Time-lapse data matching using neural networks with multiple reflections

Shang Huang and Daniel Trad

ABSTRACT

Time-lapse seismic is a complex problem in reservoir monitoring because near-surface noise, poor subsurface illumination, and the inherent weak amplitude of reservoir changes affect the quality of the interpretation. Because of its importance, geophysicists have tried many approaches to solve these problems. In this chapter, I apply deep learning methods to address these challenges. Firstly, a stacked long short-term memory (SD-LSTM) neural network adapts the near-surface baseline data to the near-surface monitor data. This assumes that differences from the near-surface are not due to changes in the reservoir but differences in the seismic experiment (acquisition and processing). A U-Net is then followed to work on differences between monitor and baseline images to suppress noise on a large scale. Furthermore, the energy from surface multiples is added during migration and shot record generation in the forward modelling step to increase subsurface illumination. A double-difference method is applied to the predicted and observed data to give a final difference. The results show that the SD-LSTM can anticipate and mitigate noise in the monitor data. The final difference between baseline and monitor models has suppressed significant noise after combining SD-LSTM, U-Net and surface multiples. The proposed method is also tested in a field dataset, DAS VSP data from the CaMI FRS project, with extended bidirectional SD-LSTM and convolutional neural networks (CNN). The output provides meaningful information and prediction for CO_2 injection migration within a target area, which matches the CO_2 injection plan.

INTRODUCTION

Time-lapse seismic monitoring, which acquires seismic data at time intervals over the same site to obtain reservoir variations, has contributed to detecting subsurface physical properties and reservoir behaviour in recent years (Wang, 1997; Koster et al., 2000; Pennington, 2000; Lumley, 2001; Arts et al., 2003; Isaac and Lawton, 2006, 2014; Chadwick et al., 2010; Wang and Morozov, 2020; Henley and Lawton, 2021). Particularly, time-lapse becomes essential in carbon capture and storage reservoirs to monitor the movement and possible leaks of CO_2, which is necessary for the safety and the success of the process and tracking oil and gas displacement effects during long-term change. In another context, time-lapse is a crucial technique for enhanced oil recovery projects to improve fluid injection efficiency.

Time-lapse seismic monitoring projects face many challenges. Seismic imaging of time-lapse changes is difficult because of their weak amplitudes, often covered by noise differences due to near-surface complexities. Variations in the acquisition system and geometry create complex wavefield differences that do not cancel by subtraction. The weak effect of reservoir changes on seismic amplitudes is overwhelmed by false anomalies. Poor subsurface illumination accentuates these problems since 4D seismic captures only parts of the reservoir changes. Some geophysical approaches (Rickett and Lumley, 2001; Ayeni and Biondi, 2010; Zhang et al., 2013; Bergmann et al., 2014; Wapenaar and Van Ijsseldijk, 2021; Fu and Innanen, 2022) try to solve the challenges above. In addition to geophysical methods, deep neural networks have become useful tools to deal with these problems (Yuan et al., 2020; Zhong et al., 2020; Li et al., 2021; Hussein et al., 2021; Alali et al., 2022; Li and Alkhalifah, 2022). Long short-term memory (LSTM) and bidirectional long short-term memory (Bi-LSTM) (Hochreiter and Schmidhuber, 1997; Graves and Schmidhuber, 2005) are two typical recurrent neural networks that have been applied in many geophysical problems to learn non-linear relationships. For example, seismic data reconstruction (Yoon et al., 2020), missing well log estimation (Pham and Wu, 2019), elastic properties and litho-fluid facies estimation (Aleardi, 2022), seismic impedance inversion and parameter estimation (Calderón-Macías et al., 2000; Moya and Irikura, 2010; Alfarraj and AlRegib, 2019; Das et al., 2019; Guo et al., 2019; Roy et al., 2020). The Bi-LSTM can work with long and dense temporal traces as the LSTM does, but it can also learn from long-term forward and backward temporal dependencies from historical data. It works for complex situations, including noise overlapping with data. Thus, these networks are suitable to be used in time-lapse seismic data.

In this chapter, I propose a data-driven method to predict and mitigate non-reservoir-related changes between baseline and monitor data, using stacked LSTM and CNN-Bi-LSTM with surface multiple reflections. The networks predict non-reservoir data variations of the monitor data from the baseline data by using as training input the data acquired above the reservoir. Then, a double difference method is applied to the migrated data to generate model residuals between predictions and observations. The time-lapse model residuals are fed into a U-Net for image denoising. The U-Net, in this chapter, acts as a filter and mask to enhance image resolution and mitigate noise on a large scale. Reservoir changes become more visible after attenuating the migration artifacts. Additionally, I take advantage of multiple reflections generated from free surface boundary conditions to broaden subsurface illumination.

THEORY

Time-lapse seismic

In time-lapse seismic, baseline data is set as the reference. After injecting CO_2 or other fluids into a target area, monitor data collected after a long period of time is used to indicate the seismic attribute variations. Based on Alali et al. (2022), the residual between monitor data $\mathbf{d}_{obsm}$ and baseline data $\mathbf{d}_{obsb}$ is

$$\delta\mathbf{d}(t) = \mathbf{d}_{obsm}(t) - \mathbf{d}_{obsb}(t) = \mathbf{n}(t) + \delta\mathbf{r}(t), \tag{1}$$

where t means time, $\mathbf{n}(t)$ denotes noise generated from near-surface change and non-repeatable data due to the change of acquisition systems, which should be eliminated. The data difference term, $\delta\mathbf{r}(t)$ denotes subsurface reservoir variations at the target layer. The aim is to determine reservoir changes from the signal, which is mixed with noise and migration artifacts. This is difficult to achieve with classical linear algorithms since artifacts and noise can have complex non-linear origins, for example, in the near-surface. Neural networks can learn non-linear patterns that characterize the noise and artifacts by using the monitor and baseline data above the targeted anomaly as training data. This network can then be used for the whole survey to predict and mitigate the noise. The input traces are baseline data, and true labels are monitor data above the reservoir change area.

Different from other research, in this chapter, two windows are constructed to predict two sets of baseline data. The first window, located at a time record of a shallow depth, matches near-surface change from baseline to estimated baseline data, given the corresponding section of the observed monitor data as labels. Even though it is called “estimated baseline data”, it should predict some near-surface noise. The other window, positioned at a time record to a greater depth, is deployed to predict another baseline data from observed baseline data. However, it is essential to note that the two sets of predicted baseline data refer precisely to the areas above the reservoir changes. The aforementioned terms individually indicate the predictions obtained from the shallow and deep windows, which are subsequently used in a double difference method. After subtraction, the near-surface noise should be reduced, and reservoir variations can be obtained.

Figures 1 to 3 present a simplified example illustrating the rationale behind utilizing shallow and deep windows in this chapter. These figures show a trace used in the workflow and the corresponding estimated results. In Figure 1, the trace used in neural network training is shown in the dashed line with the first order of surface multiples added at 0.3 and 0.8 seconds. The primary reflections depicted in this figure compare different time records for primaries and multiples, aiding in determining the shallow and deep windows in the subsequent step.

Figure 2 provides a trace comparison between the observed baseline (red line) and monitor (dashed line) with two windows. One assumption is that the observed monitor has a 0.03-second time shift lag from baseline due to the near-surface change. Additionally, its amplitude decreases to 0.9 times that of the baseline data. A reservoir change is identified at the 1-second mark in the time records.

Two windows are defined based on the primary and surface multiple distributions shown in Figure 1. Firstly, a shallow window with one primary reflection is established, ranging from 0 to 0.2 seconds. Then, a deep window is defined from 0 to 0.45 seconds, involving two primary reflections and one surface multiple reflection. Both windows are positioned above the reservoir change area, which is located at 1 second. Traces selected from shallow and deep windows are separately fed into a neural network for training and testing.

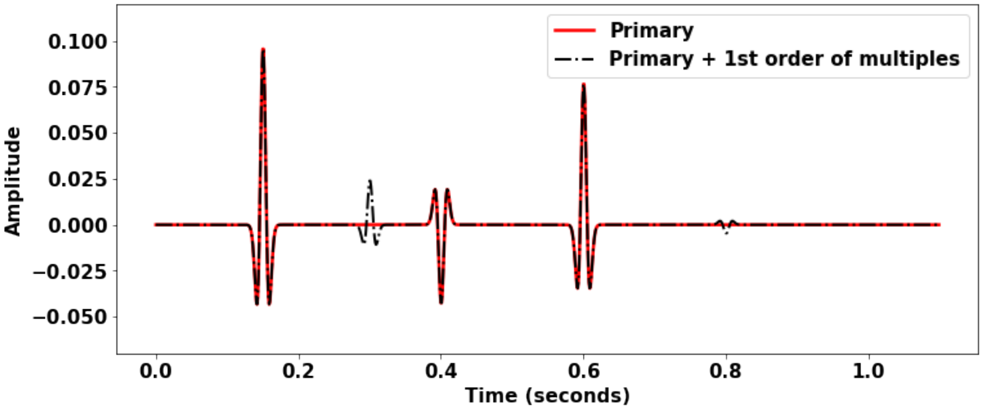


FIG. 1: Example of a trace prepared for the neural network training. Primary reflections (red line) from three subsurface reflectors are observed separately at 0.15, 0.4 and 0.6 seconds. The first order of surface multiple reflections from the first and second reflectors are acquired at 0.3 and 0.8 seconds, shown in the dashed line.

