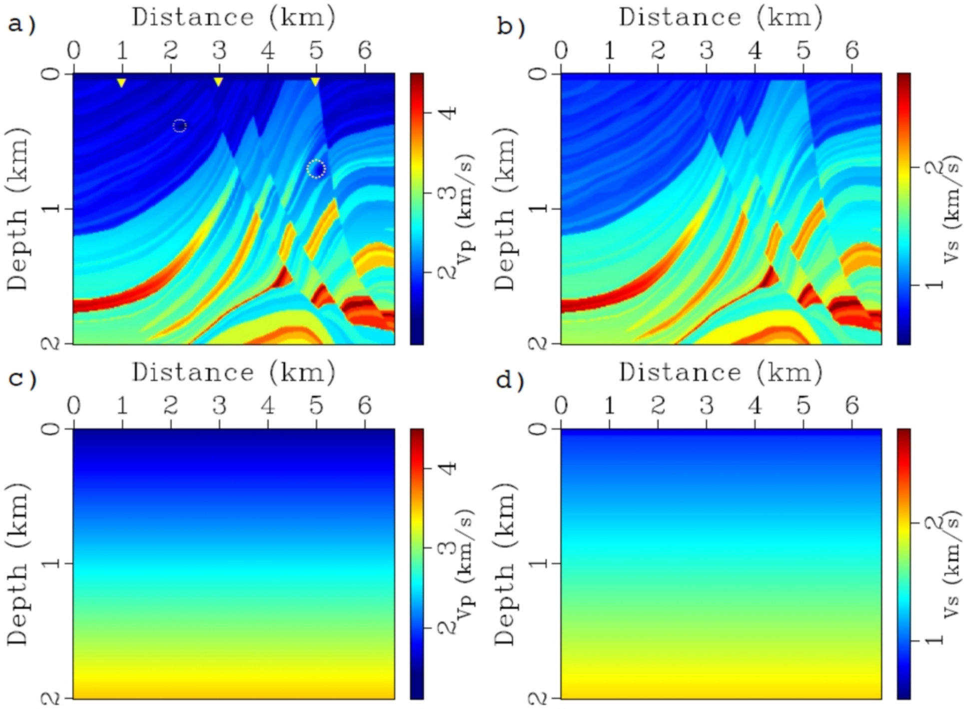

Figure 3. Velocity models. Actual vp (a) and vs (b). Initial vp (c) and vs (d). There are two low velocity zones in actual vp; actual vs=vp/1.732 + 0.1(vp-2.4). Solid triangles in (a) indicate locations of pseudo wells used in the training. The initial models are constant gradient models and vs=vp/1.732.

284x209mm (300 x 300 DPI)

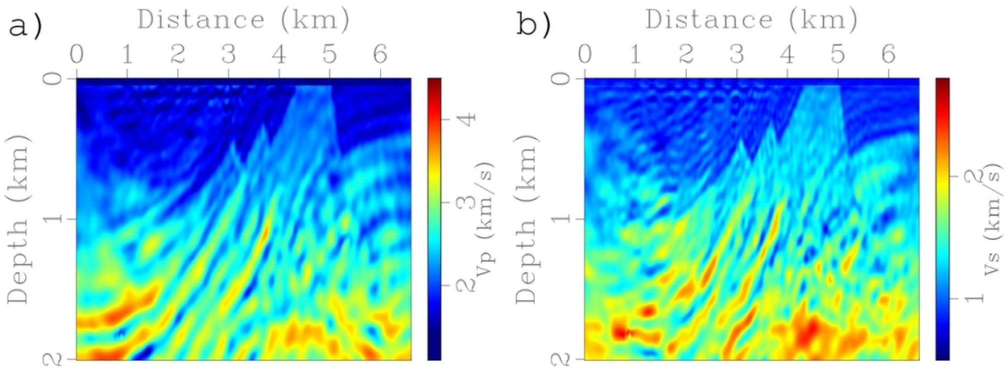

Figure 4. Estimated velocities without facies constraints. a) vp, b) vs. The inversion has apparently converged to one of the local minima.

