

Deep decomposition learning for reflectivity inversion

(October 22, 2022)

Running head: **Deep decomposition learning for reflectivity inversion**

ABSTRACT

We report a combination of classical regularization theory with a null space neural network approach based on deep decomposition learning, paying particular attention to the solution of one ubiquitous problem in seismic exploration: the recovery of full-band reflectivity from band-limited seismic traces. The method extends the popular post-processing approach by learning how to improve an initial reconstruction with estimated missing components from the null space of the forward operator, which in our case, are the missing frequency components of the reflectivity. We integrate the null space element prediction to act in conjunction with CNN-based denoising and a data-consistent algorithm. The proposed framework honours the input measurements while enforcing generalization. Numerical experiments on synthetic and real datasets show that the proposed method naturally enforces a high-resolution prediction consistent with the low-resolution input seismic traces. We compare its performance with state-of-the-art thin-bed reflectivity estimation methods.

reflectivity inversion, deep learning, inverse problem, neural networks, seismic, deconvolution

INTRODUCTION

Enhancing the resolution of band-limited seismic data via effective deconvolution techniques has always been a recurring goal in exploration seismology. Superior bandwidth content translates into potentially resolving reflectors hidden under the tuning thickness and ultimately provides more accurate structural and stratigraphic information from high-resolution images (Chopra et al., 2006, 2009). After the reflection data undergoes a sequence of processing steps (Levin, 1989), a common assumption is that the zero-offset seismic trace can be modelled as a linear system. The linear system entails the convolution of a band-limited source wavelet (i.e., the blurring kernel with most of its energy concentrated within some pass-band) with the earth's impulse response, which is typically conceived as a broadband reflectivity time series representing layers of constant material parameters (Robinson and Treitel, 1980). The deconvolution process attempting the frequency enhancement then considers the computation of an approximate solution for the discretized problem

$$\mathbf{d}_\epsilon = \mathbf{L}\mathbf{m} + \epsilon, \quad (1)$$

where $\mathbf{m} \in \mathbb{R}^n$ is the (vectorized) reflectivity time series, $\mathbf{d} \in \mathbb{R}^m$ contains the seismic traces, $\epsilon \in \mathbb{R}^m$ denotes an unknown data error represented as additive noise, and the linear forward operator $\mathbf{L} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ contains the stationary seismic wavelet properly arranged into a Toeplitz matrix

$$L_{ij} = \begin{cases} w_{i-j+1} & i \geq j \\ 0 & i < j \end{cases}. \quad (2)$$

This enables convolution via simple matrix-times-vector multiplication, with w_{i-j+1} as the $i - j + 1$ th sampling point of the wavelet. For this purpose, this paper adopts a classic

deconvolution framework, in which the noise is white and Gaussian, and the wavelet is assumed time-invariant and known or at least well approximated as a preliminary step (Ulrych et al., 1995). While a more realistic model entails an unknown and non-stationary propagating wavelet, this simplification serves as a good approximation. We point out, however, that multiple efforts have also considered time-variant and blind deconvolution frameworks that simultaneously estimate the wavelet and the reflectivity in a non-linear fashion (e.g. Kaarsen and Taxt, 1998; Kazemi and Sacchi, 2013; Gholami and Sacchi, 2013; Chen et al., 2023).

Even when the wavelet is available a priori, reflectivity inversion is an underdetermined problem due to the presence of a non-trivial kernel or null space, a consequence of the lack of low and high frequencies of the seismic wavelet. Rigorously, the missing model components of \mathbf{m} lie in the spectrum gaps describing the null space of the forward operator. Given the existence of non-uniqueness, many reflectivity series fit the acquired data equally well, despite having radically different features. Only some of these solutions might be able to accurately characterize the true earth's impulse response. Additionally, the ill-posedness of the problem makes the reconstruction process more vulnerable to noise, especially in the high and low bands where the noise contribution is stronger. Thus, a direct inversion of the Toeplitz matrix $\mathbf{m} = \mathbf{L}^{-1}\mathbf{d}_\epsilon$ is impossible, and appropriate priors upon the reflectivity must be promoted in order to reduce the null space ambiguity and determine a unique and credible approximation of \mathbf{m} .

Early attempts to perform reflectivity inversion via deconvolution are based on classic inverse-filtering theory assuming a white random reflectivity sequence with Gaussian prior distribution and a minimum-phase seismic wavelet (Berkhout, 1977; Robinson and Treitel, 1980; Scales and Smith, 1994; Yilmaz, 2001). Such restrictions render a stable and causal

wavelet inverse filter that can be applied to the data to retrieve a reflectivity estimate. Nevertheless, gaussianity yields band-limited results with broadened peaks and side-lobe artefacts that preclude closely spaced reflectors to be sharply resolved. To overcome the shortcomings that typify conventional least-squares deconvolution, many methods have been proposed for high-resolution, sparse-spike or thin-bed reflectivity inversion, including techniques that use minimum entropy (Wiggins, 1978; Sacchi et al., 1994) and l_p ($0 \leq p \geq 1$) sparsity constraints (Taylor et al., 1979; Levy and Fullagar, 1981; Oldenburg et al., 1983; Debeye and Van Riel, 1990; Sacchi, 1997; Chopra et al., 2006; Zhang and Castagna, 2011; Gholami and Sacchi, 2012).

Particularly, sparse priors are only valid through mathematically enforcing that the reflectivity series consists of a few isolated spikes comprising a plane homogeneous layered model. Under this simplified construction, the seismic trace constitutes a finite superposition of seismic wavelets. This hypothesis is suitable when just a small number of strong reflections dominate the seismogram. In reality, however, data obtained from well-logs present a stochastic pattern that is far more complex (Walden, 1985; Tenorio, 2001). Furthermore, when sub-optimally applied as a trace-by-trace process, sparse deconvolution algorithms do not consider the inherent continuity in the spatial dimensions, and we require more complicated multichannel techniques to avoid harming the signal (Idier and Goussard, 1993; Kaaresen and Taxt, 1998; Gholami and Sacchi, 2013), which still may face challenges for seismic data with complex structures.

Additionally, sparse inversion algorithms suppress random noise to some extent when Gaussian statistics are assumed for the data error, but they often have significant sensitivity to outliers that negatively impacts the results (Debeye and Van Riel, 1990). Ultimately, practical applications of high-resolution deconvolution still face several challenges related

to regularization and hyper-parameter selection, intensive demand for human-computer interaction, and the high computational cost of iterative reconstruction in large 3D seismic volumes. Despite providing state-of-the-art solutions, limitations from utilizing handcrafted priors inspire the development of alternative data-driven and learning-based seismic data processing methods.

In recent years, various geophysical problems have explored applications of supervised deep learning techniques using Convolutional Neural Networks (CNN) (LeCun et al., 2015), for they have powerful representation learning properties and the potential to process extensive seismic surveys with minimal human intervention (Yu and Ma, 2021). In the supervised regime, one trains a neural network as a universal approximator to recover model parameters from observed data with many high-quality pre-labelled solutions from a representative (training) dataset. Deep learning applications on seismic inversion include end-to-end approaches (Araya-Polo et al., 2018; Mandelli et al., 2019; Chai et al., 2021; Wu et al., 2021), in which the neural networks directly learn an inverse data-to-model mapping. Such networks bypass the use of explicit physics operators but rely on a vast amount of training samples to learn the underlying physics of the problem. To reduce the dependency on training data, learned iterative schemes (Torres and Sacchi, 2022a) incorporate physics into the learning process and replace various components of unrolled iterative reconstruction algorithms with neural network computations. Alternatively, to avoid iterations, the learned post-processing method first maps the measurements to the model space through a known physics operator (either the pseudoinverse $\mathbf{L}^\dagger : \mathbb{R}^m \rightarrow \mathbb{R}^n$ or an approximation to it) and then trains a neural network to learn a model perturbation that potentially improves this initial reconstruction (Kaur et al., 2020; Zhang et al., 2021). Even though these techniques have demonstrated remarkable empirical success, many supervised approaches still lack a

data consistency constraint to enforce that the predicted model matches the acquired data, a necessary condition for a reliable solution to the inverse problem.

Consequently, such deep learning schemes do not lead to convergent regularization strategies. Hence, the results might look realistic, but there is no way to assess their accuracy. Unsupervised approaches (Dhara and Sen, 2022; Chen et al., 2022; Kong et al., 2022) address this issue by design, but they often amplify the expensive iterative nature of traditional methods when the trainable weights are not correctly initialized.

Schwab et al. (2019) introduced regularization via null space networks as an alternative image domain restoration method to account for data consistency. By computing a projection onto the null space of the forward operator after the last weight layer of a residual architecture, it is possible to train a neural network to learn the missing components of the initial reconstruction. The null space projection ensures that the output estimates are consistent with the observed input data. As a generalization to null space networks for noisy data, Chen and Davies (2020) introduced the concept of deep decomposition learning, which attaches a complementary network to act as a denoiser on the range of the pseudoinverse. Similarly, Schwab et al. (2020) allow the null space networks to act on the orthogonal complement of the kernel by being dependent on the regularization technique that produces the initial reconstruction and demonstrate the convergence properties of these algorithms. Based on these ideas, we investigate the extension of data-consistent null space learning on the inversion of reflectivity. Specifically, we extend the deep decomposition approach by using the truncated singular value decomposition (TSVD) as an initial regularized reconstruction for approximating the low-frequency components of the model. As a second step, we trained two neural networks to recover the missing parts of the model and the "inverted" noise, respectively.

This paper is an extension of Torres and Sacchi (2022b). In the next section, we reintroduce the concept of null space networks for the convolutional model for normal incidence seismograms. To our knowledge, regularization via null space networks has not been reported in the literature in the context of full band reflectivity inversion. As a proof of concept, we empirically demonstrate the behaviour of the proposed method on a single-channel toy dataset in which the optimal solution is known and two 2D field datasets. These examples show that when the wavelet is known, the proposed deconvolution approach may reconstruct a suitable high-resolution reflectivity model from band-limited noisy data. The performance, we believe, is comparable to the state-of-the-art techniques in this category. The paper ends with a discussion of limitations and prospects for further development of this reflectivity inversion method.

METHOD

Deconvolution

Many classic deconvolution methods find approximate deterministic solutions to equation 1 by computing the reflectivity model \mathbf{m}^* that minimizes a composite objective function

$$\mathbf{m}^* = \arg \min_{\mathbf{m}} \{J(\mathbf{m}) + \lambda R(\mathbf{m})\}, \quad (3)$$

where the first term is a convex data-fidelity term, usually defined as the least-squares data error $J(\mathbf{m}) = \|\mathbf{L}\mathbf{m} - \mathbf{d}_\epsilon\|_2^2$, $R(\mathbf{m})$ represents a regularization function, and $\lambda > 0$ denotes the regularization parameter. For example, Wiener or Tikhonov regularized deconvolution can be derived assuming a Gaussian prior for the reflectivity $R(\mathbf{m}) = \|\mathbf{m}\|_2^2$, which has

analytical solution

$$\mathbf{m}_\lambda^* = [\mathbf{L}^T \mathbf{L} + \lambda \mathbf{I}]^{-1} \mathbf{L}^T \mathbf{d}_\epsilon, \quad (4)$$

where \mathbf{L}^T is the transpose of \mathbf{L} (a cross-correlation operator) and $\mathbf{I} \in \mathbb{R}^n$ is the identity operator. Analogously, when the problem allows an explicit singular value decomposition (SVD) of \mathbf{L} , the truncated SVD (TSVD) pseudo-inverse \mathbf{L}_k^\dagger can be used as an intuitive brute-force low-rank approximation to obtain a filtered solution

$$\mathbf{m}_k^* = \mathbf{L}_k^\dagger \mathbf{d}_\epsilon \quad (5)$$

$$= \mathbf{V}_k \mathbf{S}_k^{-1} \mathbf{U}_k^T \mathbf{d}_\epsilon, \quad (6)$$

with $\mathbf{U}_k \in \mathbb{R}^{m \times k}$ and $\mathbf{V}_k \in \mathbb{R}^{n \times k}$ as the top k rows of the unitary matrices $\mathbf{U} \in \mathbb{R}^{m \times m}$ and $\mathbf{V} \in \mathbb{R}^{n \times n}$, respectively, and $\mathbf{S}_k \in \mathbb{R}^{k \times k}$ as a diagonal matrix containing the largest k singular values of \mathbf{L} . TSVD and Tikhonov regularizations provide similar solutions under certain conditions (Hansen, 2010). In both cases, the regularization parameters λ and k attempt to reduce the influence of noise by suppressing the less reliable signal components. However, it is well-known that none of these techniques are suitable when the solution is discontinuous. Deconvolution via Tikhonov regularization, for instance, tends to smear the solution by enforcing that the reflectivity coefficients concentrate close to zero, considering $R(\mathbf{m}) = \|\mathbf{m}\|_2^2 = \|\mathbf{m} - \mathbf{m}_{\text{prior}}\|_2^2$ with $\mathbf{m}_{\text{prior}} = \mathbf{0}$ as a vector of all zeros. The TSVD pseudo-inverse truncates the spectrum of \mathbf{L} and only provides minimum norm solutions since no elements in the null space of \mathbf{L} are added to the solution.