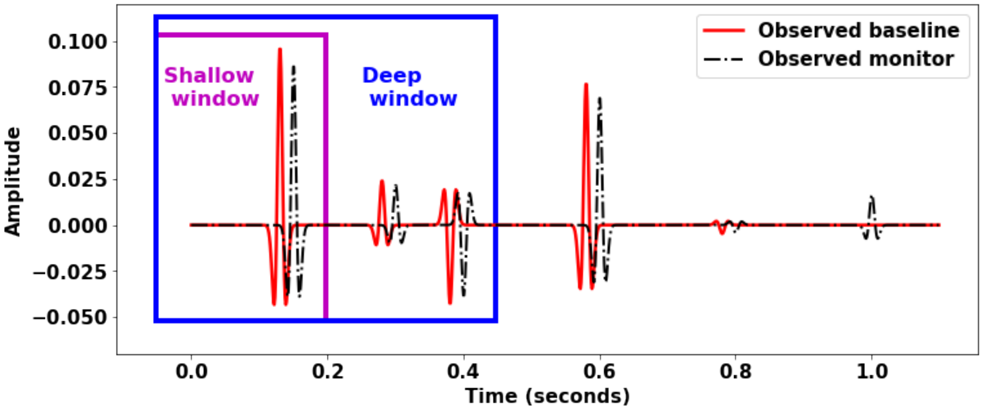
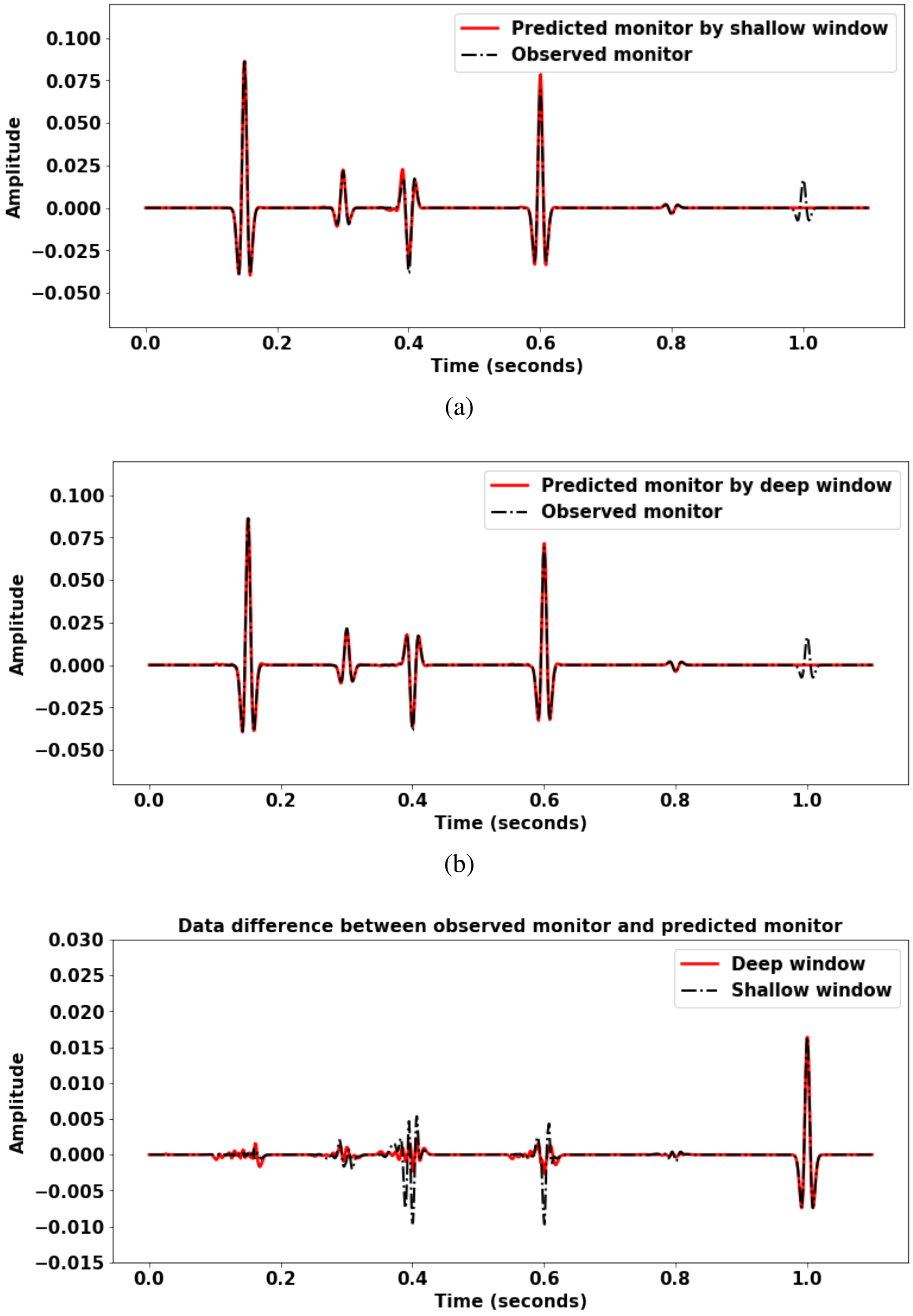


FIG. 2: Example of two windowed trace inputs for the neural network training. A shallow window (purple box) contains a primary reflection, and a deep window (blue box) consists of primaries and the first order of surface multiple reflections from the first reflector.

The results of shallow window prediction, deep window prediction and data differences are presented in Figure 3a, b and c, respectively. For ease of comparison, Figure 3c demonstrates that, before the 0.2 seconds time record, the predicted monitor obtained from the shallow window (dashed line) can give fewer artifacts than the deep window (red line).

On the other hand, after 0.2 seconds, the deep window demonstrates its capability to leverage information from surface multiples and additional primaries, resulting in more accurate trace prediction than the shallow window. From this observation, the shallow window prediction effectively handles and mitigates the impact of near-surface change, and the deep window prediction can recover the amplitude of the observed monitor data from the baseline. This rationale underlies the selection of two windows for predicting the two baseline data sets. A detailed algorithm design will be shown in the workflow section.



(a)

(b)

(c)

FIG. 3: Prediction of the trace example. Comparisons between the predicted monitor (red line) by (a) shallow window and (b) deep window, and the observed monitor (dashed line).

In the following subsections, I introduce some fundamental knowledge about LSTM, Bi-LSTM and stacked neural networks.

Recurrent neural network (RNN) and LSTM

Next, I will set the deep learning background used in this chapter. I will start from a simple LSTM framework, and then expand to a Bi-LSTM and stacked neural networks. A recurrent neural network (RNN) is an artificial neural network that uses sequential or time series data. RNN (Jordan, 1986; Rumelhart et al., 1985) is derived from a feed-forward neural network where the connections between nodes do not form a cycle and deliver information in one direction (Figure 4). On the other hand, RNN has an internal self-looped deep-learning architecture. It allows previous output to affect subsequent input and output. In other words, the current input will learn from and depend on the past sequence output. After obtaining the current output, it will be sent back into the recurrent network. The benefit of RNN is that it can process variable-length input sequences, for example, time sequences. While training an RNN model, vanishing or exploding gradient issues might occur. If the gradient is too small or large, it tends to grow or vanish when it is passed back through many time steps.

Long short-term memory (LSTM) (Hochreiter and Schmidhuber, 1997) was designed with special memory cells to store temporal information. It can remember values over arbitrary time intervals with gate structures shown in Figure 5. Also, it can avoid the vanishing gradient problem usually occurring in RNN. The gradient in LSTM contains the gate's vector of activations, allowing the network to control the gradient values better and avoid getting too small or large. This structure will enable LSTM to remember long-range features better than conventional recurrent neural networks. This capability makes LSTMs good candidates to work with seismic traces since usually, we are interested in capturing long-term dependencies along the time direction introduced by physical features across the ray paths.

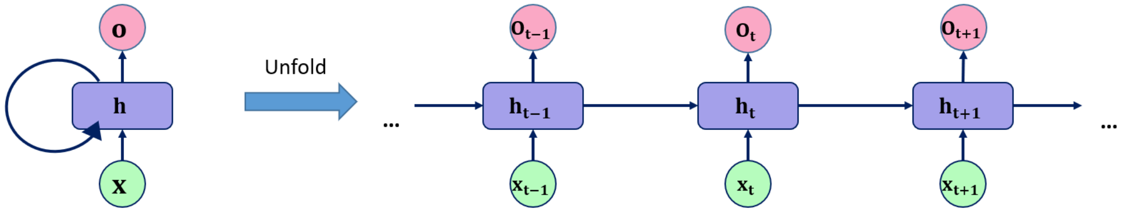


FIG. 4: Sketch for the recurrent neural network.

Within each LSTM cell (shown in Figure 5), there are four gates in total: $\mathbf{f}_t$, $\mathbf{i}_t$, $\mathbf{g}_t$ and $\mathbf{o}_t$, which are the *forget gate*, *input gate*, *candidate gate* and *output gate* cell activation vectors, respectively. They regulate the flow of information in and out of the LSTM cell. They have the same size as the hidden vector $\mathbf{h}_t$. Next, each gate vector will be illustrated in detail. The forget gate $\mathbf{f}_t$ is determined by

$$\mathbf{f}_t = \sigma(\mathbf{W}_f[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f), \tag{2}$$

where $\mathbf{h}_{t-1}$ is the hidden layer vector from the previous time step, and $\mathbf{x}_t$ means the current

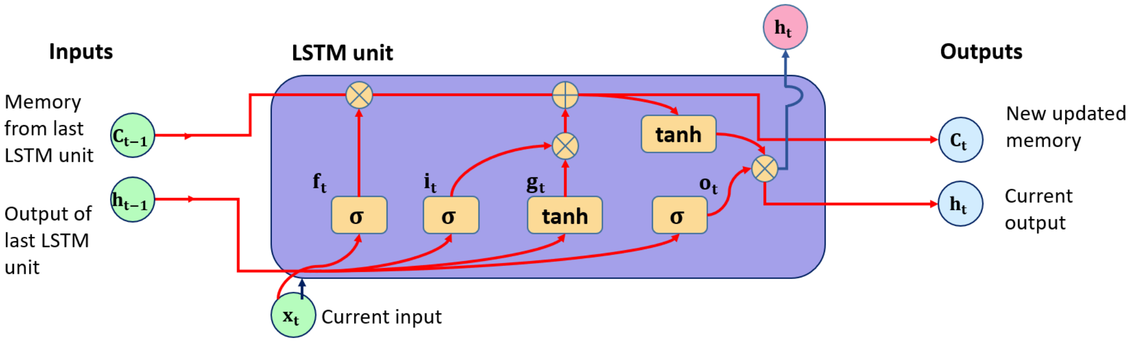


FIG. 5: LSTM unit distribution.

step input vector. σ, $\mathbf{W}_f$ and $\mathbf{b}_f$ represent the logistic sigmoid activation function, weight matrices and bias for the forget gate.

Then, the input gate $\mathbf{i}_t$ follows a similar behaviour as the forget gate but with different weight and bias:

$$\mathbf{i}_t = \sigma(\mathbf{W}_i[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i). \tag{3}$$

Another candidate gate $\mathbf{g}_t$ can be determined with a $tanh$ activation layer:

$$\mathbf{g}_t = tanh(\mathbf{W}_g[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_g). \tag{4}$$

The last gate in the LSTM unit is output gate $\mathbf{o}_t$, and it can be obtained by:

$$\mathbf{o}_t = \sigma(\mathbf{W}_o[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o). \tag{5}$$

After having all the gates, the next step is to calculate and determine a new updated memory $\mathbf{C}_t$ and current output $\mathbf{h}_t$. The former needs memory from the last LSTM unit $\mathbf{C}_{t-1}$ combined with forget gate $\mathbf{f}_t$ and the product of input gate $\mathbf{i}_t$ and candidate gate $\mathbf{g}_t$. Then, the current cell state $\mathbf{C}_t$ can be calculated by

$$\mathbf{C}_t = \mathbf{i}_t\mathbf{g}_t + \mathbf{f}_t\mathbf{C}_{t-1}. \tag{6}$$

As for the output of the LSTM cell $\mathbf{h}_t$, it is made of the product of output gate $\mathbf{o}_t$ and current cell state $\mathbf{C}_t$ after applying $tanh$ activation layer:

$$\mathbf{h}_t = \mathbf{o}_t tanh(\mathbf{C}_t). \tag{7}$$

Bi-LSTM

Unlike conventional RNNs, bidirectional RNNs (BRNNs) can deal with sequential data in both directions, forward and backward, by using two separate hidden layers (Schuster and Paliwal, 1997; Graves et al., 2013b). Based on BRNNs and LSTM, the bidirectional long short-term memory (Bi-LSTM) network (Graves and Schmidhuber, 2005) was developed to capture long sequences in reverse and forward directions. One hidden layer processes the input sequence in the forward direction. The other hidden layer handles the input in the reverse direction. It includes doing backward passes for output neurons, forward and backward states, and updating error functions and weights. Thus, the output of the current time step is obtained from both layers' hidden vectors.