117x42mm (300 x 300 DPI)

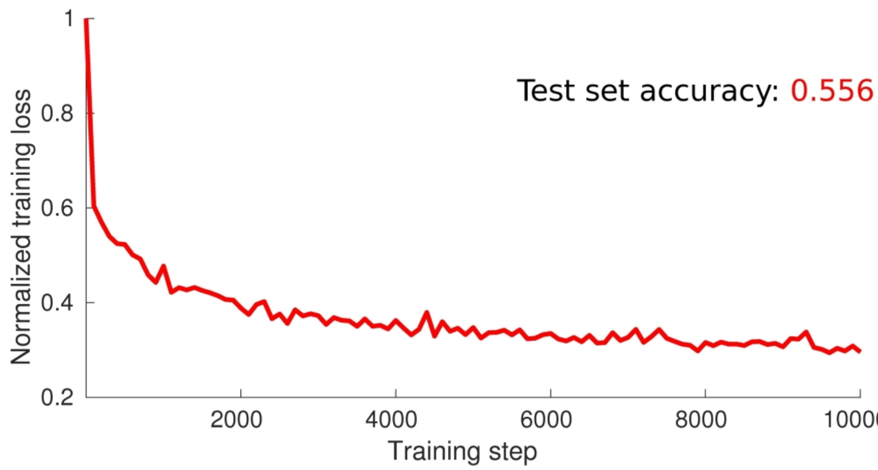

Figure 5. Normalized training loss at every 100 steps. A total of 70% training loss is achieved with a random dropout of 10% for each layer. The test set accuracy is 55.6%.

238x116mm (300 x 300 DPI)

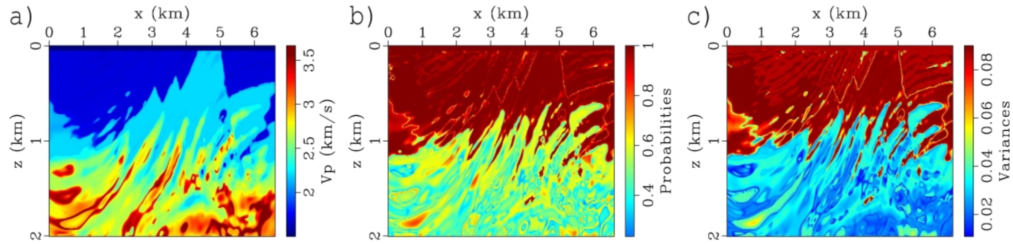

Figure 6. Classified facies. a) Converted to vp using a weighted summation ($\bar{v} = \sum_{i=1}^{n_f}p_i v_i$, v_i are given in Table 1), b) the maximum probabilities (softmax) of the classification and c) the variances of the estimated probabilities. vs is not shown here since it shares the same probability as vp.

341x80mm (300 x 300 DPI)