On the other hand, the layered earth model visualizes the unknown reflectivity series as

a train of isolated spikes represented by a set of delta functions

$$m_n = \sum_{s=0}^{S-1} a_s \delta_{n-\Gamma_s} \quad (7)$$

where S is the number of nonzero spikes of amplitude a_s , and $\Gamma \subset \{0, \dots, N-1\}$. Thus, enforcing a sparse property to the reconstructed reflectivity through the long-tailed l_1 -norm while retaining the l_2 -norm on the data-fidelity term yields efficient sparse-spike deconvolution algorithms that seek the solution to the convex problem

$$\min_{\mathbf{m}} \left\{ \frac{1}{2} \|\mathbf{L}\mathbf{m} - \mathbf{d}_{\text{obs}}\|_2^2 + \mu \|\mathbf{m}\|_1 \right\}, \quad (8)$$

which allows resolving closely spaced reflectors. Nonetheless, controlling the sparsity level of the reconstructed reflectivity is challenging. In particular, because of the constant shrinkage parameter μ , l_1 -regularization results in a biased estimate by over-penalizing large-valued reflectivity coefficients. Alternatively, we propose to solve the reflectivity inversion problem given in equation 1 with a deep decomposition technique based on null space network regularization that constructs a learnable prior from training data and possesses data consistency and denoising properties.

Deep decomposition learning for reflectivity inversion

The outset of the deep decomposition learning as a regularization technique is the active-null space decomposition of the signal, i.e., we decompose the domain of the forward operator into two sub-spaces: the measurement space and the null space. Accordingly, we might think of any reflectivity model \mathbf{m} in the domain of \mathbf{L} as being made up of two unique

orthogonal vectors,

$$\mathbf{m} = \mathbf{m}_R + \mathbf{m}_N = P_R(\mathbf{m}) + P_N(\mathbf{m}), \quad (9)$$

such that \mathbf{m}_R lies in the range of the pseudo-inverse \mathbf{L}^\dagger which is also the active space solution, and \mathbf{m}_N lies in the null space. By definition, these two components satisfy, respectively,

$$\mathbf{m}_R = \mathbf{L}^\dagger \mathbf{d}_\epsilon \quad (10)$$

$$= \mathbf{L}^\dagger \mathbf{L} \mathbf{m} + \mathbf{L}^\dagger \boldsymbol{\epsilon}, \quad (11)$$

and

$$\mathbf{L} \mathbf{m}_N = 0. \quad (12)$$

Based on this fragmentation of the model, we can define the ideal reconstruction as

$$\mathbf{m}^* = \mathbf{L}^\dagger \mathbf{d}_\epsilon - \mathbf{L}^\dagger \boldsymbol{\epsilon} + \mathbf{m}_N. \quad (13)$$

In other words, the solution is expressed in terms of a unique minimum norm least-squares solution ($\mathbf{L}^\dagger \mathbf{d}_\epsilon$) minus the "inverted" noise plus the null space vector.

As denoted in equation 9, the model components can be obtained from two orthogonal projections, P_R and P_N , defined as

$$P_R = \mathbf{L}^\dagger \mathbf{L}, \quad (14)$$

and

$$P_N = \mathbf{I} - \mathbf{L}^\dagger \mathbf{L}. \quad (15)$$

Using a physics-engaged approach promoted by the application of the above-mentioned orthogonal projections, deep decomposition learning attempts to solve equation 13 with a trained estimator $\Lambda : \mathbb{R}^m \rightarrow \mathbb{R}^n$ defined as

$$\Lambda(\mathbf{d}_\epsilon; \theta_1, \theta_2) = \mathbf{L}^\dagger \mathbf{d}_\epsilon + P_R \circ \mathbf{F}_{\theta_1} \circ \mathbf{L}^\dagger \mathbf{d}_\epsilon + P_N \circ \mathbf{N}_{\theta_2} \circ (\mathbf{L}^\dagger \mathbf{d}_\epsilon + P_R \circ \mathbf{F}_{\theta_1} \circ \mathbf{L}^\dagger \mathbf{d}_\epsilon), \quad (16)$$

where \mathbf{F}_{θ_1} and \mathbf{N}_{θ_2} are two trainable neural networks. Compared to equation 13, it is clear that the second term in the right-hand side of equation 16 tries to estimate the negative "inverted" noise by projecting the output of the network \mathbf{F}_{θ_1} onto the range of the pseudo-inverse. Likewise, the third term in equation 16 tries to estimate the null space component from the denoised input $\mathbf{L}^\dagger \mathbf{d}_\epsilon + P_R \circ \mathbf{F}_{\theta_1} \circ \mathbf{L}^\dagger \mathbf{d}_\epsilon \approx \mathbf{L}^\dagger \mathbf{d}_\epsilon - \mathbf{L}^\dagger \epsilon$. When Λ lacks the explicit denoising element ($P_R \circ \mathbf{F}_{\theta_1} \circ \mathbf{L}^\dagger \mathbf{d}_\epsilon = 0$), the estimator turns into a standard null space regularization network of the form

$$\begin{aligned} \Lambda(\mathbf{d}_\epsilon; \theta) &= \mathbf{L}^\dagger \mathbf{d}_\epsilon + P_N \circ \mathbf{N}_\theta \circ \mathbf{L}^\dagger \mathbf{d}_\epsilon \\ &= (\mathbf{I} + P_N \circ \mathbf{N}_\theta)(\mathbf{L}^\dagger \mathbf{d}_\epsilon), \end{aligned} \quad (17)$$

where the data consistency property, $\mathbf{L}\Lambda(\mathbf{d}_\epsilon; \theta) = \mathbf{d}_\epsilon$, is exactly preserved.

For noisy seismic data, the results obtained using the naive solution $\mathbf{L}^\dagger \mathbf{d}_\epsilon$ are non-interpretable. Therefore, in this work, we use TSVD for computing a regularized approximation \mathbf{L}_k^\dagger to the pseudo-inverse, where the regularization parameter is given by the number

of non-truncated singular values $k > 0$. When the true model is unavailable, or the noise level is unknown, one may refer to data-dependent strategies to obtain an approximation of the optimal truncation, such as the *optimal singular value hard thresholding* (Gavish and Donoho, 2014). The solution obtained by TSVD is the classical regularized least-square solution that will yield a reflectivity of spectral properties similar to the original data and not significant gain in bandwidth. By only recovering signal components corresponding to sufficiently large singular values, TSVD produces a stable solution that prevents small singular values of \mathbf{L} from amplifying noise. The absent components (frequencies in the null space of the operator) have an unreliable model-to-data mapping and, therefore, will be recovered in the learning stage. In other words, because of the impossibility of precisely differentiating the signal from noise in the active space, we partially recover some reflections with TSVD filtering and then predict the residual noise component with a denoising network. Next, we attempt to increase the high-frequency content of the reflectivity with the estimation of the null space element, which limits the non-uniqueness of the problem. Instead of explicitly incorporating a sparse regularizer in an iterative least-squares inversion, the spike representation of the reflectivity is embedded in the manifold of representative solutions in the training labels. The null space network will learn it during training.

Finally, the overall procedure aims to jointly seek the weights that minimize

$$E(\theta_1, \theta_2) = \frac{1}{N} \sum_{i=1}^N \|\mathbf{m}^i - \Lambda(\mathbf{d}_e^i; k, \theta_1, \theta_2)\|_2^2 + \lambda_1 \sum_{i=1}^N \|\mathbf{LF}_{\theta_1}(\mathbf{L}_k^\dagger \mathbf{d}_e^i) - \epsilon^i\|_2^2 + \lambda_2 \|\theta_2\|_2^2, \quad (18)$$

where the first term enforces "benignant" inductive bias by carrying out supervised training on a synthetic dataset $\mathcal{D} = \{(\mathbf{m}^i, \mathbf{d}_e^i, \epsilon^i)\}_{i=1}^N$ using the Mean Squared Error (MSE) loss, the second term prevents the denoising component from breaking the data consistency property.

The third term provides the null space estimator with robustness to small perturbations via weight regularization (Schwab et al., 2019).

We define \mathbf{F}_{θ_1} as a denoising convolutional neural network (DnCNN) architecture (Zhang et al., 2017) (Figure 1) designed to predict the residual noise component from a corrupted input with the operations in the hidden layers. We set \mathbf{N}_{θ_2} as an encoder-decoder neural network, shown in Figure 2, inspired by the U-net architecture (Ronneberger et al., 2015). This type of architecture is a two-stage convolutional neural network commonly used in image segmentation and reconstruction from insufficient data due to its straightforward design, high representation power, and fast convergence in training. The encoder portion of \mathbf{N}_{θ_2} implements convolutions and downsampling operations to extract high-level features from the inputs. The decoder then performs convolutional and upsampling operations on the retrieved features to produce the required outputs. Skip connections feed the extracted features from the encoder sections to their respective decoder sections to recover the information lost in the downsampling process. Compared to typical neural networks, the downsampling operations provide the network with a broad field of view of signal features at a reduced cost, and the upsampling operations produce high-quality outputs with the desired size. We add dropout layers to improve the prediction performance (Tompson et al., 2015).

[Figure 1 about here.]

[Figure 2 about here.]

A strength of the null space network component is that the estimator $\Lambda(\mathbf{d}_e; k, \theta_1, \theta_2)$ only adds missing information without introducing inconsistencies with the acquired data, even

when applied to models very different from the training data, which in principle can improve the reconstruction quality if compared to non-data-consistent approaches. Algorithm 1 summarizes the proposed scheme.

Algorithm 1 Deep decomposition estimator $\Lambda(\mathbf{d}_\epsilon; k, \theta_1, \theta_2)$

Requires:

- Forward operator \mathbf{L} (constructed with an estimated wavelet as per equation 2)
- Projection operators P_R, P_N
- Training dataset $\mathcal{D} = \{(\mathbf{m}^i, \mathbf{d}_\epsilon^i, \epsilon^i)\}_{i=1}^N$

Parameters:

$$(k, \lambda_1, \lambda_2) > 0, \text{ stopping criteria}$$

- 1: SVD of $\mathbf{L} \rightarrow \mathbf{L}_k^\dagger$
 - 2: Training stage:
 - 3: repeat for all labels until stopping criteria
 - 4: $\Lambda(\mathbf{d}_\epsilon^i; k, \theta_1, \theta_2) = \mathbf{L}_k^\dagger \mathbf{d}_\epsilon^i + P_R \circ \mathbf{F}_{\theta_1} \circ \mathbf{L}_k^\dagger \mathbf{d}_\epsilon^i + P_N \circ \mathbf{N}_{\theta_2} \circ (\mathbf{L}_k^\dagger \mathbf{d}_\epsilon^i + P_R \circ \mathbf{F}_{\theta_1} \circ \mathbf{L}_k^\dagger \mathbf{d}_\epsilon^i)$
 - 5: Minimize $\frac{1}{N} \sum_{i=1}^N \|\mathbf{m}^i - \Lambda(\mathbf{d}_\epsilon^i; k, \theta_1, \theta_2)\|_2^2 + \lambda_1 \sum_{i=1}^N \|\mathbf{L}\mathbf{F}_{\theta_1}(\mathbf{L}_k^\dagger \mathbf{d}_\epsilon^i) - \epsilon^i\|_2^2 + \lambda_2 \|\theta_2\|_2^2$
 - 6: **Output** Trained parameters θ_1, θ_2
 - 7: Inference stage:
 - 8: **Input:** Observed data \mathbf{d}_ϵ
 - 9: **Output** $\mathbf{m}_\Lambda^* = \mathbf{L}_k^\dagger \mathbf{d}_\epsilon + P_R \circ \mathbf{F}_{\theta_1} \circ \mathbf{L}_k^\dagger \mathbf{d}_\epsilon + P_N \circ \mathbf{N}_{\theta_2} \circ (\mathbf{L}_k^\dagger \mathbf{d}_\epsilon + P_R \circ \mathbf{F}_{\theta_1} \circ \mathbf{L}_k^\dagger \mathbf{d}_\epsilon)$
-

As noticed, computing the trained estimator only requires one SVD of the forward operator before training, so the technique is practically convenient. The following section presents evidence that regularization via deep decomposition can be applied to field seismic data sets and, at least in the cases exhibited here, produce plausible results.

NUMERICAL EXAMPLES

We evaluate the proposed learned regularization method on a synthetic example and two real data sets. Setting the trade-off parameters to $\lambda_1 = 10e^{-6}$ and $\lambda_2 = 10e^{-8}$ yields stable solutions for the three experiments considered in this section. Stochastic gradient descent with the adaptive moment estimation scheme (Kingma and Ba, 2014) minimizes equation 18, running 400 epochs as the stopping criteria with a learning rate of 0.001.

Synthetic example: single-channel deconvolution

Relying on the convolutional model of the seismic trace, we attempt to recover a known full-band reflectivity sequence from a zero-offset seismic trace. To train the estimator, we generate 5000 1D random reflectivity samples and obtain the corresponding data by convolving a 60 Hz Ricker wavelet and adding the observational noise. For this, we incorporate different realizations of unstructured Gaussian noise to the clean data samples such that the signal-to-noise ratio equals 20%. The maximum amplitude of the wavelet is unity. The deconvolution process uses the exact wavelet and signal-to-noise ratio.