The forward function of Bi-LSTM with inputs of M units and N hidden units is shown below:

$$\mathbf{h}_t^n = \sum_{m=1}^{M_I} x_t^n w_{mn} + \sum_{n'=1,t>0}^{N_H} \alpha_{t-1}^{n'} w_{n'n}, \tag{8}$$

$$\alpha_t^n = \Theta_n(h_t^n), \tag{9}$$

where $\mathbf{h}_t^n$ is the output of the current time step at each unit n, and x_t denotes the sequence input. w_{mn} represents the weight of the input m to hidden unit n, and $w_{n'n}$ means the weight of hidden unit n towards hidden unit n'. Θ_n means the activation function of the hidden unit n. The activation function of hidden unit n at time step t is given by α_t^n.

The backward calculation is

$$\frac{\delta O}{\delta w_n k} = \sum_{t=1}^{T} \frac{\delta O}{\delta h_t^n} \alpha_t^n, \tag{10}$$

$$\frac{\delta O}{\delta \alpha_t^n} = \Theta_n'(h_t^n)(\sum_{k=1}^{K} \frac{\delta O}{\delta h_t^n} w_{nk} + \sum_{n'=1,t>0}^{H} \frac{\delta O}{\delta h_{t+1}^{n'}} w_{nn'}), \tag{11}$$

where O denotes an objective function with unit K output.

For the bidirectional LSTM, we need to consider the forward and backward flows in two separate layers (Du et al., 2020). The final output can be obtained by

$$h_t = \gamma h_t^f + \beta h_t^b, \tag{12}$$

$$y_t = \sigma(h_t), \tag{13}$$

where h_t^f is the forward LSTM layer output which takes time sequences from x_1 to x_T, h_t^b denotes the backward LSTM layer output which takes the reverse time sequences from x_T to x_1. γ and β represent the importance of forward LSTM and backward LSTM, and satisfy $\gamma + \beta = 1$. h_t is the sum of two LSTM outputs, and y_t is the prediction after using an activation function. In this chapter, the prediction at this stage involves separately estimating the baseline and monitor data from two distinct windows. For the output dense layer, a linear activation function is selected, considering that seismic traces can contain negative values.

Stacked LSTM and Bi-LSTM

Based on the research from Graves et al. (2013a), and Cui et al. (2018), a deep LSTM can be generated by stacking several LSTM hidden layers on top of each other. This scheme also works for the Bi-LSTM networks. The output of one Bi-LSTM hidden layer will be fed into the subsequent Bi-LSTM hidden layer as the input. Figure 7 shows the stacked Bi-LSTM mechanism. The input of every hidden layer should consist of both the forward and backward layers at the level below. Stacked Bi-LSTM can detect and build up an effectively high level of sequential data representations.

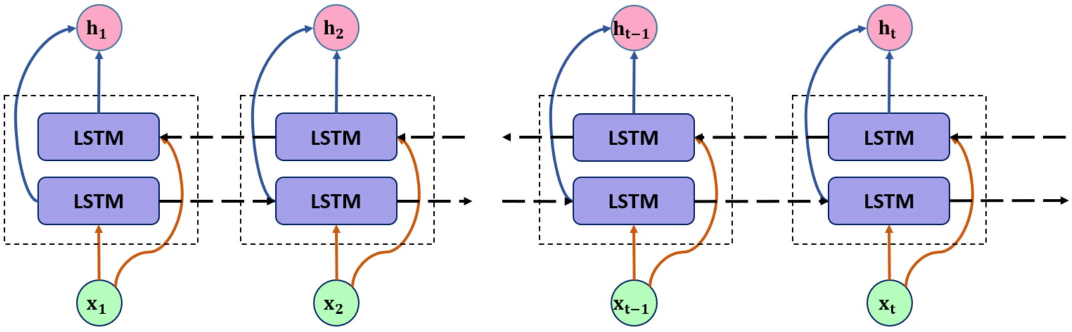


FIG. 6: Workflow for Bi-LSTM algorithm.

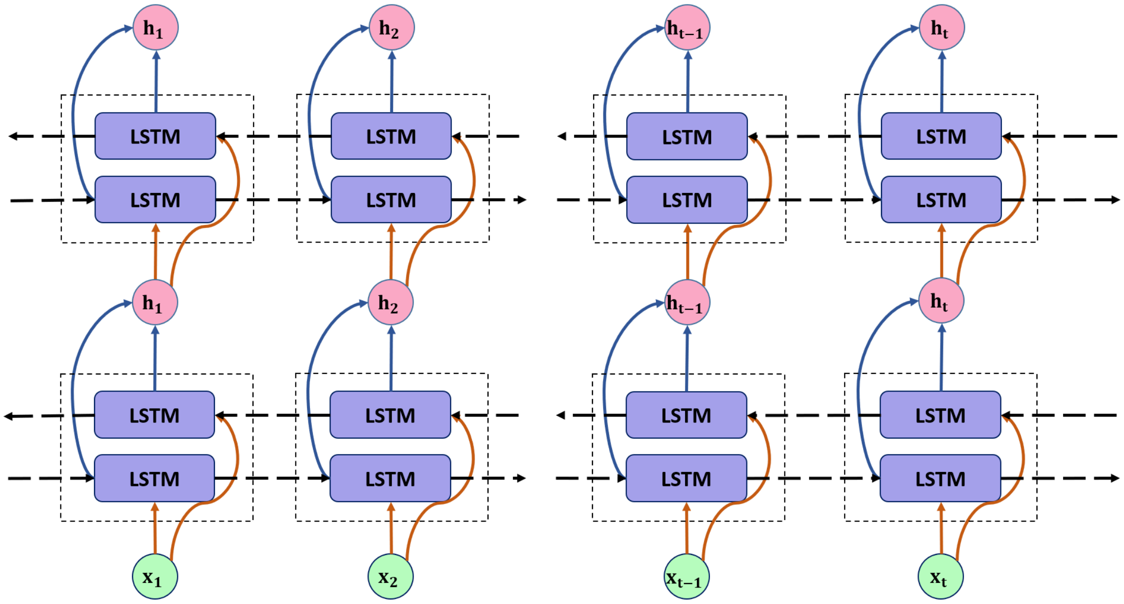


FIG. 7: Stacked Bi-LSTM.

The subsequent section will present a detailed description of the proposed method's workflow and parameter configuration. This will provide a comprehensive understanding of the methodology. Additionally, the parameter settings, such as learning rate, batch size, hidden layer size and the number of epochs, will be discussed to ensure reproducibility and optimize the method's performance.

WORKFLOW

The workflow presented in this chapter, depicted in Figure 8, consists of four distinct steps that delineate implementing the proposed method. In the first step, I extend the idea of Alali et al. (2022), training stacked LSTM neural networks using different trace parts of observed baseline data $\mathbf{d}_{obsb}$ by windowing: the shallow part of traces is used for predicting baseline data $\mathbf{d}_{calcb_s}$, and the deep part of traces above the reservoir change is trained to determine another predicted baseline data $\mathbf{d}_{calcb_d}$ but with multiples included. During the training, different parts of baseline data are the input and the corresponding parts of monitor data $\mathbf{d}_{obsm}$ are set as labels. The reason for training the network with data above the reservoir is that time-lapse requires minimizing the differences in acquisition and other near-surface variations while keeping the differences produced by fluid changes in the reservoir. Training on the shallow part can compensate only for variations in near-surface velocity. Without these corrections, the time shift in the deeper part will contain a mix of both effects. In summary, neural networks shift the predicted baseline close to the observed monitor data without affecting the anomalies sought to detect. To compare the effectiveness of the proposed method against the traditional approach, I apply a matching filter with a filter size of 21 temporal steps to the baseline shot records. An analysis of the results is provided in the subsequent sections.

After obtaining the two sets of predicted baseline data, the next step is calculating the migration difference by an elastic reverse time migration with surface-related multiples added. Then, A double-difference method, calculated from the difference between two residuals of observed and predicted migration images, is applied to determine the time-lapse reservoir changes. The double-difference concept originated from tomography and inversion (Watanabe et al., 2004; Asnaashari et al., 2015; Zhou et al., 2010). To address the limitations of the direct data difference approach ($\mathbf{d}_{obsm} - \mathbf{d}_{obsb}$), I propose an extension of the data difference formulation with two additional sets of predicted baseline data. This modification is necessary as the direct data difference method yields unsatisfactory results. To improve the accuracy of the subtraction process and mitigate the influence of near-surface noise, I introduce a more complex technique. So, the difference between the two sets of observation and prediction data is:

$$\delta\mathbf{d} = (\mathbf{d}_{obsm} - \mathbf{d}_{obsb}) - (\mathbf{d}_{calcb_s} - \mathbf{d}_{calcb_d}), \tag{14}$$

where $\mathbf{d}_{obsm}$ and $\mathbf{d}_{obsb}$ denote observed monitor and baseline data, respectively, and $\mathbf{d}_{calcb_s}$ and $\mathbf{d}_{calcb_d}$ represent predicted data by using shallow and deep windows separately. The predicted baseline using a shallow window is subtracted from the direct data difference to reduce the near-surface noise. Additionally, to account for the contribution of multiple reflections, the predicted baseline obtained using a deep window is added to the resulting difference. This addition helps to incorporate the valuable information provided by multiple reflections. Then, the time-lapse model changes $\delta\mathbf{m}_{time-lapse}$ can be derived:

$$\delta\mathbf{m}_{time-lapse} = \delta\mathbf{m}_{obs} - \delta\mathbf{m}_{calc}, \tag{15}$$

where $\delta\mathbf{m}_{obs}$ and $\delta\mathbf{m}_{calc}$ are calculated from the first and second parentheses in equation 14, respectively.