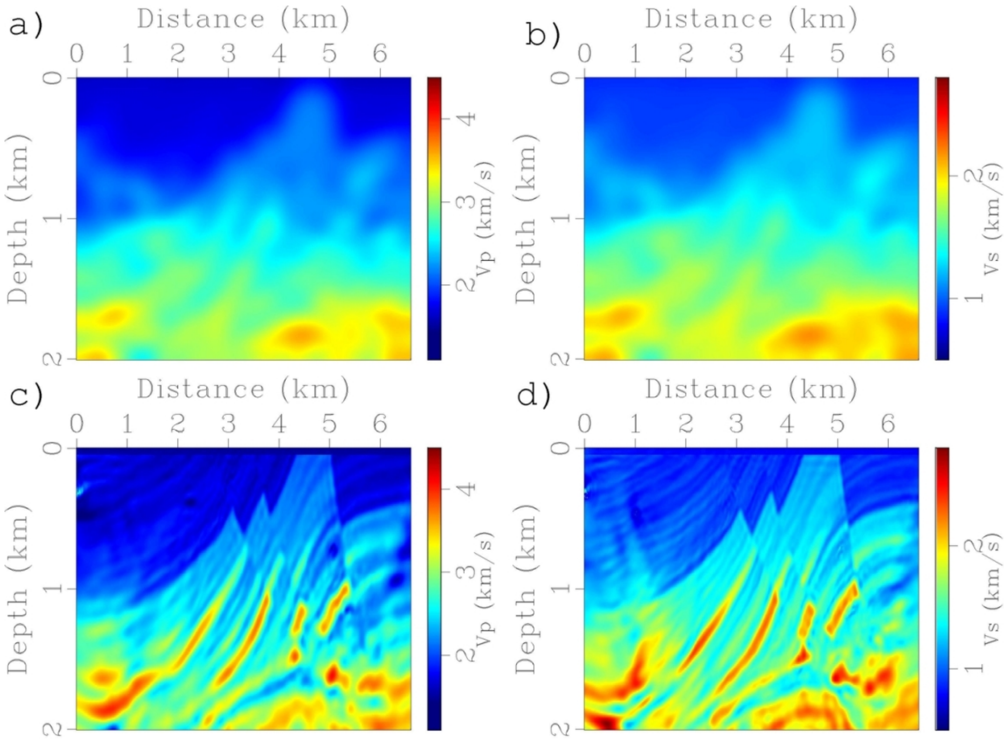

Figure 7. Estimated distribution of facies converted to vp (a) and vs (b) and final inverted vp (c) and vs (d). (a) and (b) are smoothed versions of the original estimates (e.g., vp in Figure 6a) and used as initials for obtaining (c) and (d). (c) and (d) are inverted using an L2 norm based elastic FWI.

117x86mm (300 x 300 DPI)

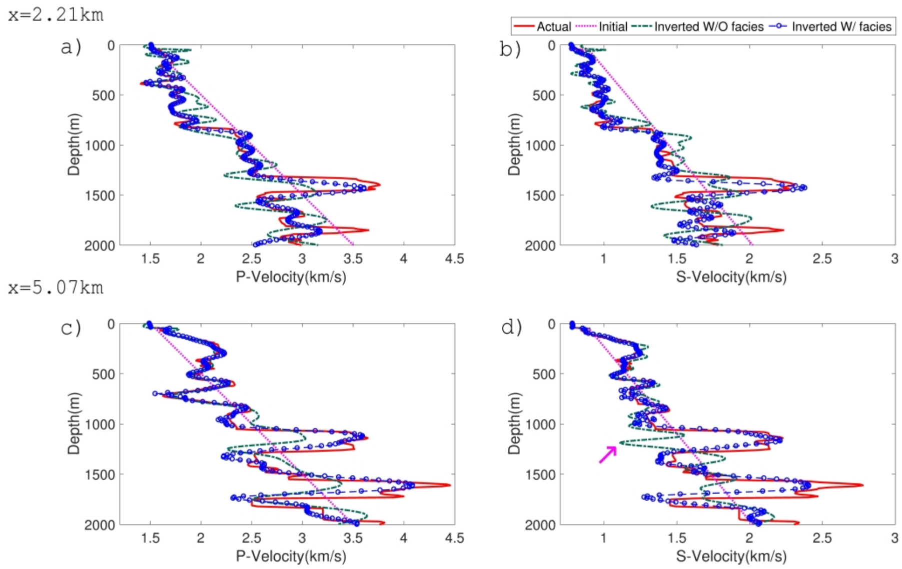

Figure 8. Vertical profiles across the low-velocity zones. The inverted velocities are far from the actual ones without using facies as constraints and vs suffers from a severe cycle-skipping problem as the pink arrow indicates in d). Facies constraints can eliminate artifacts caused by cycle skips.

507x299mm (300 x 300 DPI)