Figure 3 shows the results for a test reflectivity model different from all training samples but generated using the same random procedure. For a quantitative evaluation of the results, we calculate the reconstruction accuracy ($\text{dB} = 10 \times \log_{10} \|\mathbf{m}\|_2^2 / \|\mathbf{m} - \mathbf{m}^*\|_2^2$, where \mathbf{m} and \mathbf{m}^* are the true and inverted reflectivity models, respectively. Figure 3b shows the initial TSVD solution, where some of the most prominent reflections are partially recovered, but a residual band-limited component masks the non-resolvable events. Moreover, oscillations are visible in the result, courtesy of the truncated expansion. For comparison, Figure 3c displays a sparse-spike deconvolution obtained as a solution of equation 8 via 100 iterations of the FISTA solver with regularization parameter $\mu = 0.5$. Even though the inversion with sparsity promotion enables a full-band solution, the noisy data impedes successfully retrieving the correct amplitude and positioning of some events, and the solution presents spurious spikes. Figure 3d shows the result obtained with deep decomposition regularization, where we conclude that, both visually and quantitatively, the learned estimator inversion produces higher quality results and fewer spurious events compared to the two previous techniques. Finally, Figures 4a and 4b show the amplitude spectrum of the true

and the estimated result and the solution's range and null space components, respectively. Both figures show that deep decomposition regularization approximately recovers the true missing frequency components of the original signal.

[Figure 3 about here.]

[Figure 4 about here.]

Albeit synthetically, this example demonstrates that the proposed algorithm provides a broad-band, sparse and spiky reflectivity solution that, under controlled conditions, is more faithfully resolved than the filtered TSVD and spike-sparse deconvolution.

Field data examples

We also adopted the proposed estimator to predict full-band reflectivity models from two real 2D datasets. In a 1D trace-by-trace deconvolution technique, thin layers can be vertically subtle, and the desired spatial coherence might be easily lost (Wu et al., 2021). Instead of using 1D models and noisy traces, the networks were trained with synthetically generated 2D reflectivity models and their corresponding noisy data slices to improve lateral continuity and robustness to noise. We follow Torres and Sacchi (2022a) to generate synthetic reflectivity models (as shown in Figure 5) considering the spatial correlation imposed by depositional processes while mimicking fractured, faulted and folded sedimentary structures that can take arbitrary shapes and orientations. The training samples reflect that physical, nonrandom processes produce the earth's geology, and subsurface layers can have many arbitrary orientations, including sharp discontinuities.

[Figure 5 about here.]

Despite dealing with different geological settings, both experiments used the same reflectivity labels in the training data to test the generalization properties of the proposed estimator. We notice high-resolution results even when there is a significant shift in the data distribution.

Alberta foothills seismic data

The first dataset entails a 2D seismic section from Alberta, Canada, in the foothills of the Rocky Mountains (Chopra et al., 2009). It consists of 300 traces digitized in increments of 1 ms. Figure 6 shows the estimated wavelet and its frequency spectrum.

[Figure 6 about here.]

Figure 7 portrays the results, where for completeness, we also compare the proposed method to the output of a thin-bed spectral inversion (TSI) (the thin-bed reflectivity inversion method used here is commercially available as ThinMan) (Chopra et al., 2006) (Figure 7c) and the solution obtained with a learned post-processing residual U-net architecture (ResUnet) (Figure 7d) given by $\mathbf{m}_{\text{ResUnet}}^* = (\mathbf{I} + \mathbf{U}_\theta)(\mathbf{L}_k^\dagger \mathbf{d}_e)$, in which \mathbf{U}_θ has the same design as \mathbf{N}_{θ_2} . For consistency, the ResUnet was trained by minimizing the MSE loss $\sum_i^N \|\mathbf{m}_i - (\mathbf{L}_k^\dagger \mathbf{d}_e^i - \mathbf{U}_\theta(\mathbf{L}_k^\dagger \mathbf{d}_e^i))\|_2^2$ using the same training data, number of epochs and learning rate as the proposed estimator.

No additional information (e.g., sonic logs) is available in the studied area, which limits the validation of the results. However, the deep decomposition output compares favourably with the conventional TSI solution, which provides confidence that the proposed technique has revealed substantial reflectivity information hidden in the original seismic section. Moreover, the predicted deconvolution output of the deep estimator shows more continuous re-

flectors and fewer noise artifacts than TSI due to an improvement in the lateral continuity and the explicit denoising component. At first glance, the deep decomposition prediction performs similarly to the ResUnet solution. Nonetheless, the residual data panel shown in Figure 7e presents almost no signal for the proposed solution. In other words, the reconvolution of the deep decomposition result with the wavelet is almost identical to the input data. It indicates the proposed deconvolution’s ability to better preserve the overall solution’s fidelity to the measured data. On the other hand, a moderate amount of residual error is visible for both ResUnet and TSI re-simulated outputs (Figures 7f and 7g). The degree of approximation shows clearly in the comparison of a representative middle trace in terms of the Pearson correlation coefficient (Figure 8)

[Figure 7 about here.]

[Figure 8 about here.]

The deep decomposition estimator offers versatility from the image perspective as we can decompose the prediction of the proposed deconvolution operator in the three panels shown in Figure 9 for visual inspection, according to equation 13. Figure 9a displays the minimum norm solution alone, anticipated to fail due to its lack of resolution. The inverted noise panel (see Figure 9b) mainly presents uncorrelated and band-limited noise. The predicted null space component (Figure 9c) exhibits the thin-beds and high-frequency features.

[Figure 9 about here.]

Lastly, Figure 10 compares the average power spectral density of the inverted reflectivities. We see that the three deconvolution methods yield an increase in the spectral content

for both low and high frequencies compared to the input data, which significantly improves the resolution.

[Figure 10 about here.]

Penobscot seismic data: a blind well test

We further test the deconvolution estimator in a third example comprising a 2D section of the Penobscot 3D survey (crossline number 1155), an offshore dataset recorded at a 4ms sampling rate. Figure 11 shows the statistically extracted wavelet and its frequency spectrum.

[Figure 11 about here.]

Figure 12 shows the predicted output along with the input data. In general, the deconvolved solution is superior if we keep in mind the quality of the input seismic data, not only in terms of extra reflection cycles but also in the fault detail. While there is an evident loss of detail in the data's deeper part (approximately after 2 seconds), the amplitudes and lateral coherence of the reflectors are well preserved throughout the deconvolved traces.

[Figure 12 about here.]

Additionally, we notice the data's relative reliability since the spectrum's overall shape (see Figure 13) within the seismic main frequency band is approximately preserved and enhanced.

[Figure 13 about here.]

The studied area also contains one well log that we use to compare the reflectivity profiles. We first calculate the impedance and reflectivity curves from the P-wave sonic and density logs. Then, the reflectivity is converted to the time domain using the P-wave log data. The final synthetic trace is obtained by convolving the well-log-derived reflectivity with the estimated wavelet. We observe a good correlation between the seismic sections (both the input data and the deconvolved result) and the synthetic seismogram, especially in key horizons. For more detail, Figure 14 shows a close-up view of the well-derived synthetic trace and the real data and predicted traces in the exact location of the well-log data.

[Figure 14 about here.]

Furthermore, we validate the inverted results in Figure 15 by comparing the predicted reflectivity convolved with a known bandpass wavelet with the synthetic seismogram. In general, the predicted reflectivity recovers the most significant features, and we observe an appropriate degree of agreement with the well-derived reflectivity despite not applying any stretch or squeeze of the well-log in this process. A point worth considering is that a perfect well-tie is impossible because we used a simplistic physics model that neglects more complex phenomena such as a time and space-variant source wavelet and a dispersive media (White and Hu, 1998).

[Figure 15 about here.]

Similar to the previous example, Figure 16 displays the individual elements of the deconvolved prediction in Figure 12b. Once again, the TSVD output (Figure 16a) presents the inevitable band-limited deconvolution along with a general loss in resolution with depth. Figure 16b exhibits artefacts in the inverted noise panel, and Figure 16c presents thin-bed

reflectors as the predicted null space components.

[Figure 16 about here.]

Uncertainty quantification

In the previous examples, the deconvolution problems were solved deterministically and did not provide any uncertainty on the obtained solution. Uncertainty estimates, however, may give valuable additional insight into reflectivity estimation by revealing the reliability of the predictions, which might impact subsequent decision-making. In this section, we focus specifically on epistemic uncertainty quantification, which, in our neural network setting, refers to the uncertainty associated with estimating the trainable parameters of the deep decomposition estimator conditioned on the training data. Namely, several parameter configurations $\Theta = [\theta_1, \theta_2]$ may explain the data for a particular training dataset, rendering different predictions on the input seismic at inference time. Many methods have been proposed to encode the uncertainty of neural networks over the model parameters (Abdar et al., 2021), including Bayesian neural networks (BNNs) adopting variational inference to approximate Bayesian statistics. BNNs usually double the number of parameters per layer compared to their non-Bayesian counterparts since each trainable weight is replaced by a trainable Gaussian weight distribution parametrized by its standard deviation and mean. The high parameter dimensionality of BNNs leads to slow convergence in training which might hurt performance.

As an efficient alternative that does not increase the number of parameters and requires minimal modification of the original estimator without retraining, we implement Monte Carlo Dropout (MC dropout). Unlike regular dropout regularization that only applies

dropout layers at training time, MC dropout activates dropout during test time to sample from the posterior parameter distribution. Gal and Ghahramani (2016) show the connection between dropout and approximate inference in a Gaussian process, and propose MC dropout to approximate the exact posterior inference relying on Bernoulli distributed weights.

Concretely, MC dropout utilizes layers $\mathbf{W}_i^{\text{do}} \in \Theta$ defined as

$$\mathbf{W}_i^{\text{do}} = \mathbf{W}_i \text{diag}(\mathbf{z}_i), \quad (19)$$

where \mathbf{W}_i are the weight matrices for each i -th convolutional layer before dropout is applied, $\text{diag}(\mathbf{z}_i)$ is the randomly generated dropout mask, and $\mathbf{z}_i \sim \text{Bernoulli}(p_i)$ are the randomly activated coefficients with dropout probability p_i , which can be fixed or set as a learnable parameter. We then perform Monte Carlo integration to approximate an empirical unbiased estimator

$$\hat{\mathbb{E}}(\mathbf{m}) \approx \frac{1}{T} \sum_{t=1}^T \Lambda(\mathbf{d}_\epsilon; k, \hat{\Theta}_t) \quad (20)$$

$$\approx \frac{1}{T} \sum_{t=1}^T \Lambda(\mathbf{d}_\epsilon; k, \hat{\theta}_{1t}, \hat{\theta}_{2t}), \quad (21)$$

such that $\Lambda(\mathbf{d}_\epsilon; k, \hat{\theta}_{1t}, \hat{\theta}_{2t})$ are realizations of independent draws of random dropout masks for the same estimator Λ . In practice, equation 21 is equivalent to running T stochastic forward passes through the trained estimator, all with the same input, and computing the mean of the dropout realizations (Gal and Ghahramani, 2016). Finally, we use the pixel-wise variance to indicate instabilities in the reconstruction for a particular input, which

contains information about the uncertainty of the model parameters

$$\hat{\sigma}^2(\mathbf{m}) = \frac{1}{T} \sum_{t=1}^T (\Lambda(\mathbf{d}_\epsilon; k, \hat{\theta}_{1_t}, \hat{\theta}_{2_t}) - \hat{\mathbb{E}}(\mathbf{m})) \odot (\Lambda(\mathbf{d}_\epsilon; k, \hat{\theta}_{1_t}, \hat{\theta}_{2_t}) - \hat{\mathbb{E}}(\mathbf{m})), \quad (22)$$

where \odot represents element-wise multiplication. This strategy is similar in spirit to the one proposed by Velis (2008) for quantifying the uncertainty of stochastic inversion in sparse deconvolution via fast simulated annealing.

In this example, we use the Penobscot dataset from the previous section and set $p_i = 20\%$ and $T = 200$. Figures 17a and 17b show the averaged solution and the pixel-wise standard deviation of the inferred reflectivities, respectively.

[Figure 17 about here.]

Since this real dataset is an out-of-distribution sample, we notice relatively high model uncertainty in all model regions. The estimator can also express an increased uncertainty in the deeper parts of the studied area, which corresponds to partially imaged reflectors and less resolvable features in the input data due to amplitude loss. Ultimately, this result allows us to flag unreliably reconstructed reflectors and quantify seismic structural uncertainty. In addition, due to the data consistency property of the deep decomposition estimator, multiple realizations from the deep null space network can be used to stochastically explore the null space of the forward operator and obtain a family of solutions that honour the data. However, the uncertainty maps will be partially biased since the approximated posterior is still dependent on the training dataset used to extract the prior information to solve the inverse problem and other hyper-parameters such as the dropout rate. Changing these variables could significantly modify the calculated uncertainty.

DISCUSSION AND FUTURE WORK

Some points deserve special consideration when using the proposed technique to deconvolve noisy seismograms. First, since we adopt the convolutional model for the seismogram, it is required that initial processing recovers amplitudes and removes multiples as well as possible. Surface-consistent deconvolution should also be implemented, allowing us to assume an invariant wavelet across traces.

Due to the necessity of constructing the forward operator from the test data, deconvolution via deep decomposition can be classified as an acquisition-restricted learning method. In other words, because of the dependency on the forward operator and its null space components, the deep estimator needs to be trained every time the signature of the estimated source wavelet changes. This does not represent a disadvantage in practice since SVD is only calculated once before training, which renders the algorithm’s training phase relatively cheap, depending on the number of training samples. Transfer learning could potentially ease the computational burden of the training stage, but the success of such an approach needs further investigation and escapes the scope of this paper. Likewise, the direct computation of the projection operators is intractable for large-scale problems. Iterative solvers such as the conjugate gradient method can help calculate approximations to these projections, for example, by solving the linear system

$$P_N(\mathbf{x}) \approx \arg \min_{\mathbf{m}} \|\mathbf{m} - \mathbf{x}\|_2^2 \quad \text{s.t. } \mathbf{Lm} = 0, \quad (23)$$

with \mathbf{x} as an initial guess model. Another option is to train a neural network to approximate the projection operators as proposed in Kuo et al. (2022).

l_p norms can be easily incorporated into the proposed training loss to investigate the use of explicit sparsity-promoting norms in combination with the deep decomposition deconvolution operator to measure the resolving power in the context of sparse reflectivity inversion. While we only consider white Gaussian noise, significant non-Gaussian noise may appear as low-rank components, affecting the proposed method’s performance. In this case, the denoising element can be adapted to give \mathbf{F}_{θ_1} a broader range to annihilate non-Gaussian and correlated noise.