After applying reverse time migration and double-difference on migrated images, the final step is using a U-Net (Ronneberger et al., 2015) to denoise predicted reflectivity model residuals. This U-Net architecture was first mentioned in Chapter 3 with additional skip connection layers. This chapter uses it to learn patterns from time-lapse reflectivity differences and mitigate artifacts in the migrated images. In this chapter, the input channel contains the differences in reflectivity between observed and predicted migrated images. The additional skip connection layer works as an identity mapping, to transfer information and patterns from the input. It is important to note that, for the real field data example with DAS, stacked data is utilized instead of migration images. This decision is made due to the absence of information regarding the imaging velocity. Therefore, stacked data is employed as a suitable alternative for the purpose of the analysis. Additionally, I use stacked

Bi-LSTM with convolutional neural network blocks in the DAS example because bidirectional neural networks can handle better the complexity of DAS data for the near-surface part.

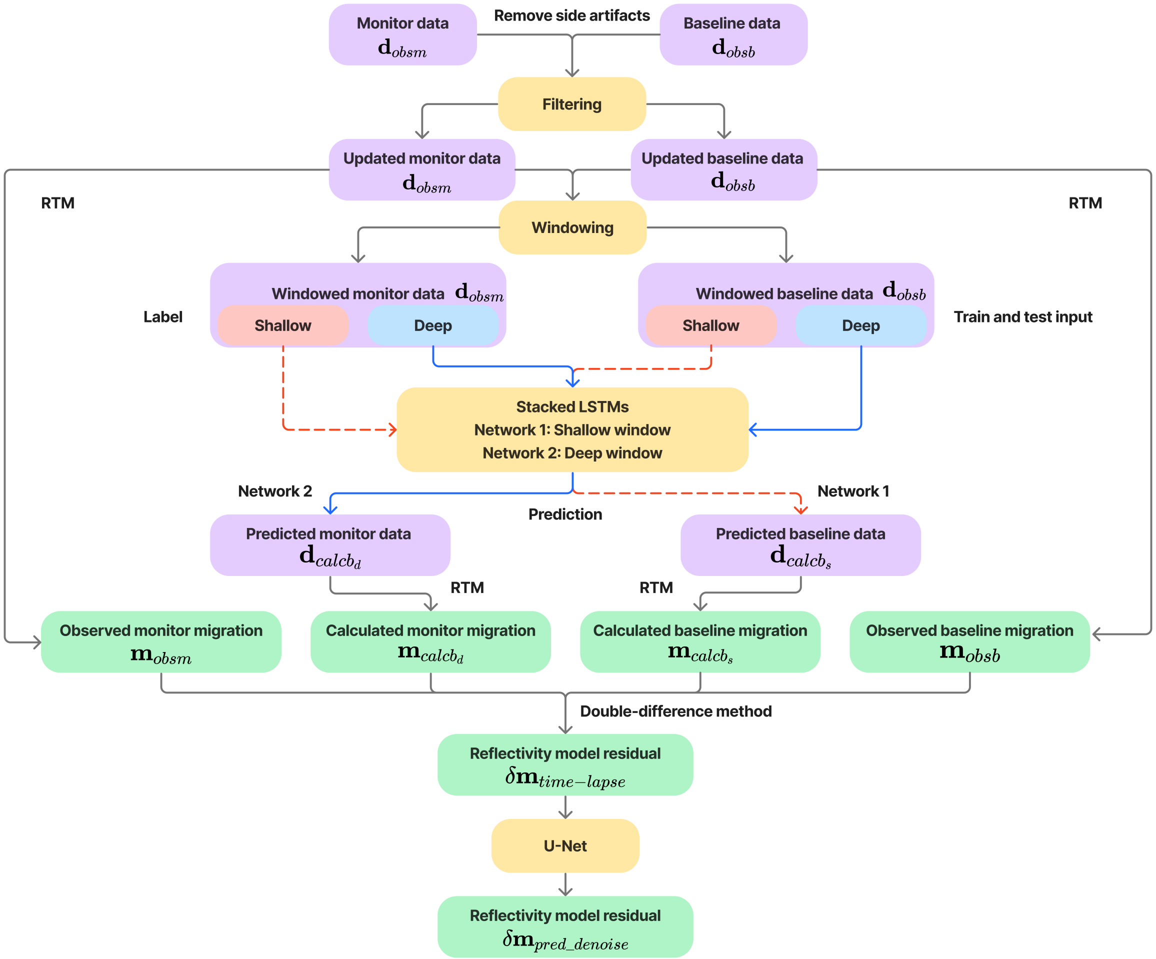


FIG. 8: Workflow for predicting time-lapse reflectivity change.

NEURAL NETWORK SETTINGS

The main assumption in this chapter is that monitor and baseline acquisition systems are the same but with near-surface velocity changes in the monitor data. Five geology models are used for training and testing, including a horizontal-layered model, a curve-fault model, the Marmousi model, the Overthrust model, and a field DAS example. A slight reservoir anomaly is located at the deep part of each model to generate weak amplitudes of data difference. The train and test rates are set to 0.8: 0.2.

A stacked LSTM neural network is used to predict data anomaly above reservoir changes for the first four synthetic examples. The hyperparameters for the first stacked LSTM are listed as follows. The batch size is 128, the number of epochs is 50, the learning rate starts from 0.01, and the hidden layer size is 100. Two layers of LSTM are applied with a 0.1% dropout. The final output activation function is a $tanh$ function.

As for the DAS VSP data, a complex CNN-Bi-LSTM is applied to handle noisy data for extracting and learning key features of signals from noise. The 1D CNN-Bi-LSTM has a more profound architecture compared with the two-layer stacked LSTM mentioned above. Figure 9 shows the neural network architecture's hyperparameters in each step box. It combines two 1D CNN blocks with stacked Bi-LSTM layers to suppress significant noise through training. Convolutional layers can extract amplitude information from the seismic traces and their output is fed into the stacked Bi-LSTM. After that, an LSTM layer with long units is set to strengthen the training process and provide stabilization. ReLu activation functions are used in the ConvNet and LSTM. Lastly, A dense layer with a linear activation function outputs the final prediction. The learning rate also starts from 0.01 with a "patience" hyperparameter of 15 iterations. The number of epochs is 100, with a batch size of 512.

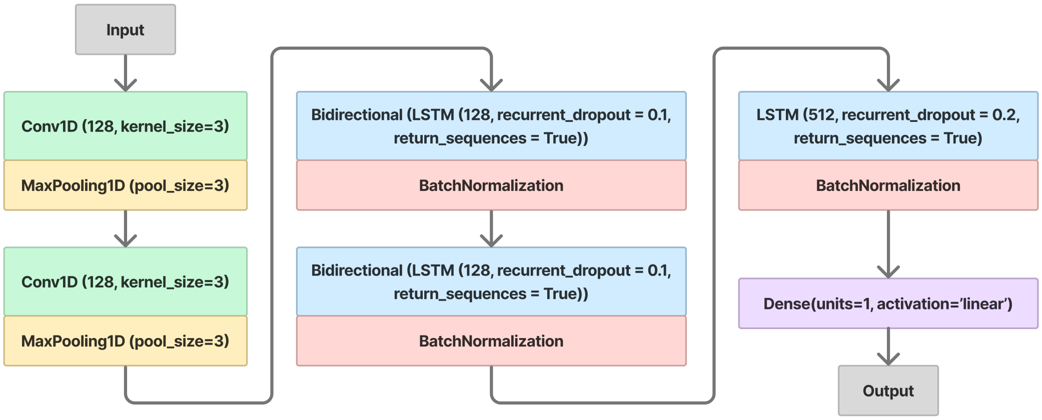


FIG. 9: Stacked Bi-LSTM with convolutional layers architecture.

In the synthetic examples in this chapter, I also use a U-Net to suppress the noise in the predicted model differences. The total number of input windows varies depending on the model's size, which is extracted partly from the predicted reflectivity residual. The true (known) reflectivity difference is set as the label. The train and test rates are also set to 0.8: 0.2. After trials with different window sizes, empirically, I decided to use 64×64 points, which seems a proper size to capture features in the reservoir anomaly. Learning rates for this U-Net start from 0.001 with "patience" of 20 iterations. There are 200 epochs in total, with a batch size of 64. The root-mean-squared error (RMSE) and peak-signal-to-noise ratio (PSNR), introduced in Chapter 3, are applied to measure network performance. Similarly to the mean square error, the root mean square error (RMSE) penalizes large errors when given a large-scale data set and evaluates the quality of predictions. It is the square root of the average of the squared differences between the estimated and the actual value of the variable/feature.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\mathbf{d}^i_{calc} - \mathbf{d}^i_{obs})^2}, \tag{16}$$

where n is the total number of samples, $\mathbf{d}_{calc}$ are the predicted shot records, and $\mathbf{d}_{obs}$ represents the observed data.

NUMERICAL EXAMPLE

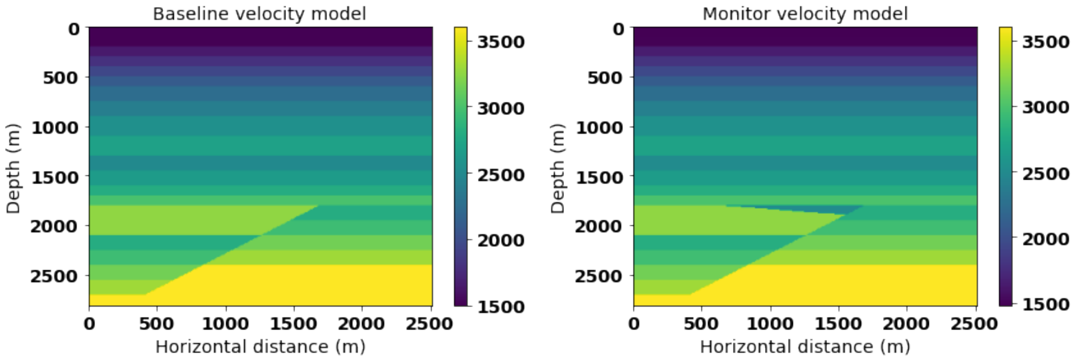
In this section, I show five examples to illustrate the generalization of the proposed workflow for different subsurface structures with weak reservoir anomalies. The first four are synthetic models with the same spatial interval, time sampling rate and total record time. The models have near-surface velocity changes from the baseline data to the monitoring data, which could come, in a real case, from changes in the acquisition geometry (not in this case) or changes in the near-surface conditions, like for example, in the water table level or freezing conditions. This setting is designed to test the attenuation of near-surface differences that account for most non-reservoir-related changes between baseline and monitor surveys. The proposed stacked LSTM can learn minor changes and time shifts from the perturbations and predict baseline data close to monitor data at shallow depth. The fourth example shows the versatility of the proposed method by showcasing its ability to generalize to similar models or situations using pre-trained features. This highlights the robustness and adaptability of the approach, as it can effectively estimate CO_2 migration paths for data collected from other parts of the same survey. For the last example, a field DAS dataset, the monitor data has a denser acquisition system than the baseline data. Based on the data above, the proposed method using CNN-Bi-LSTM neural network gives reliable predictions on the shape of the CO_2 injection area.