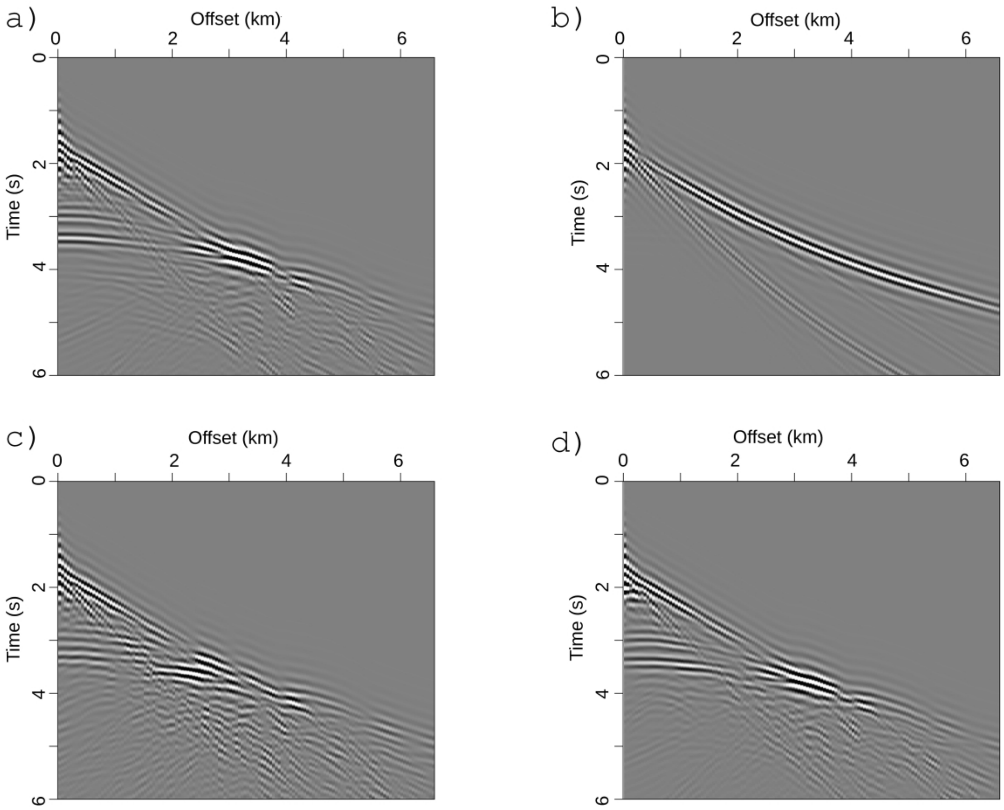

Figure 9. Data comparison (vertical component). a) Observed, b) initial, c) inverted without facies constraints and d) inverted with facies constraints. The shot gathers are plotted at the same scale so that a direct comparison can be made.

306x247mm (300 x 300 DPI)

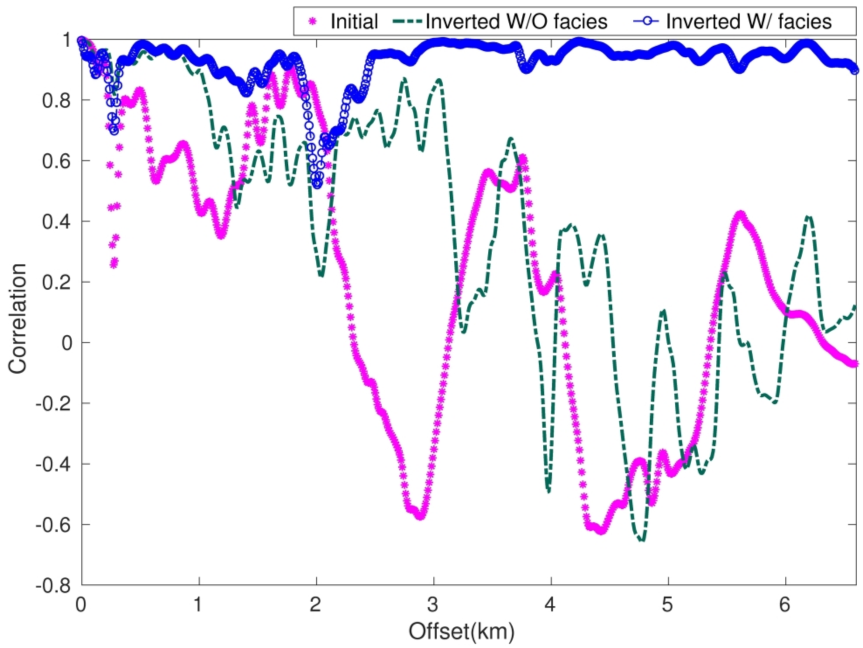

Figure 10. Correlation of the predicted and observed data (\$\frac{\textbf{u}\cdot\textbf{d}}{\sqrt{\textbf{u}\cdot \textbf{u}}\sqrt{\textbf{d}\cdot \textbf{d}}}\$). The initial model cannot provide accurate prediction in the far-offsets. The adaptive-selection objective function fails when the predicted and observed data are far from each other. The inverted model of the proposed approach can provide accurate prediction at the far-offsets.

270x182mm (300 x 300 DPI)

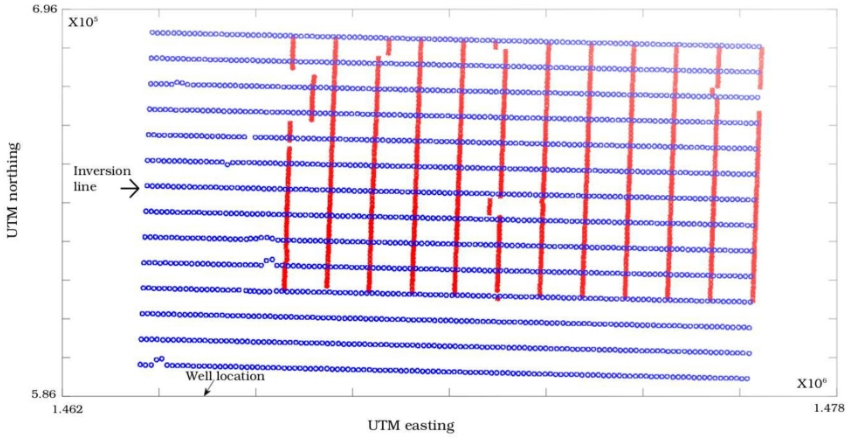

Figure 11. 3D survey geometry. We choose one 2D line for inversion as indicated by the arrow. There is a well at the edge of the survey area.