Post-processing regularization via null space filtering is not new to the geophysics community. Deal and Nolet (1996) proposed a similar method, coined null space shuttles, in which desired features are imposed on an initial inverted solution by applying non-linear filters in the null space to preserve the observed data. Variations of this method include successful applications to tomography (de Wit et al., 2012; Osypov et al., 2013) and full waveform inversion (Keating and Innanen, 2021). The most significant differences between the method of Deal and Nolet (1996) and the proposed approach relate to the treatment of the filter design. Compared to the handcrafted filtering in the null space shuttles method, we make no explicit a priori assumptions about the properties of the solution. Instead, we use only reliable data in combination with the additional knowledge learned by the network from the training samples. Thus, deconvolution via deep decomposition can be regarded as a data-driven version of the null space shuttle approach.

In the proposed method, our constraints are based on knowledge of the physical process embedded in the forward operator, so we dispense the use of wells to prevent adding bias at an early stage (Li et al., 2022), although having a single well control point might be helpful for wavelet extraction. In the second example, we use well data for quality control but not in the inversion process itself. Once high-resolution seismic is obtained using the suggested

approach, conventional inversion employing well-log data can be performed. Downton et al. (2020) show that using a small number of well logs and a physics-based data augmentation technique can create multiple realizations of area-specific data labels. Future research should include additional well-log data information in the estimator training process to guiding the reconstruction of the missing frequencies.

One characteristic of the proposed deconvolution method is that it is possible to estimate the prediction uncertainty of the inversion by performing several forward passes using different weight seeds via MC dropout. While easy to train and compute, MC dropout is a simplistic approximation of Bayesian variational inference. In this regard, future work should explore reflectivity inversion via deep decomposition integrating uncertainty quantification approaches that may be more efficient and robust to out-of-distribution and noisy data. For example, recent research (Siahkoohi et al., 2022) advocates for invertible neural networks leveraging (conditional) normalizing flows to parameterize surrogate conditional distributions and enable efficient variational inference. As in all uncertainty quantification techniques, accuracy and computational efficiency have a trade-off. Explaining this trade-off for the presented reflectivity inversion problem is a question of ongoing research. Future work should also adapt the proposed framework to tolerate wavelet phase inaccuracies, extensions to 3D and a forward modelling operator that considers a more complex physics framework.

CONCLUSIONS

In this work, we investigated the use of regularizing networks via the deep decomposition approach to deal with the always-desirable goal of estimating full-band reflectivity from band-limited seismic data. We illustrate how learned null space regularization adds reason-

able estimates from the null space, improving classic regularization solutions. Additionally, we combined the deep decomposition learning method with TSVD, which helps produce clean inputs for the efficient training of the null space network. As a result, we produce an approximate high-resolution deconvolution operator for a specific and predetermined forward operator. Despite this limitation, the enhancement in resolution caused by the null space network formulation may be key to achieving widespread adoption of deep learning in full-band reflectivity estimation. We demonstrate the method’s effectiveness in multiple datasets and obtained reflectivity sections with no spurious artifacts and good lateral continuity. The inversion of 2D field datasets illustrates that the algorithm works in the presence of noise and is practical. An important research direction is identifying ways of efficiently integrating deep null space regularization with more significant problems where the direct computation of the orthogonal projections is prohibitively expensive.

REFERENCES

- Abdar, M., F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, V. Makarenkov, and S. Nahavandi, 2021, A review of uncertainty quantification in deep learning: Techniques, applications and challenges: *Information Fusion*, **76**, 243–297.
- Araya-Polo, M., J. Jennings, A. Adler, and T. Dahlke, 2018, Deep-learning tomography: The Leading Edge, **37**, 58–66.
- Berkhout, A. J., 1977, Least-squares inverse filtering and wavelet deconvolution: *Geophysics*, **42**, 1369–1383.
- Chai, X., G. Tang, K. Lin, Z. Yan, H. Gu, R. Peng, X. Sun, and W. Cao, 2021, Deep learning for multitrace sparse-spike deconvolution: *Geophysics*, **86**, V207–V218.
- Chen, D., and M. E. Davies, 2020, Deep decomposition learning for inverse imaging problems: Presented at the Proceedings of the European Conference on Computer Vision (ECCV).
- Chen, H., M. Sacchi, H. H. Lari, J. Gao, and X. Jiang, 2023, Nonstationary seismic reflectivity inversion based on prior-engaged semi-supervised deep learning method: *Geophysics*, **0**, 1–72.
- Chen, H., M. D. Sacchi, and J. Gao, 2022, Seismic reflectivity inversion via a regularized deep image prior: Second International Meeting for Applied Geoscience & Energy, 2586–2590.
- Chopra, S., J. Castagna, and O. Portniaguine, 2006, Seismic resolution and thin-bed reflectivity inversion: *CSEG recorder*, **31**, 19–25.
- Chopra, S., J. Castagna, and Y. Xu, 2009, Thin-bed reflectivity inversion and some applications: *First Break*, **27**, 55–62.

- de Wit, R. W. L., J. Trampert, and R. D. van der Hilst, 2012, Toward quantifying uncertainty in travel time tomography using the null-space shuttle: *Journal of Geophysical Research: Solid Earth*, **117**.
- Deal, M. M., and G. Nolet, 1996, Nullspace shuttles: *Geophysical Journal International*, **124**, 372–380.
- Debeye, H. W. J., and P. Van Riel, 1990, L_p-norm deconvolution: *Geophysical Prospecting*, **38**, 381–403.
- Dhara, A., and M. K. Sen, 2022, Physics-guided deep autoencoder to overcome the need for a starting model in full-waveform inversion: *The Leading Edge*, **41**, 375–381.
- Downton, J. E., O. Collet, D. P. Hampson, and T. Colwell, 2020, Theory-guided data science-based reservoir prediction of a North Sea oil field: *The Leading Edge*, **39**, 742–750.
- Gal, Y., and Z. Ghahramani, 2016, Dropout as a bayesian approximation: Representing model uncertainty in deep learning: *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, JMLR.org, 1050–1059.
- Gavish, M., and D. L. Donoho, 2014, The optimal hard threshold for singular values is $4/\sqrt{3}$: *IEEE Transactions on Information Theory*, **60**, 5040–5053.
- Gholami, A., and M. D. Sacchi, 2012, A fast and automatic sparse deconvolution in the presence of outliers: *IEEE Transactions on Geoscience and Remote Sensing*, **50**, 4105–4116.
- , 2013, Fast 3D blind seismic deconvolution via constrained total variation and GCV: *SIAM Journal on Imaging Sciences*, **6**, 2350–2369.
- Hansen, P. C., 2010, Discrete inverse problems: Society for Industrial and Applied Mathematics.

- Idier, J., and Y. Goussard, 1993, Multichannel seismic deconvolution: IEEE Transactions on Geoscience and Remote Sensing, **31**, 961–979.
- Kaarsen, K. F., and T. Taxt, 1998, Multichannel blind deconvolution of seismic signals: Geophysics, **63**, 2093–2107.
- Kaur, H., N. Pham, and S. Fomel, 2020, Improving the resolution of migrated images by approximating the inverse hessian using deep learning: Geophysics, **85**, WA173–WA183.
- Kazemi, N., and M. D. Sacchi, 2013, Sparse multichannel blind deconvolution: Geophysics, **79**, V143–V152.
- Keating, S. D., and K. A. Innanen, 2021, Null-space shuttles for targeted uncertainty analysis in full-waveform inversion: Geophysics, **86**, R63–R76.
- Kingma, D. P., and J. Ba, 2014, Adam: A method for stochastic optimization. (cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015).
- Kong, F., F. Picetti, V. Lipari, P. Bestagini, X. Tang, and S. Tubaro, 2022, Deep prior-based unsupervised reconstruction of irregularly sampled seismic data: IEEE Geoscience and Remote Sensing Letters, **19**, 1–5.
- Kuo, J., J. Granstedt, U. Villa, and M. A. Anastasio, 2022, Computing a projection operator onto the null space of a linear imaging operator: tutorial: Journal of the Optical Society of America A, **39**, 470.
- LeCun, Y., Y. Bengio, and G. Hinton, 2015, Deep learning: Nature, **521**, 436–444.
- Levin, S. A., 1989, Surface-consistent deconvolution: Geophysics, **54**, 1123–1133.
- Levy, S., and P. K. Fullagar, 1981, Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution: Geophysics, **46**, 1235–1243.
- Li, C., Q. Zhu, and B. Roy, 2022, Seismic super resolution method for enhancing strati-

graphic interpretation: Second International Meeting for Applied Geoscience & Energy, Society of Exploration Geophysicists and American Association of Petroleum Geologists, 165–169.

Mandelli, S., F. Borra, V. Lipari, P. Bestagini, A. Sarti, and S. Tubaro, 2019, Seismic data interpolation through convolutional autoencoder: 2018 SEG International Exposition and Annual Meeting, SEG 2018, 4101–4105.

Oldenburg, D. W., T. Scheuer, and S. Levy, 1983, Recovery of the acoustic impedance from reflection seismograms.: *Geophysics*, **48**, 1318–1337.

Osypov, K., Y. Yang, A. Fournier, N. Ivanova, R. Bachrach, C. E. Yarman, Y. You, D. Nichols, and M. Woodward, 2013, Model-uncertainty quantification in seismic tomography: Method and applications: *Geophysical Prospecting*, **61**, 1114–1134.

Robinson, E., and S. Treitel, 1980, *Geophysical signal analysis*: Prentice-Hall. Prentice-Hall Foundations of Earth Science Series.

Ronneberger, O., P. Fischer, and T. Brox, 2015, U-net: Convolutional networks for biomedical image segmentation: CoRR, [abs/1505.04597](https://arxiv.org/abs/1505.04597).

Sacchi, M. D., 1997, Reweighting strategies in seismic deconvolution: *Geophysical Journal International*, **129**, 651–656.

Sacchi, M. D., D. R. Velis, and A. H. Cominguez, 1994, Minimum entropy deconvolution with frequency-domain constraints: *Geophysics*, **59**, 938–945.

Scales, J., and M. Smith, 1994, *Introductory geophysical inverse theory*: Samizdat Press. Center for Wave Phenomena.

Schwab, J., S. Antholzer, and M. Haltmeier, 2019, Deep null space learning for inverse problems: Convergence analysis and rates: *Inverse Problems*, **35**.

—, 2020, Big in Japan: Regularizing Networks for Solving Inverse Problems: *Journal of*

- Mathematical Imaging and Vision, **62**, 445–455.
- Siahkoohi, A., G. Rizzuti, R. Orozco, and F. J. Herrmann, 2022, Reliable amortized variational inference with physics-based latent distribution correction.
- Taylor, H. L., S. C. Banks, and J. F. McCoy, 1979, Deconvolution with the l1 norm: Geophysics, **44**, 39–52.
- Tenorio, L., 2001, Modeling non-gaussian reflectivities: Generalizing wiener-levinson deconvolution: Geophysics, **66**, 1913–1920.
- Tompson, J., R. Goroshin, A. Jain, Y. LeCun, and C. Bregler, 2015, Efficient object localization using convolutional networks: CVPR, IEEE Computer Society, 648–656.
- Torres, K., and M. Sacchi, 2022a, Least-squares reverse time migration via deep learning-based updating operators: Geophysics, **87**, S315–S333.
- Torres, K., and M. D. Sacchi, 2022b, Deep Null Space Regularization for Seismic Inverse Problems: 83rd EAGE Annual Conference & Exhibition, European Association of Geoscientists & Engineers, 1–5.
- Ulrych, T. J., D. R. Velis, and M. D. Sacchi, 1995, Wavelet estimation revisited: The Leading Edge, **14**, 1139–1143.
- Velis, D. R., 2008, Stochastic sparse-spike deconvolution: Geophysics, **73**.
- Walden, A. T., 1985, Non-gaussian reflectivity, entropy, and deconvolution: Geophysics, **50**, 2862–2888.
- White, R. E., and T. Hu, 1998, How accurate can a well tie be?: The Leading Edge, **17**, 1065–1071.
- Wiggins, R. A., 1978, Minimum entropy deconvolution: Geoexploration, **16**, 21–35.
- Wu, X., S. Yan, Z. Bi, S. Zhang, and H. Si, 2021, Deep learning for multi-dimensional seismic impedance inversion: Geophysics, **86**.

Yilmaz, O., 2001, Seismic data analysis, 2 ed ed.: Society Of Exploration Geophysicists, volume **10** of Investigations in Geophysics.

Yu, S., and J. Ma, 2021, Deep Learning for Geophysics: Current and Future Trends: Reviews of Geophysics, **59**.

Zhang, K., W. Zuo, Y. Chen, D. Meng, and L. Zhang, 2017, Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising: IEEE Transactions on Image Processing, **26**, 3142–3155.

Zhang, R., and J. Castagna, 2011, Seismic sparse-layer reflectivity inversion using basis pursuit decomposition: Geophysics, **76**, R147–R158.

Zhang, W., J. Gao, T. Yang, X. Jiang, and W. Sun, 2021, Least-squares reverse time migration using convolutional neural networks: Geophysics, **86**.