Horizontal-layered model

The horizontal-layered model contains several flat layers with a thrust anomaly at around 1800 meters. The total size of the model is 281×251 gridpoints with 10 meters spatial interval (Figure 10). I simulated 50 shots with 126 receivers each at 10 meters in depth. The recorded time is 3 seconds with a 0.0015 seconds time sampling rate. At the uppermost portion of the monitor velocity model, two layers are introduced to simulate near-surface variations, with velocity values ranging from 1480 m/s to 1500 m/s, whereas, at the same location, the baseline model has a velocity of 1500 m/s. This near-surface modification in the monitor survey simulates the weather layer and soil change on the near-surface.

In Figure 11, I show the first four shots for the observed baseline (Figure 11a), the observed monitor (Figure 11b), the predicted baseline by a shallow window (Figure 11c) and the predicted baseline by a deep window (Figure 11d), respectively. According to the workflow in the previous section, a shallow part of observed baseline traces d_{calcb_s}, set before 0.975 seconds, is used to predict the calculated baseline. The other calculated baseline d_{calcb_d} is predicted from a deeper part of observed baseline traces, between 0.525 and 1.425 seconds. Both the shallow or deep parts of the traces are above the reservoir anomaly, which occurs after 1.6 seconds in Figure 11.

Figure 12 shows data differences between two selected shots from Figure 11. The data differences between the observed monitor and predicted baseline by a deep window (Figure 12b), have significant amplitudes after 1.6 seconds due to changes in the velocity model simulating a CO_2 injection. Compared to the data difference between the observed monitor and baseline (Figure 12a), predicted shot records can suppress time shifts and noise in the shallow section in Figure 12b. The RMSE values for Figure 12b are smaller than the data



(a)

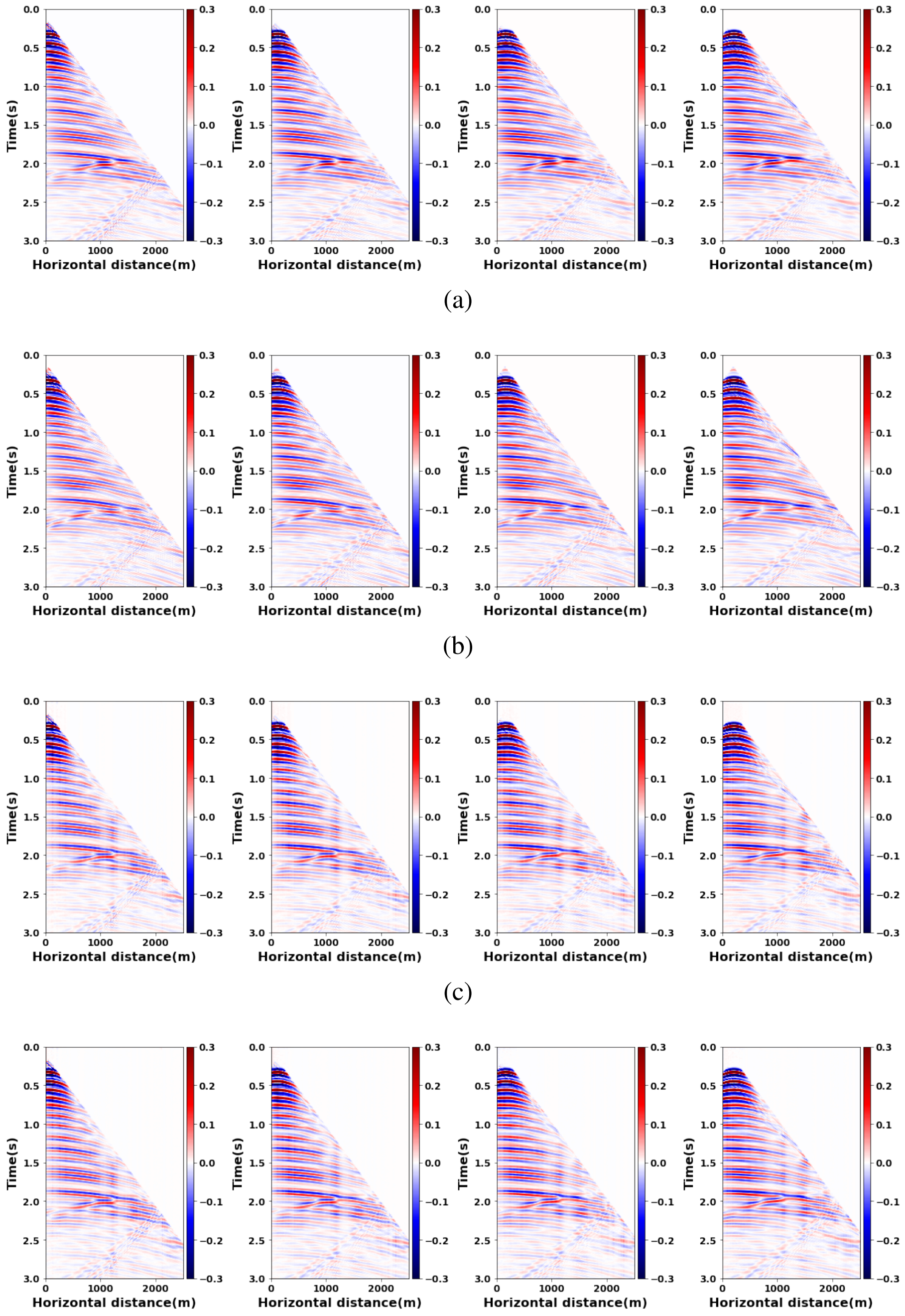
(b)

FIG. 10: Velocity model for (a) baseline and (b) monitor systems.

difference between acquired data sets (Figure 12a). This observation illustrates that predicted shot records can mitigate time shifts in the monitor data and highlight the reflections from the reservoir anomaly. Figure 12c shows the result after processing with the matching filter. Its RMSE values are slightly larger than those of the stacked LSTM approach. For comparison, Figure 12d illustrates the target difference without the near-surface changes in the acquired monitor data. Clear reflection differences can be seen between 1.6 and 2.25 seconds. Upon observation, it was found that the continuity of reflection differences is more prominent than the proposed method. However, the proposed stacked LSTM can reduce some noise between the acquired baseline and monitor data. At the same time, it also tends to distort the signal of reflection differences slightly. Despite this limitation, the stacked LSTM model is still preferred as it improves the overall quality of the predicted data.

For further illustration, two traces are extracted from shots 3 and 10, shown in Figure 13. The predicted baseline by a shallow window (blue line) gives more details for shallow time series than the predicted baseline using a deep window (green line). For example, the predicted baseline is closer to observed monitor data (dashed red line) between 1.4 and 1.5 seconds in Figure 13a. On the other hand, because the predicted baseline by a deep window is trained with a deeper time series above the reservoir anomaly, it should contain some information about time shifts and amplitude patterns from the deep section of monitor data. For instance, the deep-window predicted baseline could forecast time shifts close to observation monitor data above the reservoir change between 1.5 and 1.6 seconds.

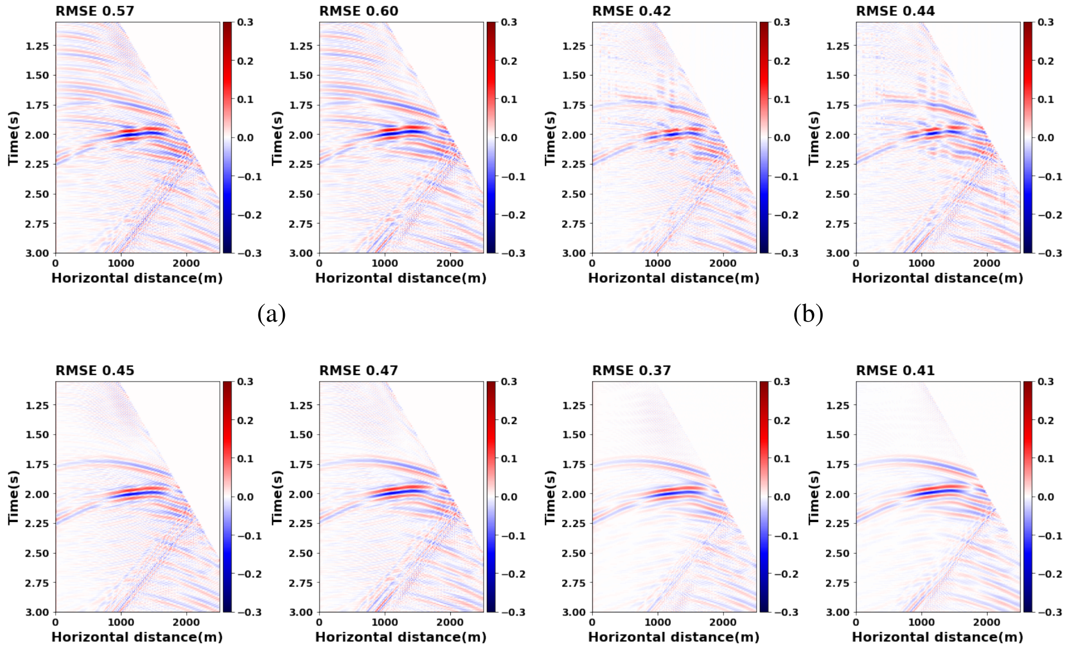
Migration results are shown in Figure 14. Observed and predicted data are migrated with a smoothed baseline velocity model to avoid information leakage. Figure 14e is designated as the target outcome to be attained by the prediction, as it is computed without considering the presence of near-surface changes. Compared with the observed difference migration result (Figure 14b), double difference migration (Figure 14c) has improved amplitude for reservoir change at 1800 m depth. This migrated image is then put into the U-Net, and the prediction is shown in Figure 14d. Based on the fundamental shape of the reservoir anomaly given in the double difference migration, the forecast after U-Net has recovered most of the anomaly with suppressed artifacts from other reflection events. Even though some noise is generated from windowing, the image has enhanced accuracy



(a) (b) (c) (d)

FIG. 11: Horizontal-layered example. Four shots are selected to show data obtained from (a) observed baseline, (b) the observed monitor, (c) the predicted baseline by a shallow window and (d) the predicted baseline by a deep window.