402x192mm (300 x 300 DPI)

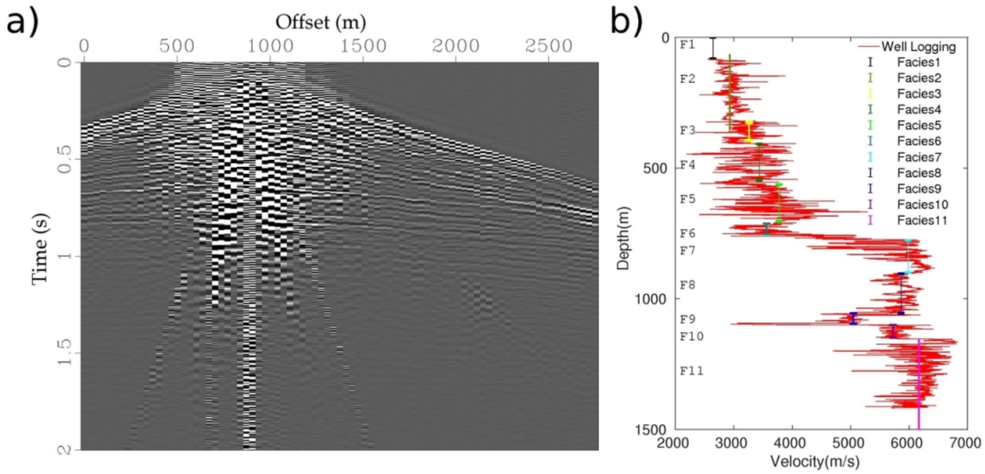

Figure 12. a) An example from the raw data set (vertical component) and b) extracted facies from the P-wave well log. The shot gather is noisy and lacks low frequencies, which is challenging for FWI. Line segments in different colors indicate the interpreted facies.

120x58mm (300 x 300 DPI)

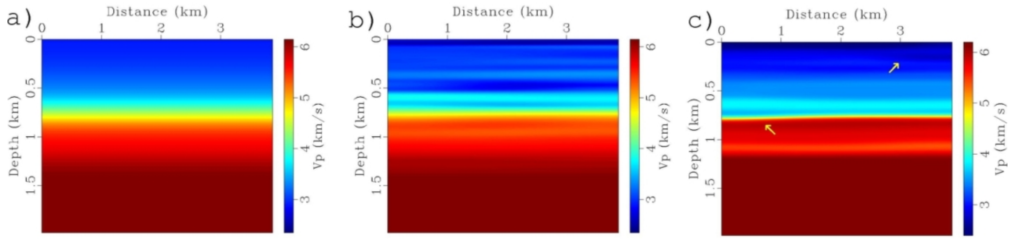

Figure 13. P-wave velocities. a) Initial vp from the well logs, b) estimated vp without facies constraints and c) estimated vp with facies constraints. Arrows point to lateral variations in the estimates.

120x28mm (300 x 300 DPI)

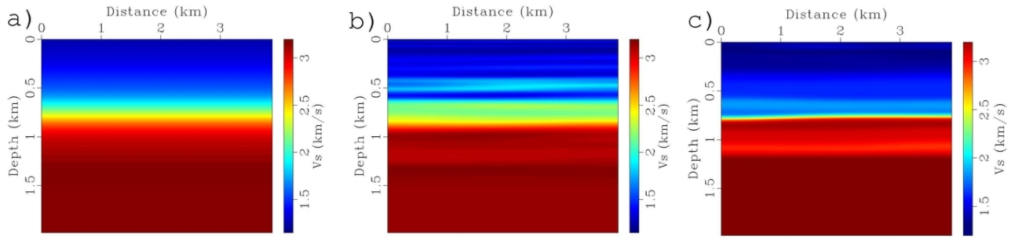

Figure 14. S-wave velocities. a) Initial vs from the well logs, b) estimated vs without facies constraints, and c) estimated vs with facies constraints.