LIST OF FIGURES

1	Convolutional neural network architecture of the denoiser \mathbf{F}_{θ_1} with corresponding depth of 17. \mathbf{I} refers to the original image size. The number at the bottom of each convolutional layer indicates the number of channels. The convolutional kernels are of size 3×3 . ReLU and BN stand for <i>rectified linear unit</i> and <i>batch normalization</i> , respectively.	36
2	U-net based architecture of the null space neural network \mathbf{N}_{θ_2} . \mathbf{I} refers to the original image size. The number at the bottom of each convolutional layer indicates the number of channels. The convolutional kernels are of size 3×3 . Max-pooling has kernel size of 2×2	37
3	Single-channel deconvolution example. True reflectivity \mathbf{m} (black) superposed with (a) the noisy trace \mathbf{d}_ϵ , (b) the TSVD result, (c) the FISTA result, and (d) the deep null space regularization.	38
4	Comparison of (a) the amplitude spectrum of the true reflectivity model and the deep null space regularization solution, and (b) the amplitude spectrum of the individual range (orange) and null space (gray) retrieved components.	39
5	Multiple realizations of reflectivity models (left column) and noisy data (right column) used for training the estimator to perform 2D thin-bed reflectivity inversion.	40
6	(a) Extracted source wavelet from the Alberta foothills seismic data. (b) Amplitude spectrum.	41
7	Alberta foothills results: (a) field data example portraying the input band-limited data \mathbf{d}_ϵ , (b) full-band reflectivity inverted by the proposed method, (c) classical thin-bed reflectivity inversion via the TSI algorithm, (d) ResUnet result. Panels (e), (f) and (g) show the data residual $\mathbf{d}_\epsilon - \mathbf{L}\mathbf{m}^*$, taking \mathbf{m}^* as the results in (a), (b) and (c), respectively, and the Pearson correlation coefficient (mean \pm standard deviation) of the traces with respect to the input data on top.	42
8	Representative trace form the Alberta foothills dataset showing the Pearson correlation coefficient (bottom) between the re-convolved results and the input data.	43
9	Individual components of the deep decomposition solution for the Alberta foothills dataset: (a) the band-limited solution obtained via TSVD ($\mathbf{L}_k^\dagger \mathbf{d}_\epsilon$), (b) residual noise leaking into the solution through the pseudo-inverse ($P_R \circ \mathbf{F}_{\theta_1} \circ \mathbf{L}^\dagger \mathbf{d}_\epsilon$), and (c) the predicted null space component $P_N \circ \mathbf{N}_{\theta_2} \circ (\mathbf{L}^\dagger \mathbf{d}_\epsilon + P_R \circ \mathbf{F}_{\theta_1} \circ \mathbf{L}^\dagger \mathbf{d}_\epsilon)$	44
10	Normalized average power spectral density for the Alberta foothills data example and three deconvolution techniques for real data thin-bed reflectivity estimation: deep decomposition learning solution $\Lambda(\mathbf{d}_\epsilon; k, \theta_1, \theta_2)$, TSI solution (Chopra et al., 2006), and ResUnet.	45
11	(a) Extracted source wavelet from the Penobscot seismic data. (b) Amplitude spectrum.	46

12	(a) Field data example from the Penobscot dataset. (b) Full-band reflectivity inverted by the proposed method.	47
13	Normalized average power spectral density for the Penobscot data example and the decomposition learning solution $\Lambda(\mathbf{d}_\epsilon; k, \theta_1, \theta_2)$	48
14	Close-up view of the Penobscot dataset. Results showing the well-derived synthetic trace (lef), the input measurements (middle) and the predicted result (right) for the trace where the well log data is located. Interpreted horizons are shown for reference.	49
15	Validation panel for the Penobscot dataset. (a) well-log derived reflectivity in time domain. (b) Simulated data from the well-log derived reflectivity in (a) (in blue) and simulated data from the predicted reflectivity (in red). The simulations are the result of convolving each reflectivity profile with a 60 Hz Ricker wavelet.	50
16	Individual components of the deep decomposition solution for the Penobscot dataset: (a) the band-limited solution obtained via TSVD ($\mathbf{L}_k^\dagger \mathbf{d}_\epsilon$), (b) residual noise leaking into the solution through the pseudo-inverse ($P_R \circ \mathbf{F}_{\theta_1} \circ \mathbf{L}^\dagger \mathbf{d}_\epsilon$), and (c) the predicted null space component $P_N \circ \mathbf{N}_{\theta_2} \circ (\mathbf{L}^\dagger \mathbf{d}_\epsilon + P_R \circ \mathbf{F}_{\theta_1} \circ \mathbf{L}^\dagger \mathbf{d}_\epsilon)$	51
17	(a) Estimated mean after MC dropout running 200 forward passes through the deep estimator for the Penobscot field data. (b) Point-wise standard deviation among samples.	52

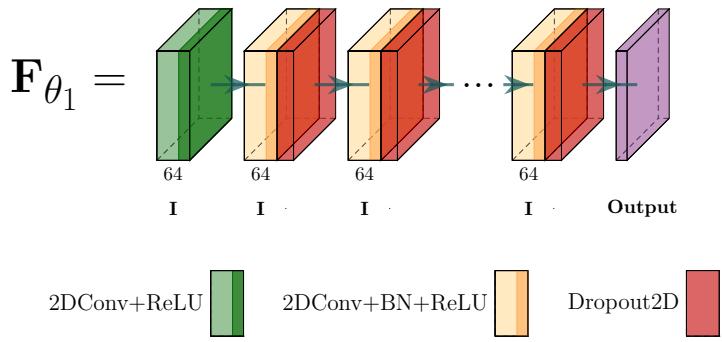


Figure 1: Convolutional neural network architecture of the denoiser \mathbf{F}_{θ_1} with corresponding depth of 17. \mathbf{I} refers to the original image size. The number at the bottom of each convolutional layer indicates the number of channels. The convolutional kernels are of size 3×3 . ReLU and BN stand for *rectified linear unit* and *batch normalization*, respectively.

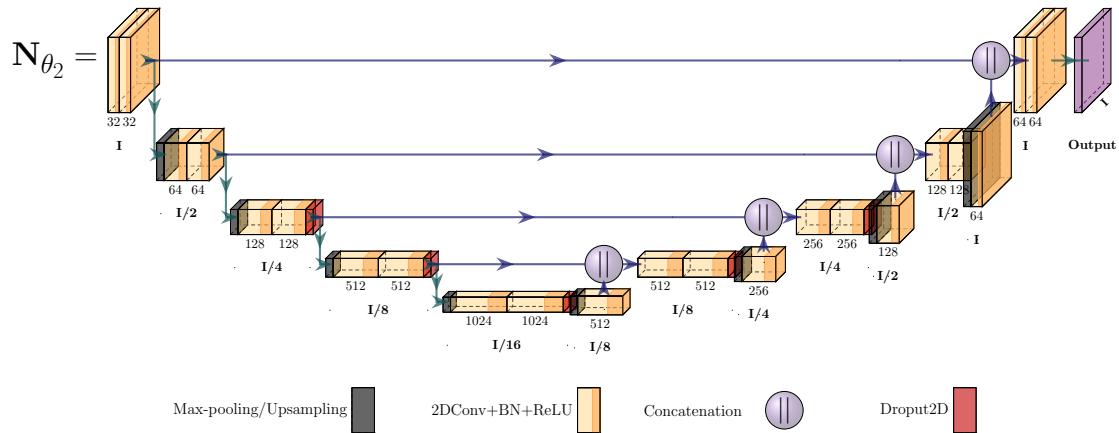


Figure 2: U-net based architecture of the null space neural network \mathbf{N}_{θ_2} . \mathbf{I} refers to the original image size. The number at the bottom of each convolutional layer indicates the number of channels. The convolutional kernels are of size 3×3 . Max-pooling has kernel size of 2×2 .

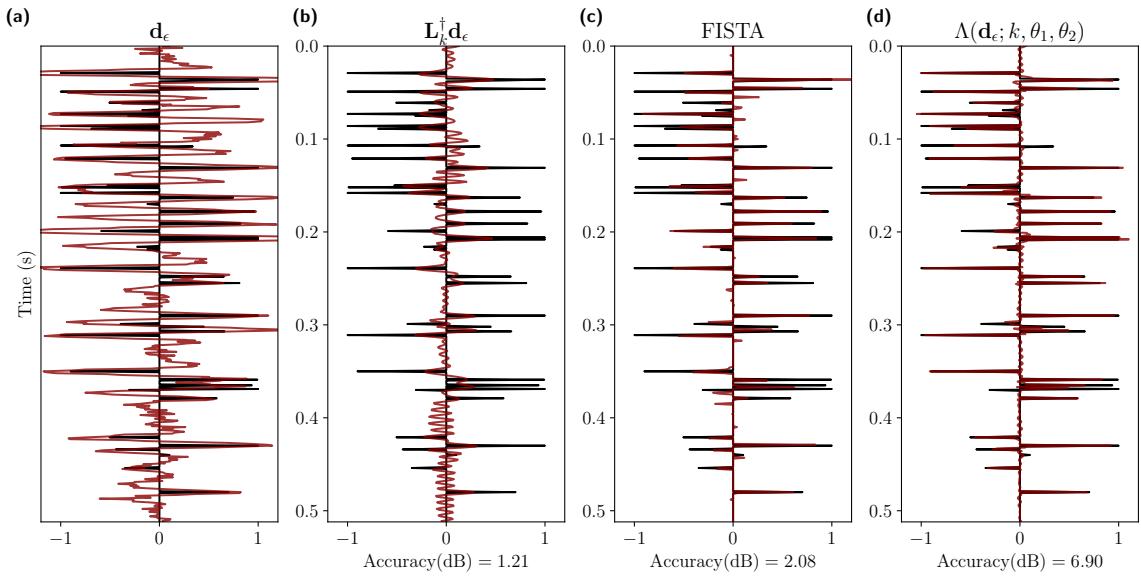


Figure 3: Single-channel deconvolution example. True reflectivity \mathbf{m} (black) superposed with (a) the noisy trace \mathbf{d}_ϵ , (b) the TSVD result, (c) the FISTA result, and (d) the deep null space regularization.

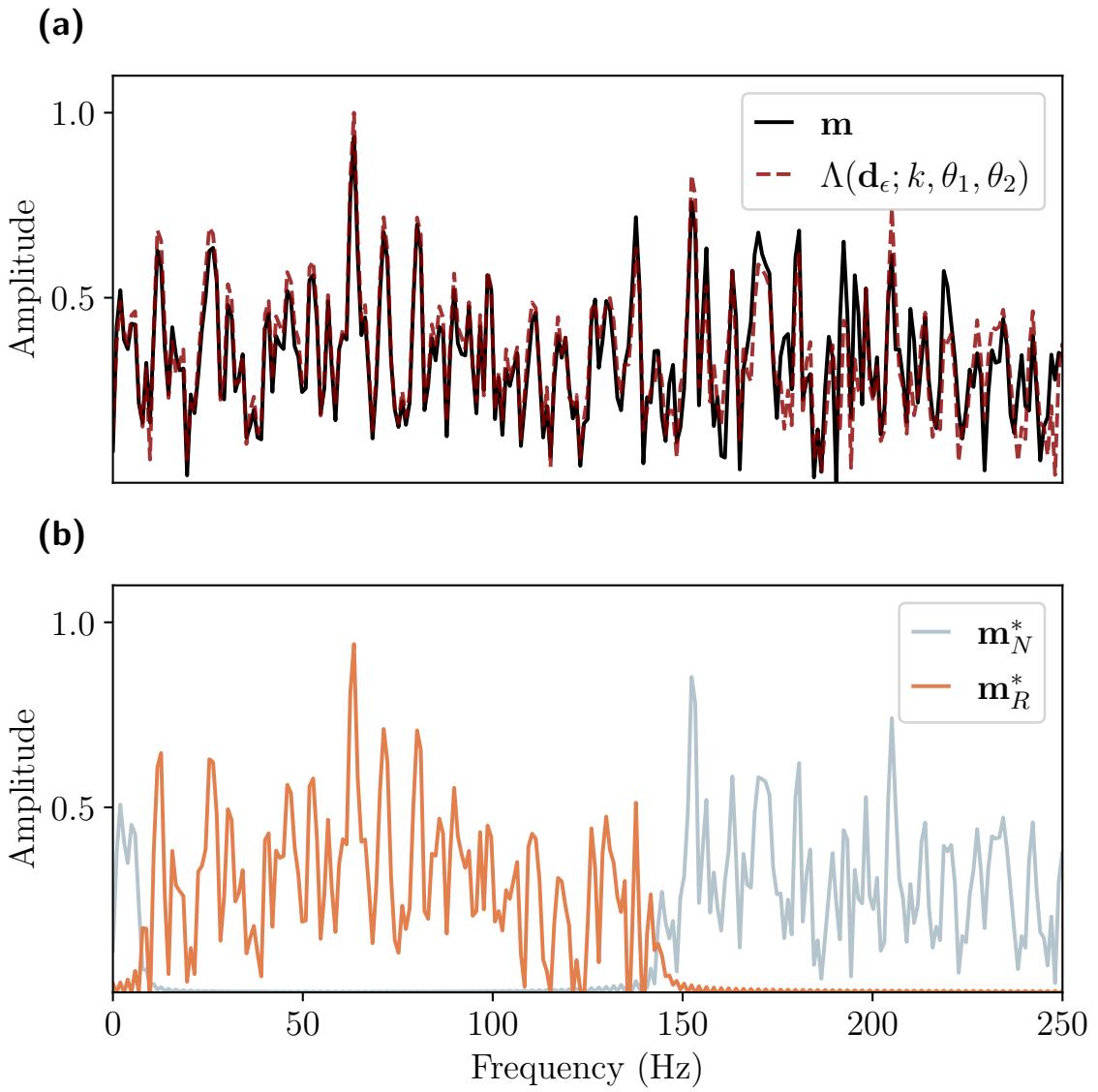


Figure 4: Comparison of (a) the amplitude spectrum of the true reflectivity model and the deep null space regularization solution, and (b) the amplitude spectrum of the individual range (orange) and null space (gray) retrieved components.