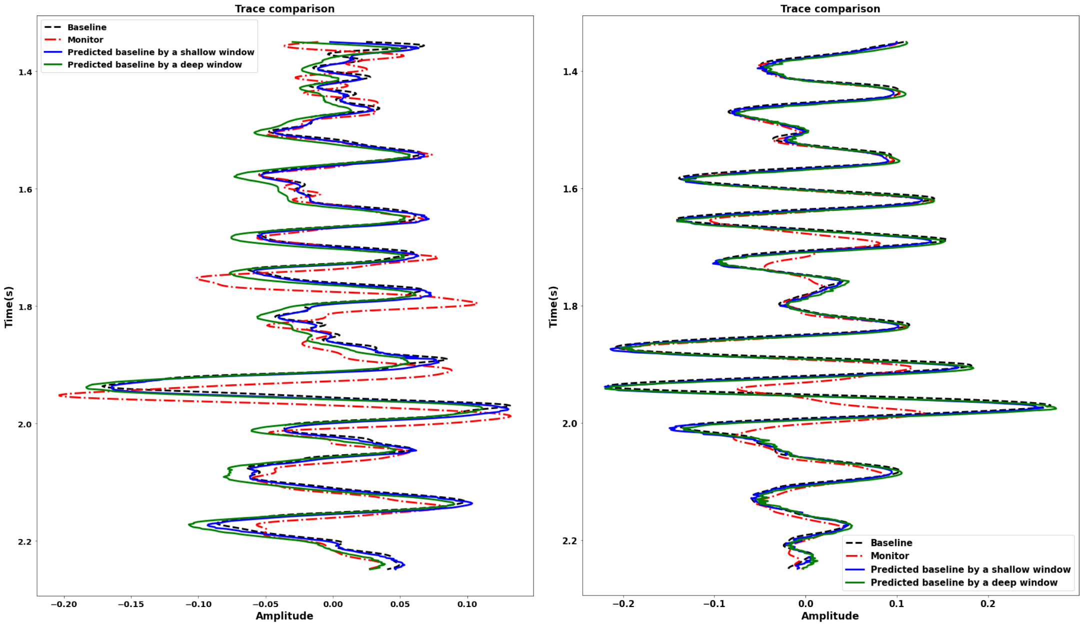
and resolution. The PSNR of the U-Net result is 21.95 dB, which indicates better quality than that of the observed difference, which has a PSNR of 19.06 dB. Additionally, the double-difference approach has a PSNR of 19.19 dB, demonstrating that the U-Net is more effective in reducing noise and improving image quality. The result shows a profound improvement, but it depends on the quality of input channels. The prediction could be biased and inaccurate if the resolution and amplitude of subsurface structures in the input are poor.



(a) (b)

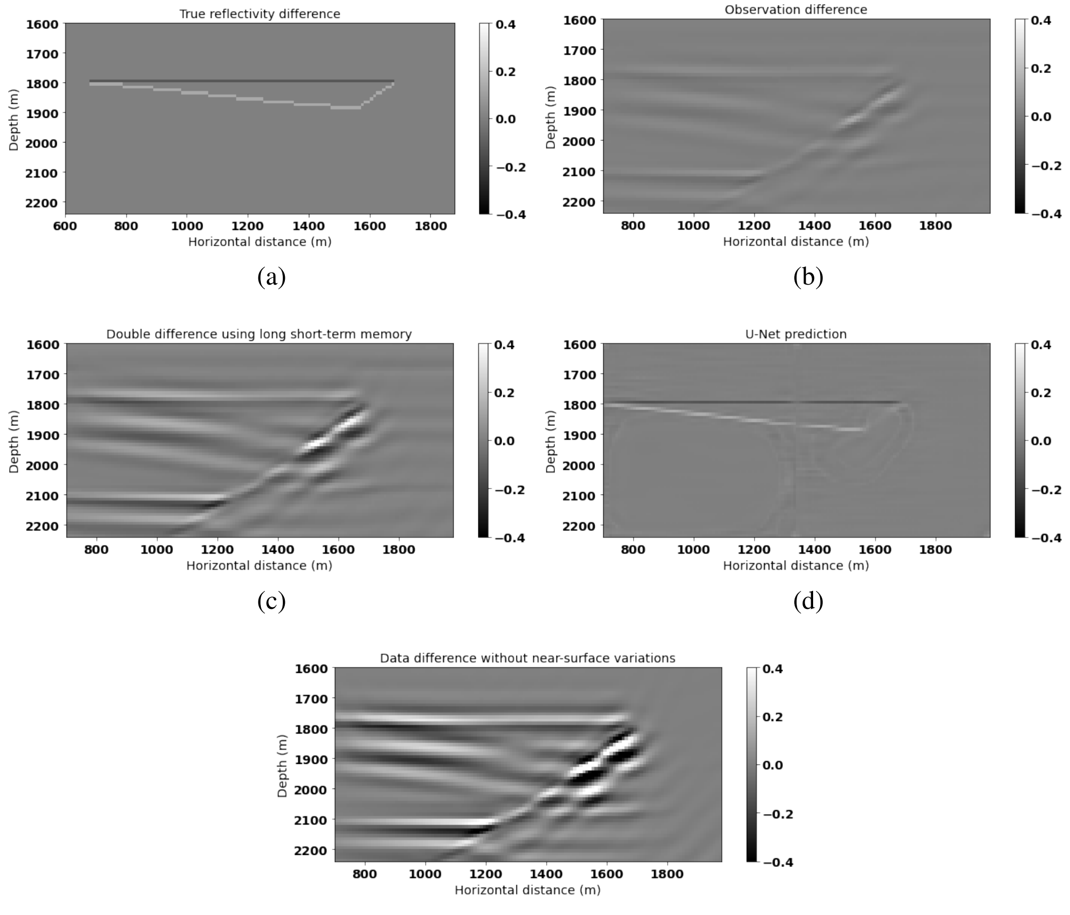
(c) (d)

FIG. 12: Horizontal-layered example. Two shots are extracted from Figure 11 to show data differences between (a) the observed monitor and baseline, (b) the observed monitor and predicted baseline by a deep window, (c) the observed monitor and predicted monitor using a matching filter, and (d) the target difference excluding near-surface change in the observed monitor data.



(a) (b)

FIG. 13: Trace comparison for shots (a) No.3 and (b) No.10.



(a)

(b)

(c)

(d)

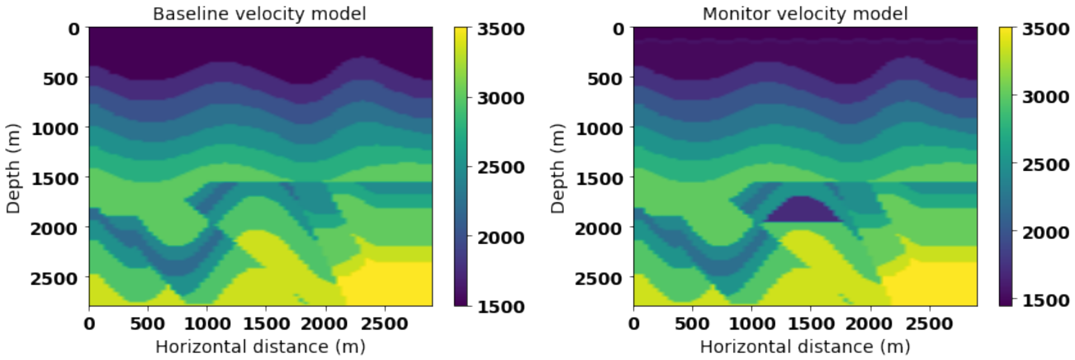
(e)

FIG. 14: Horizontal-layered example's (a) true reflectivity difference. Migration differences are generated by (b) the difference between the observed monitor and baseline, (c) the double-difference method, and (d) U-Net prediction. (e) Target migration difference without near-surface change.

Curve-fault model

Inspired by velocity models of an open source FWI project (Deng et al., 2021), I combined two synthetic models into one complete curve-fault model (Figure 15) with 279×290 gridpoints. A curved near-surface anomaly was added in the monitor velocity, whose velocity varies from 1450 to 1550 m/s. Fifty-seven shots are simulated with 145 receivers located at 10 meters depth. The observed monitor data has curved near-surface variations at around 70 meters depth. Changes in velocity, simulating a CO_2 injection, are designed between depths of 1700 and 2000 meters with 1700 m/s velocity. Training parts are extracted from traces above two seconds where the reservoir change is located. A shallow window is chosen between 0.225-1.275 seconds, and a deep window ranges from 0.450 to 1.500 seconds, where the reflections of the reservoir change occur after 1.75 seconds.

In Figure 16, the differences between monitor and baseline data for two shots are compared. Figure 16b shows the differences between observed and predicted baseline data.



(a)

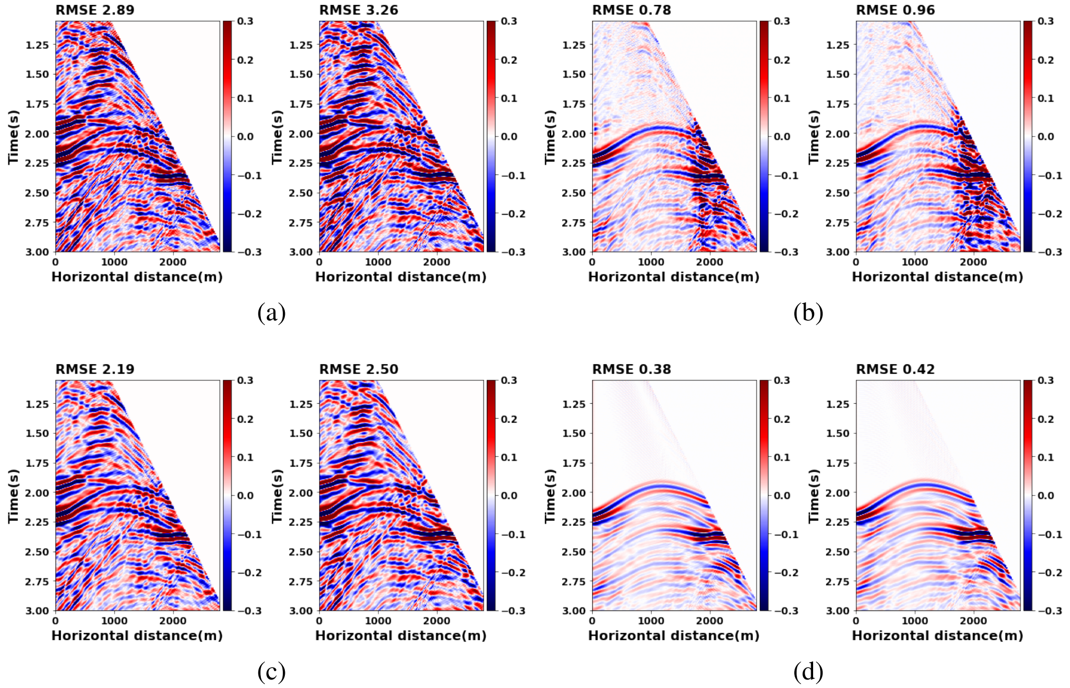
(b)

FIG. 15: Velocity model for (a) baseline and (b) monitor systems.