120x28mm (300 x 300 DPI)

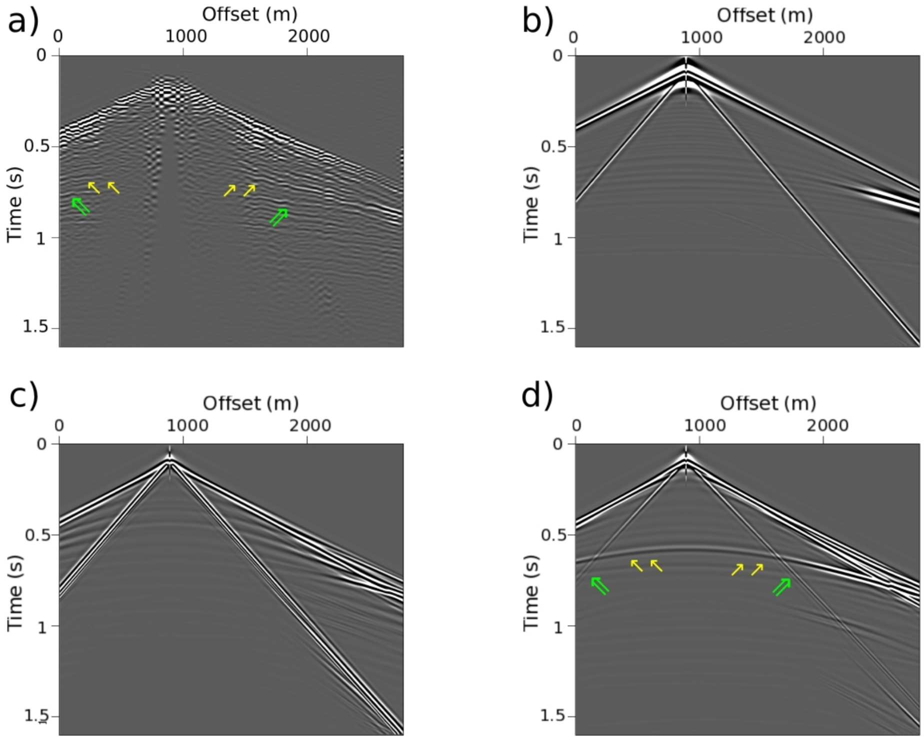

Figure 15. Data comparison (vertical component). a) Observed, b) initial, c) inverted without facies constraints and d) inverted with facies constraints. The proposed facies constraints can predict some high-contrast layers and generate strong reflections hidden in the noisy seismic data. The arrows point to some of these reflections. The shot gathers are plotted at the same scale so that a direct comparison can be made.

335x244mm (300 x 300 DPI)

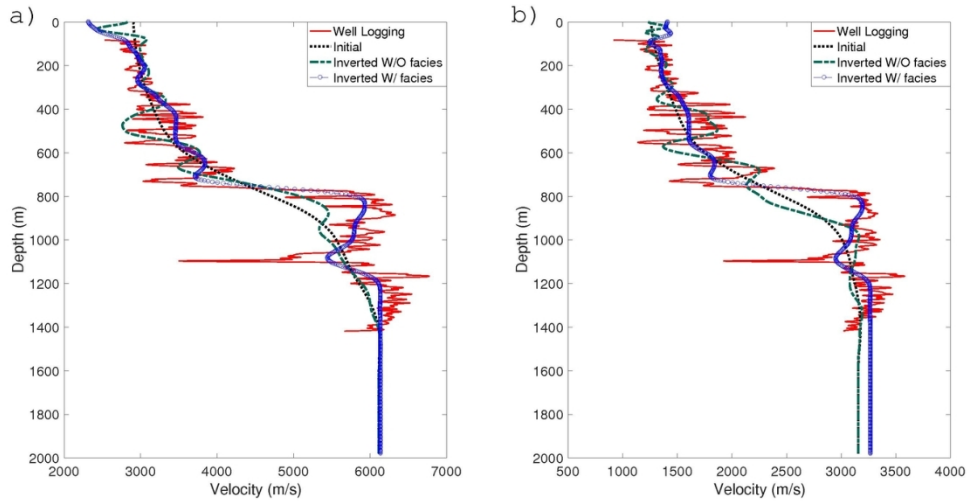

Figure 16. Vertical profiles of the smoothed well log, initial model, inverted model without facies constraints and inverted model with facies constraints. a) vp and b) vs.

117x63mm (300 x 300 DPI)

DATA AND MATERIALS AVAILABILITY

Data associated with this research are available and can be obtained by contacting the corresponding author.