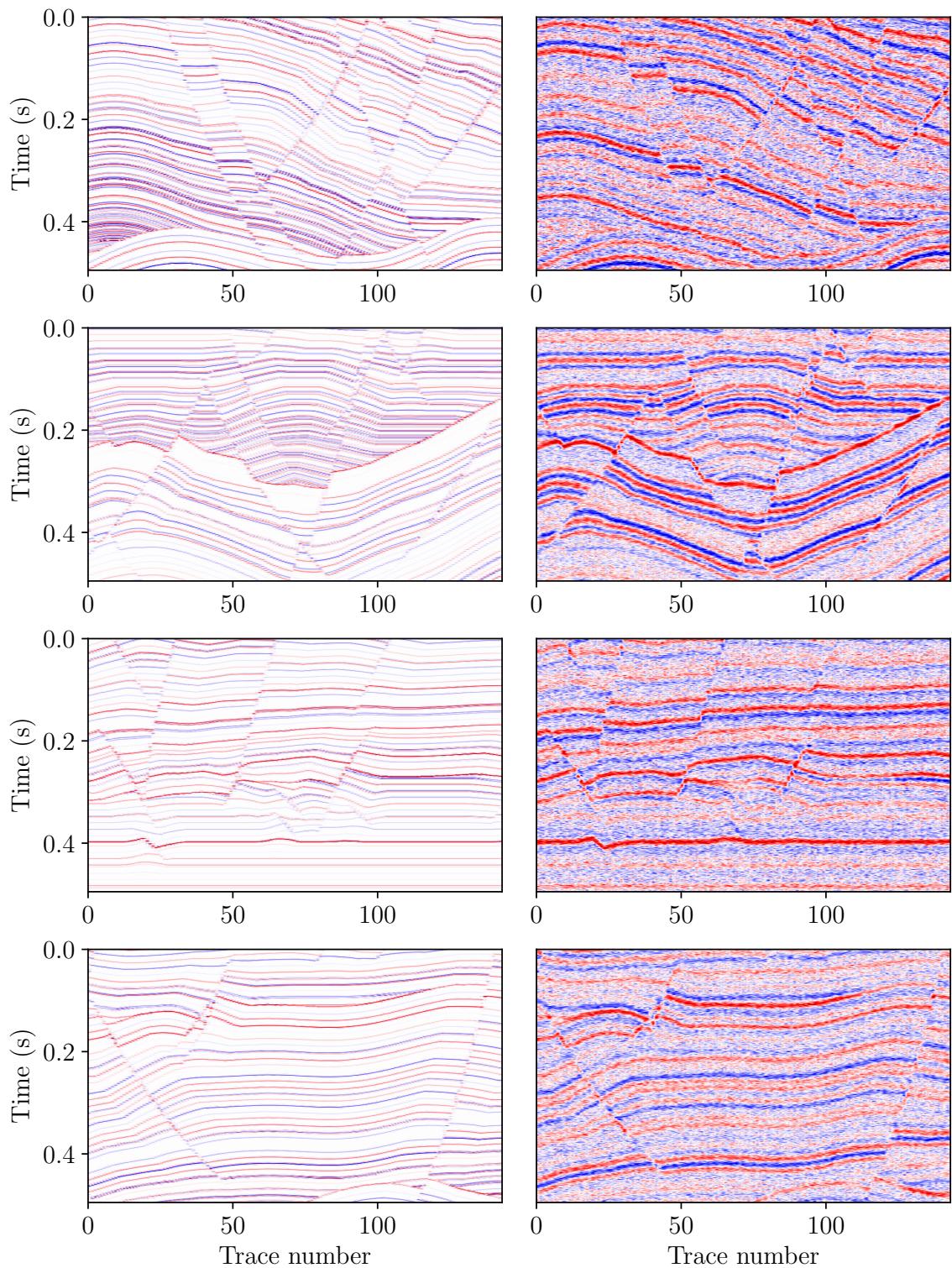


Figure 5: Multiple realizations of reflectivity models (left column) and noisy data (right column) used for training the estimator to perform 2D thin-bed reflectivity inversion.

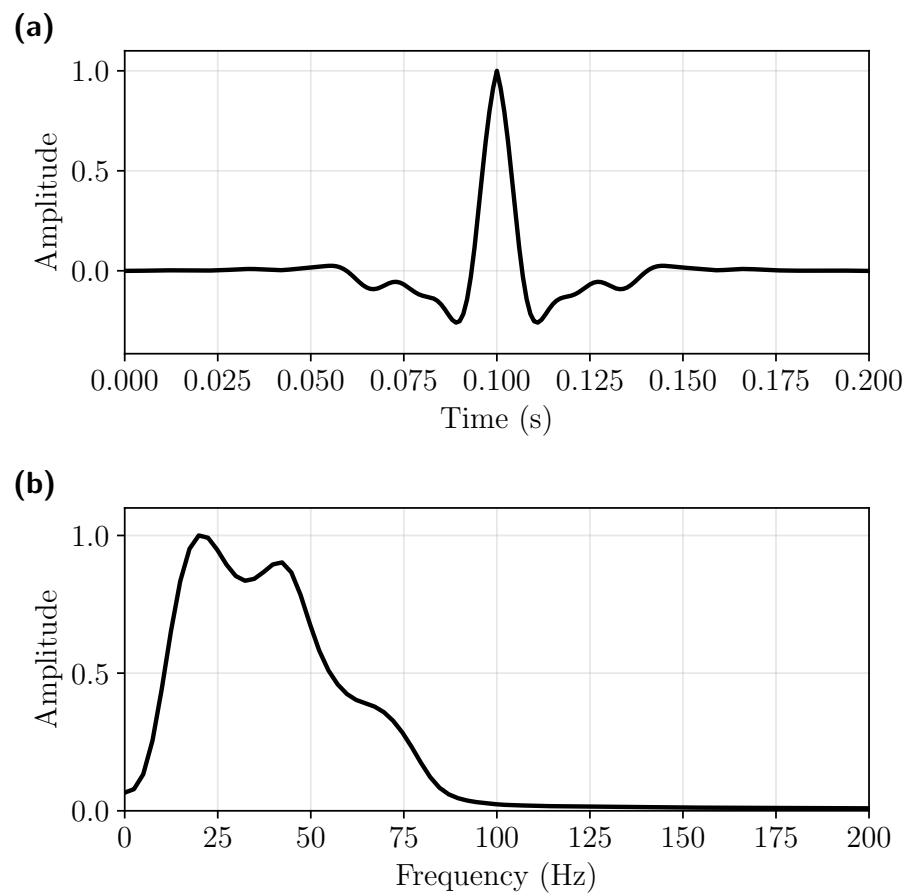


Figure 6: (a) Extracted source wavelet from the Alberta foothills seismic data. (b) Amplitude spectrum.

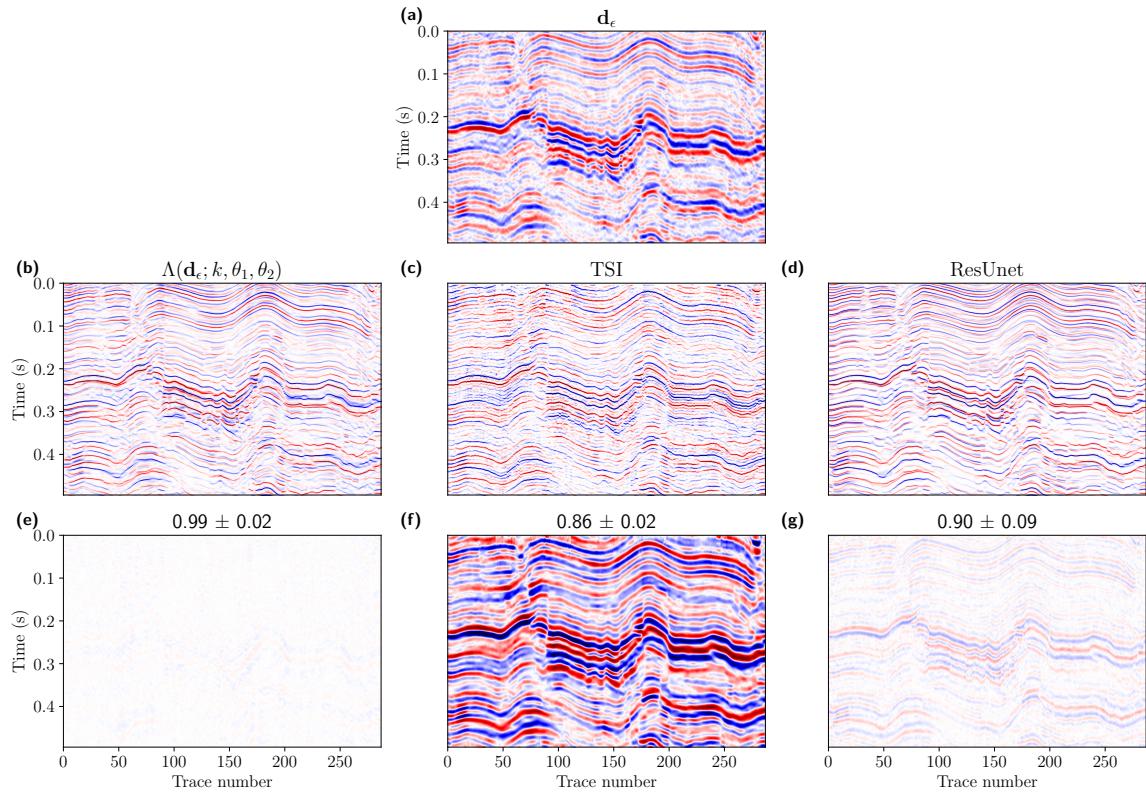


Figure 7: Alberta foothills results: (a) field data example portraying the input band-limited data \mathbf{d}_ϵ , (b) full-band reflectivity inverted by the proposed method, (c) classical thin-bed reflectivity inversion via the TSI algorithm, (d) ResUnet result. Panels (e), (f) and (g) show the data residual $\mathbf{d}_\epsilon - \mathbf{L}\mathbf{m}^*$, taking \mathbf{m}^* as the results in (a), (b) and (c), respectively, and the Pearson correlation coefficient (mean \pm standard deviation) of the traces with respect to the input data on top.

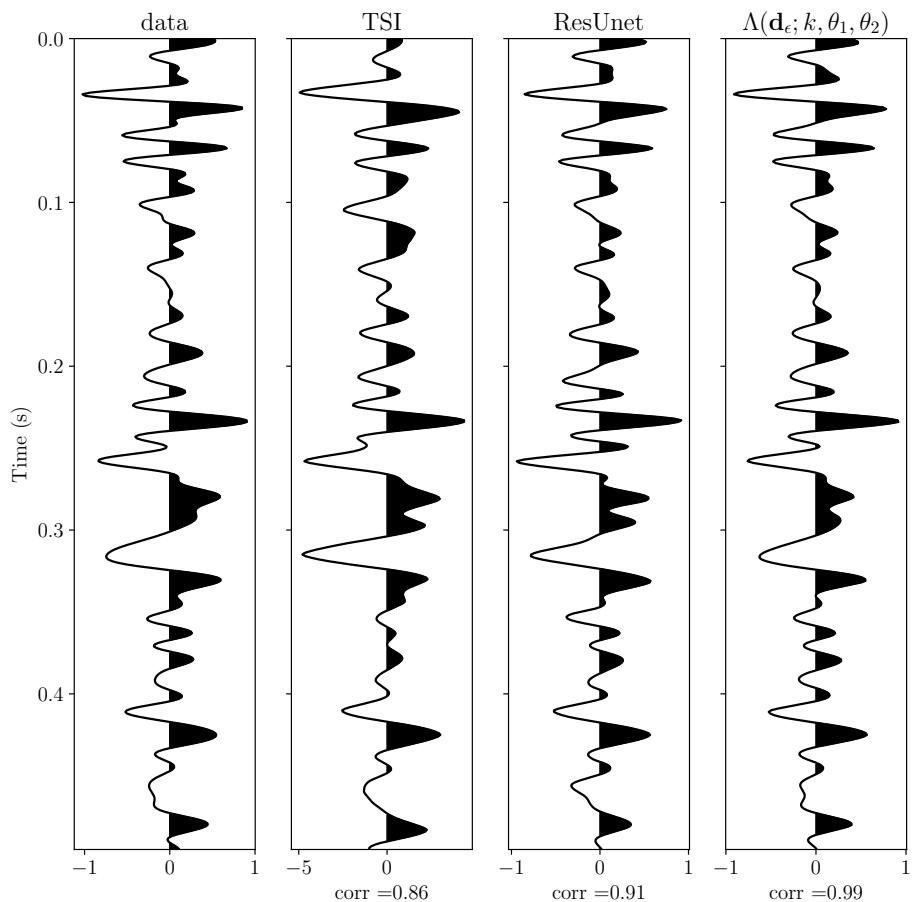


Figure 8: Representative trace from the Alberta foothills dataset showing the Pearson correlation coefficient (bottom) between the re-convolved results and the input data.