This difference has fewer artifacts at the top of the shot records but shows more significant reflections from the CO_2 injection area. The RMSE value for the first shot in Figure 16b is 0.78, which is smaller than the result obtained from the classical approach with a matching filter, which is 2.19. Additionally, Figure 16b result is closer to the target difference, which excludes near-surface changes, as shown in Figure 16d. This indicates that the proposed method can effectively mitigate noise between the acquired baseline and monitor data while preserving significant reflection differences. The results suggest that the proposed method is superior to the matching filter in terms of accuracy and reliability in this example.

A trace comparison in Figure 17 shows a similar result as the previous example in that the shallow-window predicted baseline (blue) and deep-window predicted baseline (green) provide different details in transferring the observed baseline to the observed monitor above the reservoir change location. For example, the predicted baseline provides appropriate amplitudes that are close to the observed monitor above 1.4 seconds. After 1.4 seconds, the deep-window predicted baseline (green), learned with additional time-shift patterns from the observed monitor data, has a precise prediction of amplitudes and slight phase shifts above the reservoir change area.

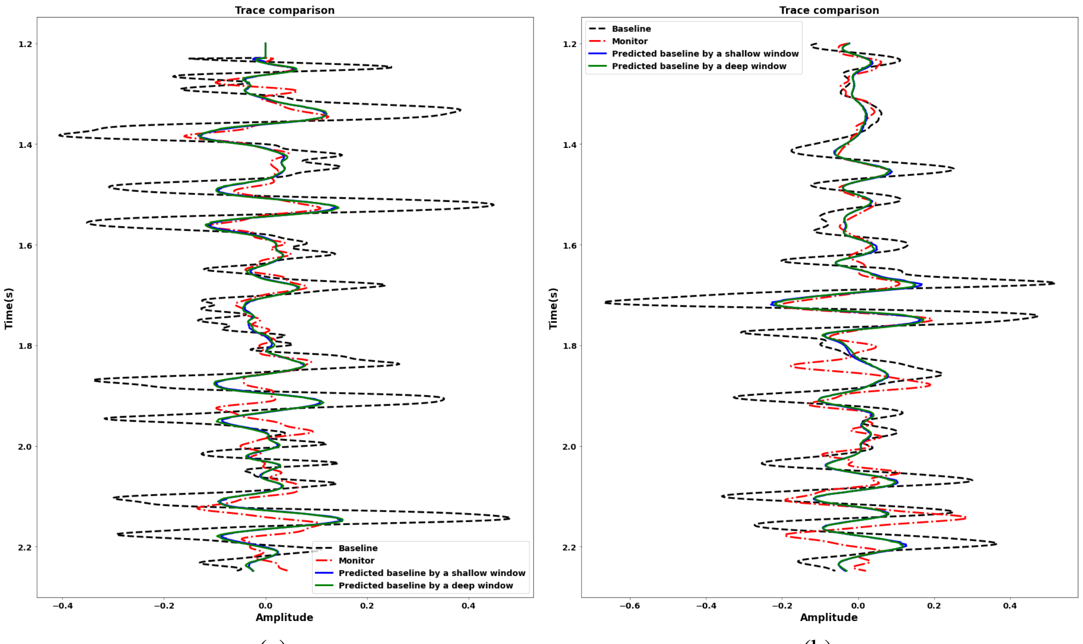
After obtaining the calculated baseline shots by two windows, the double difference migration (Figure 18c) has larger amplitudes for the anomaly structure compared with the observed data difference migration (Figure 18b). The result is also close to the ideal target migration difference shown in Figure 18e, which is the result from velocity models without the near-surface difference. Accordingly, after using U-Net, the reflectivity difference prediction (Figure 18d) removes other irrelevant reflection events but keeps the dome structure that generates velocity change with enhanced resolution. The U-Net prediction yields the highest PSNR value of 22.71 dB, outperforming both the observed difference (20.68 dB) and the double-difference approach (21.03 dB). The double-difference approach's PSNR is higher than the observed difference. This observation indicates that the double-difference method helps improve the prediction of reservoir anomalies.



(a) (b)

(c) (d)

FIG. 16: Curve-fault example. Two selected shots of data differences between (a) observed monitor and baseline, (b) observed monitor and predicted baseline by a deep window, (c) observed monitor and predicted monitor using a matching filter, and (d) the target difference excluding near-surface change in the observed monitor data.

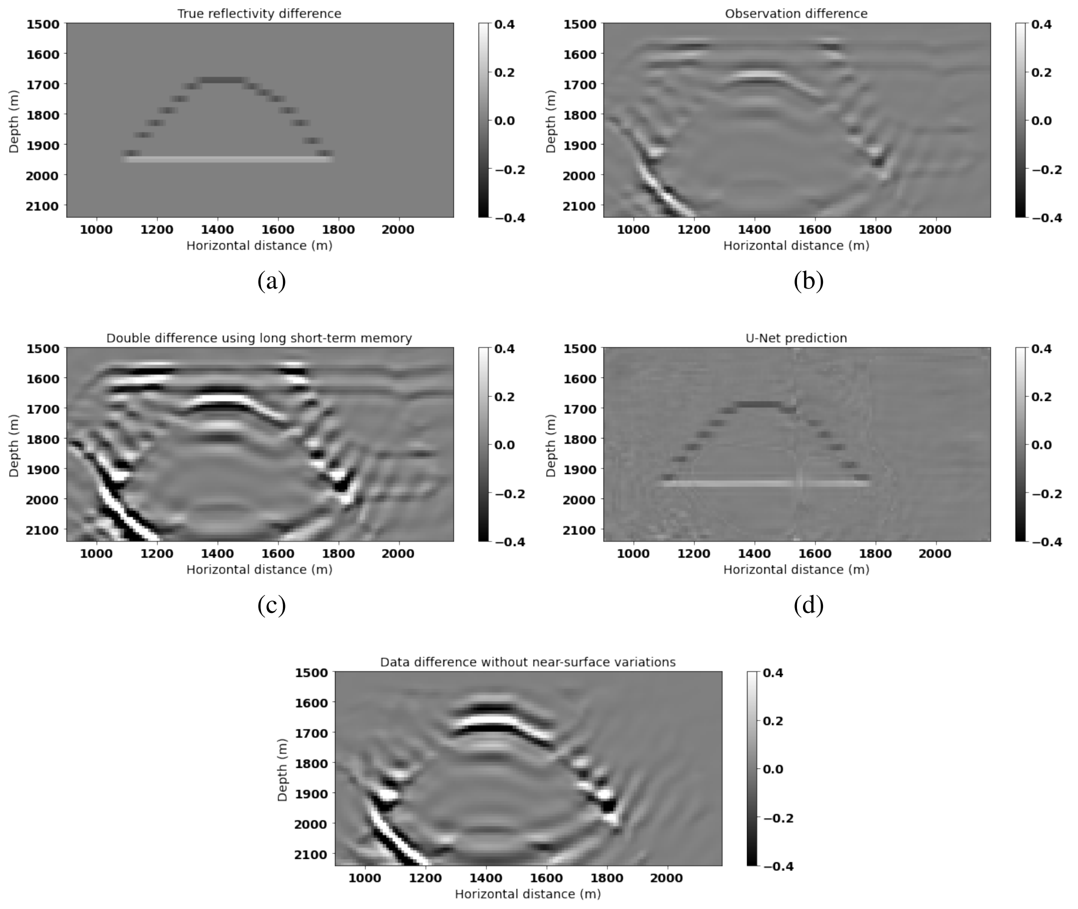


(a) (b)

FIG. 17: Trace comparison for shots (a) No.3 and (b) No.10.

Marmousi model

In this example, a part of the Marmousi model (Versteeg, 1994) is extracted and modified by adding several horizontal layers on the top (Figure 19). Now the model size changes

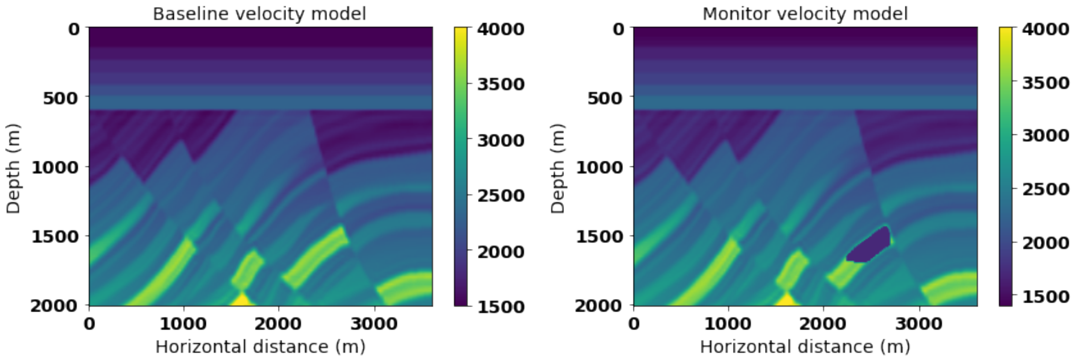


(a) (b) (c) (d) (e)

FIG. 18: Curve-fault example's (a) true reflectivity difference. Migration differences are generated by (b) the difference between the observed monitor and baseline, (c) double-difference method, and (d) U-Net prediction. (e) Target migration difference without near-surface change.

to 201×361 with 72 shots and 181 receivers, respectively. Three thin layers are located near the surface in the monitor velocity model, with velocity varying from 1400 to 1500 m/s to simulate near-surface changes. A velocity change, simulating CO_2 injection, is introduced at a depth of 1500 m with 1700 m/s. This would be a good location for CO_2 injection because the model has a caprock on the top and a fault on the side to cut off the CO_2 migration path. The shallow window is defined within the time range of 0.075-1.125 seconds, while the deep window spans from 0.075 to 1.425 seconds. The reservoir change reflections occur after 2 seconds.