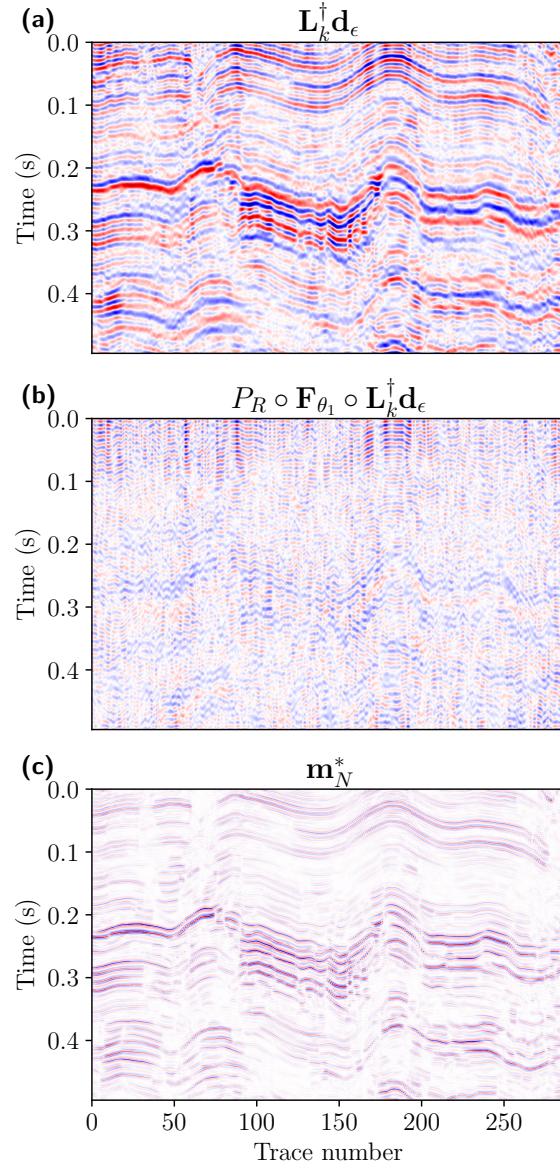


Figure 9: Individual components of the deep decomposition solution for the Alberta foothills dataset: (a) the band-limited solution obtained via TSVD ($\mathbf{L}_k^\dagger \mathbf{d}_\epsilon$), (b) residual noise leaking into the solution through the pseudo-inverse ($P_R \circ \mathbf{F}_{\theta_1} \circ \mathbf{L}_k^\dagger \mathbf{d}_\epsilon$), and (c) the predicted null space component $P_N \circ \mathbf{N}_{\theta_2} \circ (\mathbf{L}^\dagger \mathbf{d}_\epsilon + P_R \circ \mathbf{F}_{\theta_1} \circ \mathbf{L}^\dagger \mathbf{d}_\epsilon)$.

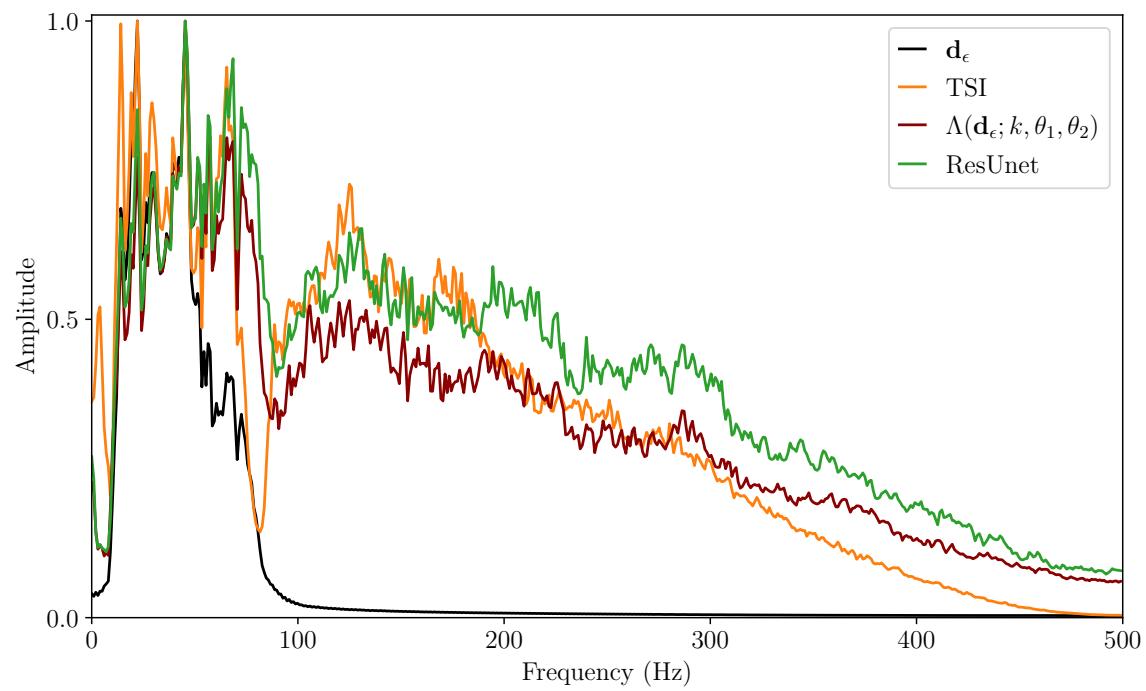


Figure 10: Normalized average power spectral density for the Aberta foothills data example and three deconvolution techniques for real data thin-bed reflectivity estimation: deep decomposition learning solution $\Lambda(\mathbf{d}_\epsilon; k, \theta_1, \theta_2)$, TSI solution (Chopra et al., 2006), and ResUnet.

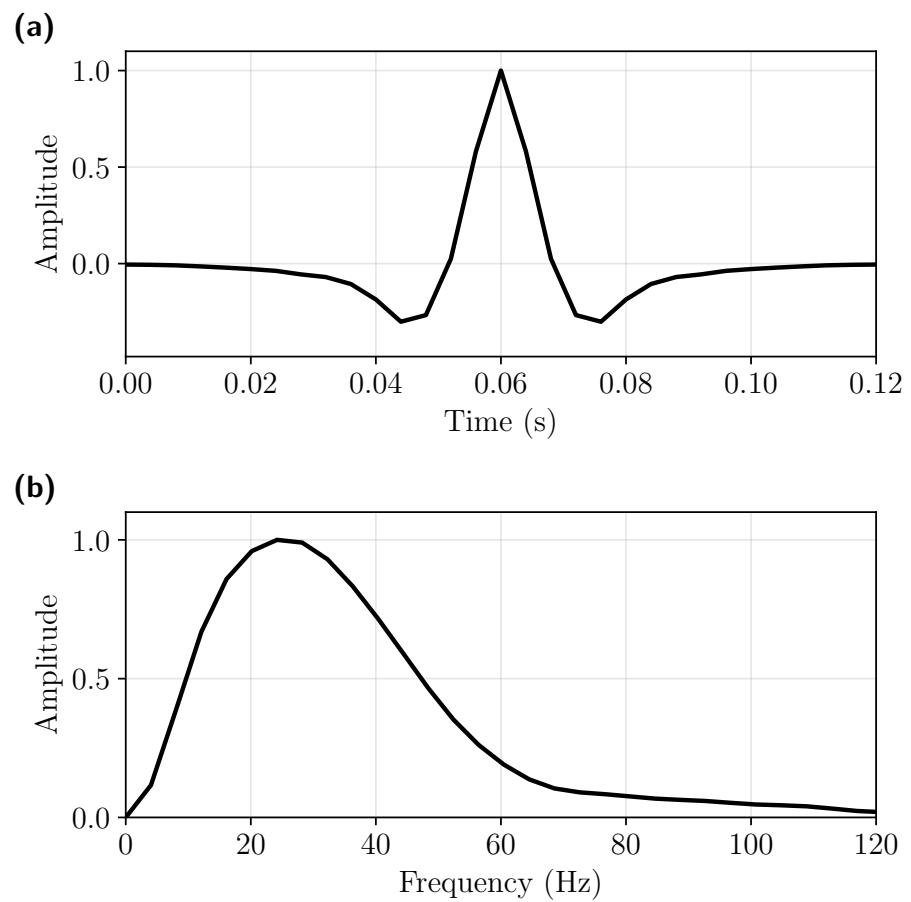


Figure 11: (a) Extracted source wavelet from the Penobscot seismic data. (b) Amplitude spectrum.

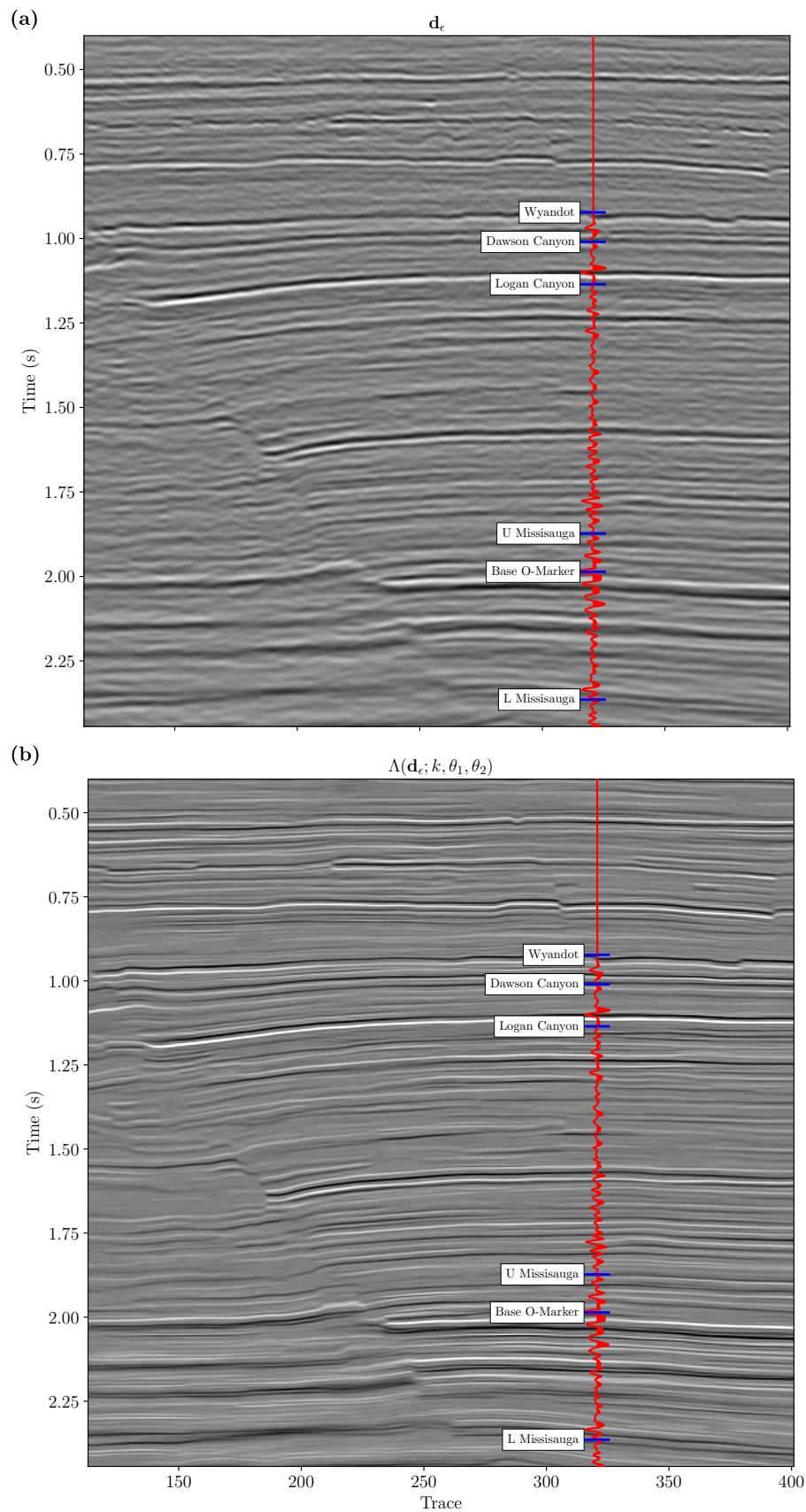


Figure 12: (a) Field data example from the Penobscot dataset. (b) Full-band reflectivity inverted by the proposed method.

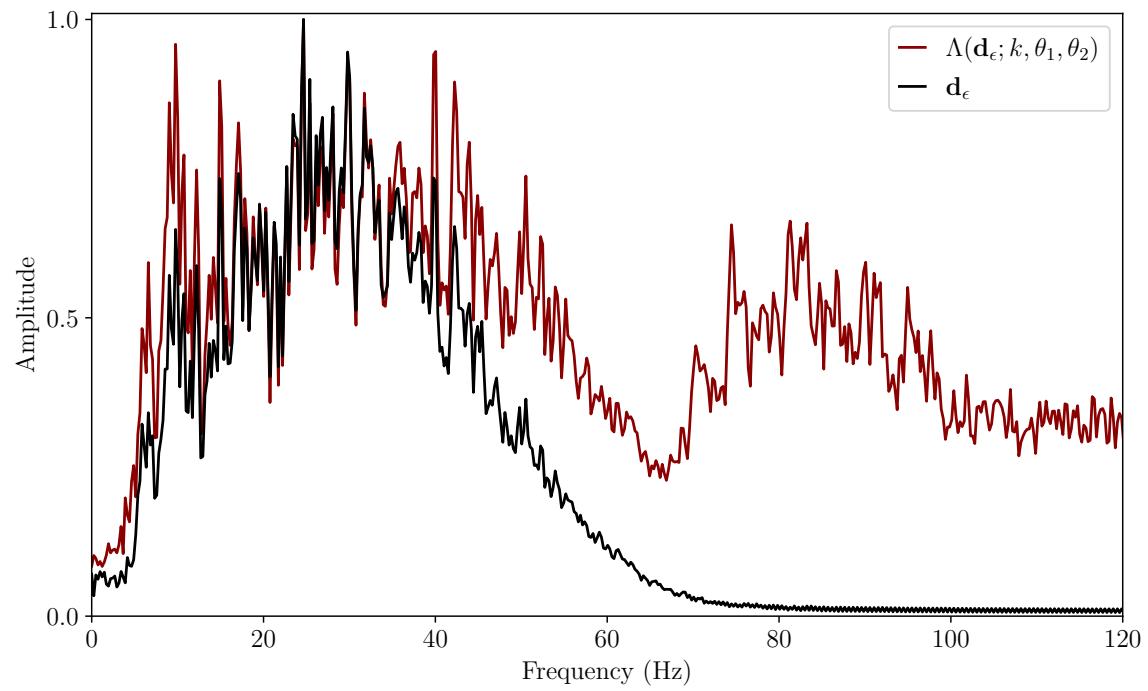


Figure 13: Normalized average power spectral density for the Penobscot data example and the decomposition learning solution $\Lambda(\mathbf{d}_\epsilon; k, \theta_1, \theta_2)$.

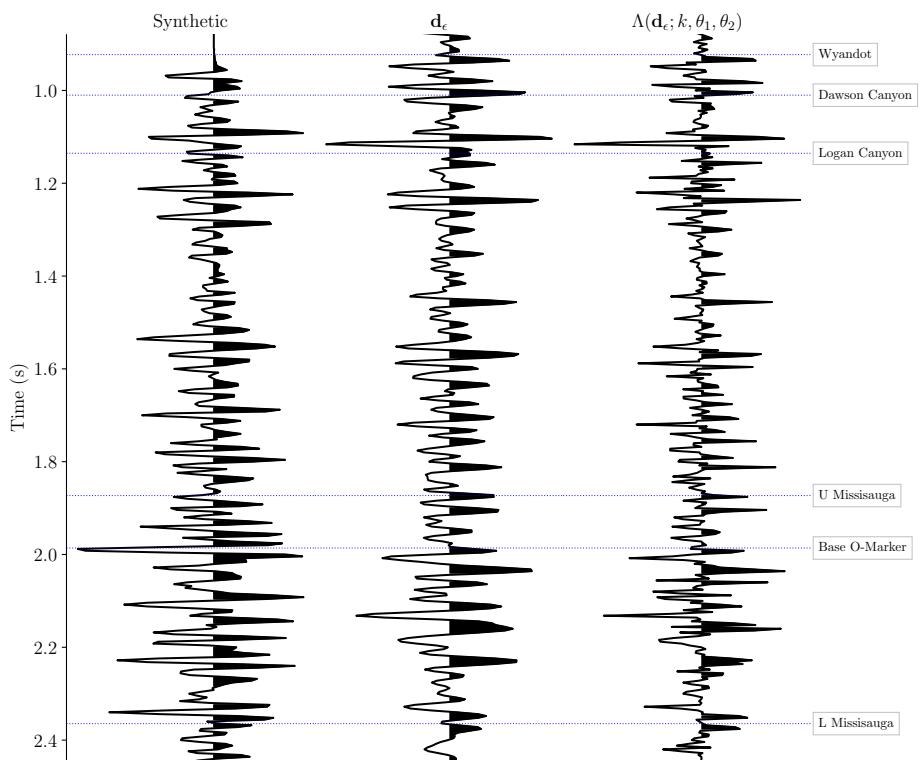


Figure 14: Close-up view of the Penobscot dataset. Results showing the well-derived synthetic trace (lef), the input measurements (middle) and the predicted result (right) for the trace where the well log data is located. Interpreted horizons are shown for reference.

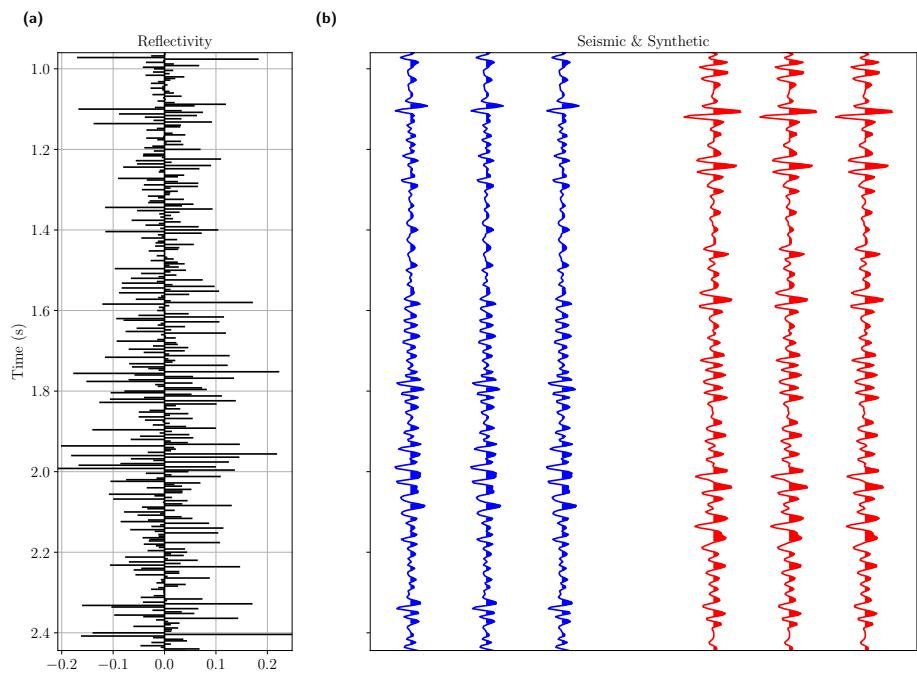


Figure 15: Validation panel for the Penobscot dataset. (a) well-log derived reflectivity in time domain. (b) Simulated data from the well-log derived reflectivity in (a) (in blue) and simulated data from the predicted reflectivity (in red). The simulations are the result of convolving each reflectivity profile with a 60 Hz Ricker wavelet.

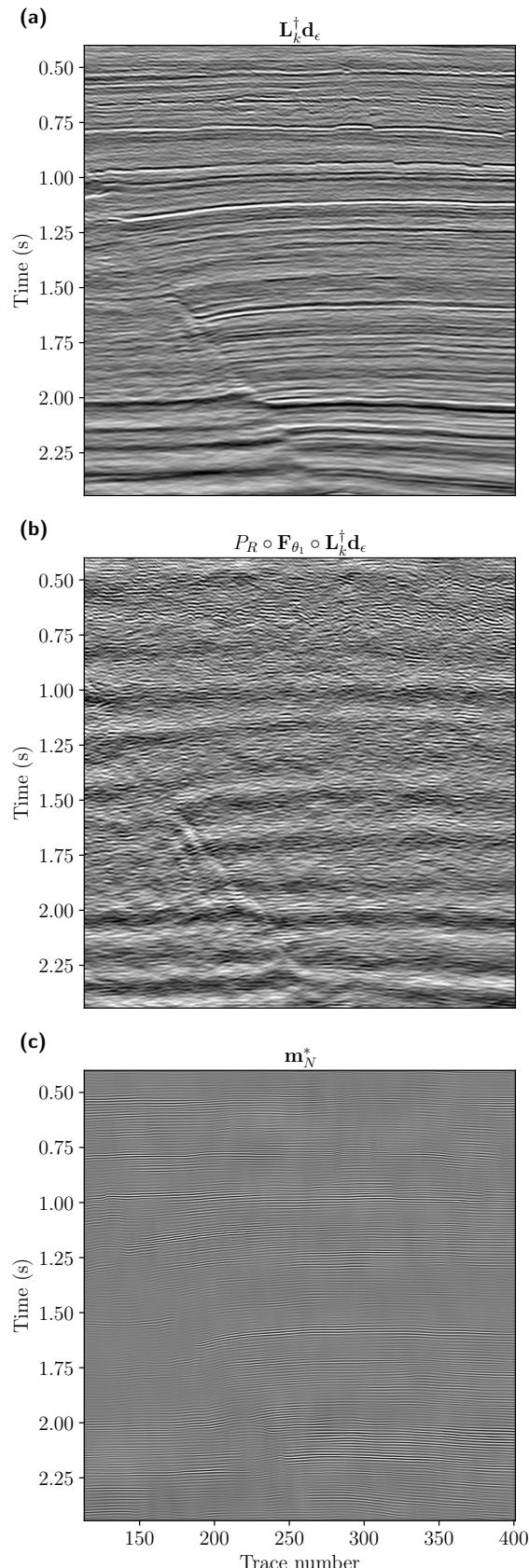


Figure 16: Individual components of the deep decomposition solution for the Penobscot dataset: (a) the band-limited solution obtained via TSVD ($L_k^\dagger d_\epsilon$), (b) residual noise leaking into the solution through the pseudo-inverse ($P_R \circ F_{\theta_1} \circ L_k^\dagger d_\epsilon$), and (c) the predicted null space component $P_N \circ N_{\theta_2} \circ (L^\dagger d_\epsilon + P_R \circ F_{\theta_1} \circ L_k^\dagger d_\epsilon)$.

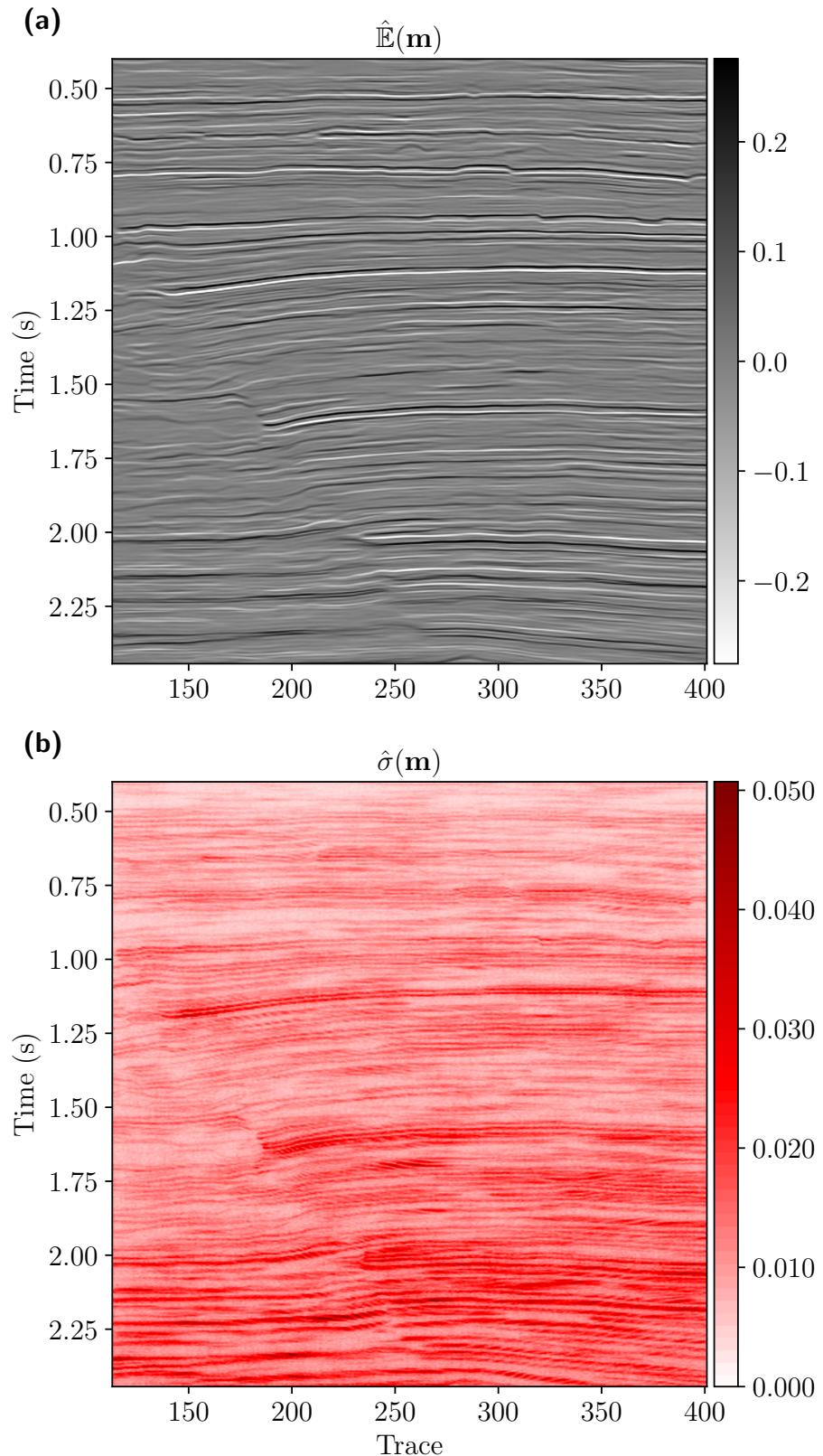


Figure 17: (a) Estimated mean after MC dropout running 200 forward passes through the deep estimator for the Penobscot field data. (b) Point-wise standard deviation among samples.