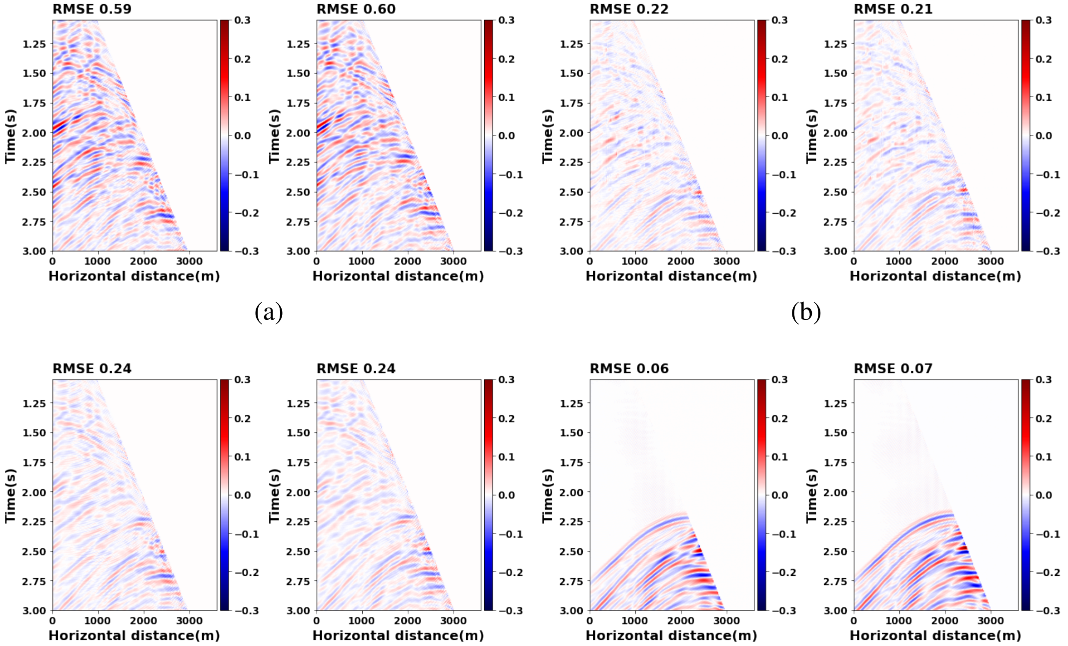
Since the pretended CO_2 injection area is located deep in the structure compared with the previous two examples, the data difference appears later in the shot record, mainly after 2 seconds. The predicted baseline data (Figure 20b) has fewer artifacts above 1.75 seconds compared with the matching filter result (Figure 20c). This observation is consistent with the target difference illustrated in Figure 20d. Thus, in this example, the proposed method again provides a better approach to predicting the reflections from the CO_2 injection area



(a) (b)

FIG. 19: Velocity model for (a) baseline and (b) monitor systems.

than the matching filter. The predicted shot record data by the proposed method shows fewer artifacts on the top of the time record, and the subtraction gives more significant reflections from the CO_2 injection area.

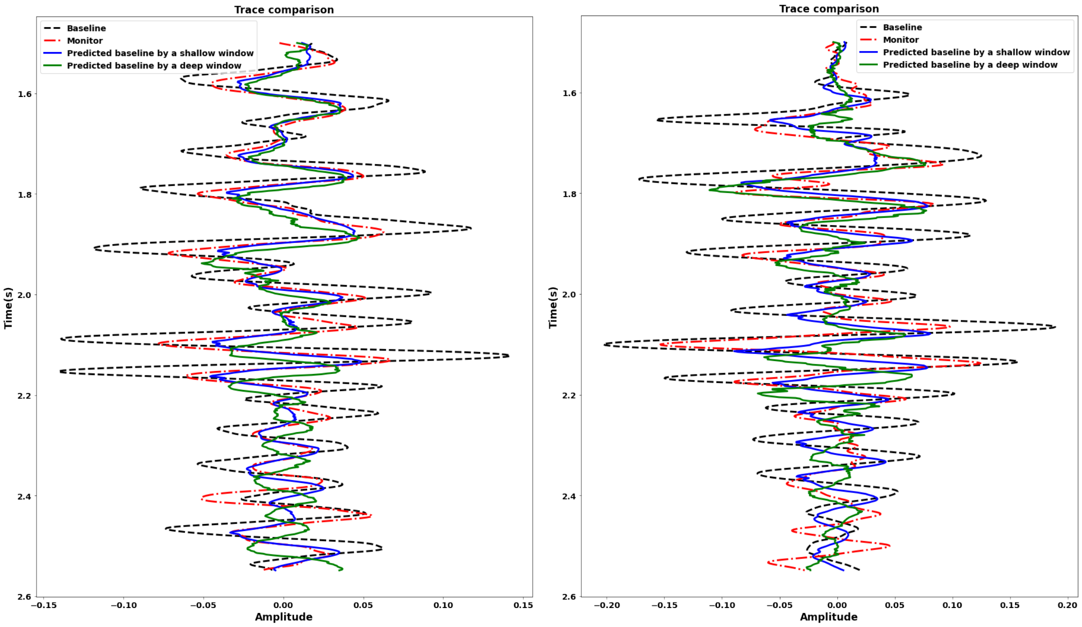


(a) (b)

(c) (d)

FIG. 20: Marmousi example. Two selected shots of data differences between (a) observed monitor and baseline, (b) observed monitor and predicted baseline by a deep window, (c) observed monitor and predicted monitor using a matching filter, and (d) the target difference excluding near-surface change in the observed monitor data.

To further investigate the performance of the method, a trace comparison between the predicted and observed data is presented in Figure 21. The predicted baseline by a deep window (green line) predicts amplitude and phase shifts closer to the observed monitor (dashed red line) than the predicted baseline by a shallow window (blue line) above 2 seconds. After 2 seconds, except for reflections of the reservoir change, the shallow-window predicted baseline produced higher accuracy in estimating time shifts. Conversely, the deep-window predicted baseline offers more precise amplitude estimations.



(a) (b)

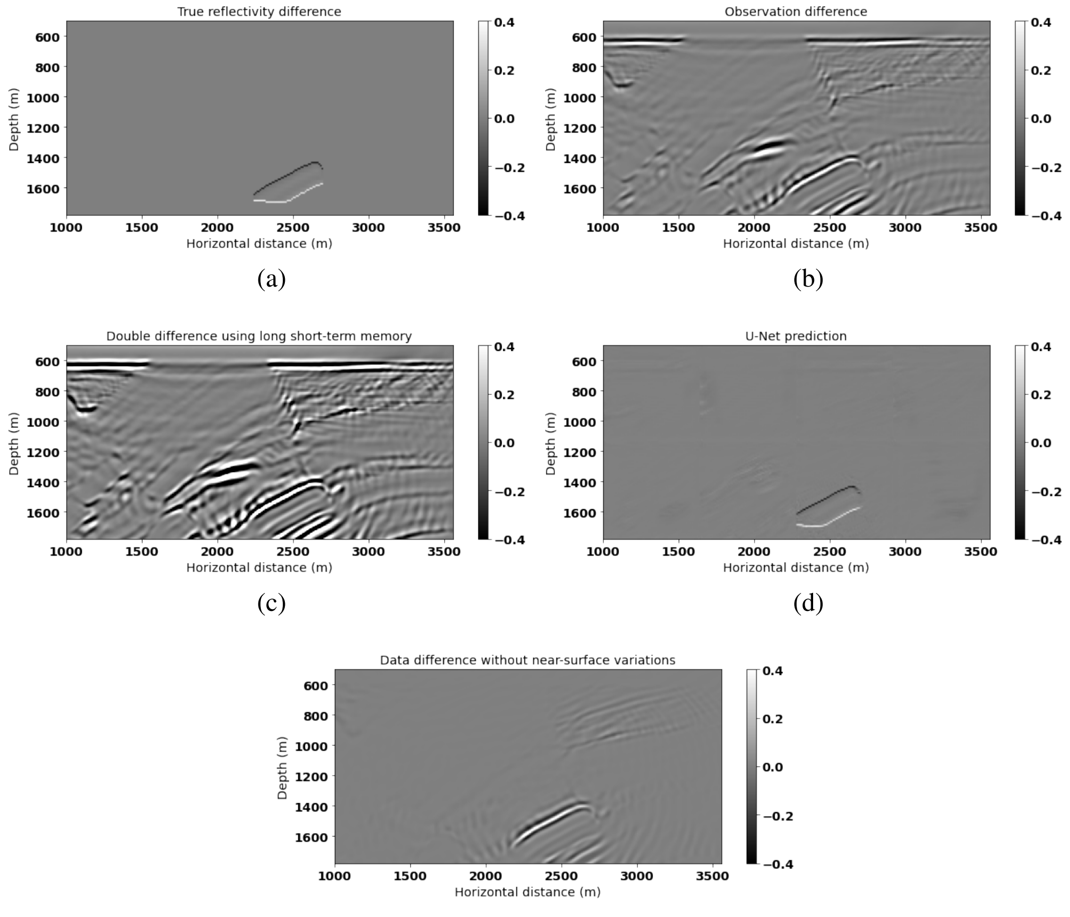
FIG. 21: Trace comparison for shots (a) No.5 and (b) No.10.

Figure 22d presents the U-Net prediction of the CO_2 injection area. The prediction reveals a clear change in the reservoir with the high resolution based on the double-difference migrated image shown in Figure 22c. The amplitude of the top boundary of the reservoir change is close to the expected target (Figure 22e), which does not account for near-surface changes. Although some noise appears above the reservoir, which could be attributed to either the U-Net learning from the noise or irrelevant reflection events, the amplitude of the predicted CO_2 injection area is more significant than the noise. In addition, the PSNR of the U-Net prediction is 33.37 dB, which is higher than the observed difference. In these results, the proposed double difference produces enhanced amplitudes for reflecting reservoir anomalies. Additionally, this U-Net-based method can effectively extract and predict CO_2 injection anomalies with high precision and accuracy, even in complex geological environments.

Table 1: PSNR (dB) comparison for examples

Migration difference	Observed	Double difference	U-Net
Horizontal-layered model	19.06	19.19	**21.95**
Curve-Fault model	20.68	21.03	**22.71**
Marmousi model	17.78	17.20	**33.37**

Table 1: PSNR (dB) comparison for synthetic examples.



(a)

(b)

(c)

(d)

(e)

FIG. 22: Marmousi example's (a) true reflectivity difference. Migration differences are generated by (b) the difference between the observed monitor and baseline, (b) double-difference method, and (d) U-Net prediction. (e) Target migration difference without near-surface change.

Overthrust model

In this section, a simulation of CO_2 injection migration was conducted on four slices of the Overthrust model (Aminadeh, 1997). A CO_2 injection site was simulated by creating velocity changes at the top of a possible trap at around 1400 meters depth, where the velocity is 1500 m/s. As we choose different slices of the 3D model, we can see the velocity anomaly taking different shapes. The objective of the tests in this subsection could be interpreted in two different ways. One is to pretend these changes in the velocity anomalies correspond to changes in the size of the injection as time passes. In other words, I use a 3D model as if it were a 2D model with historical data, and use the network to predict updated reservoir migration areas. The other interpretation is that we can have detailed information about the size of the anomaly in one section (slice) of the model, and use that information to predict the changes on a different sector (slice). Either interpretation is approximately consistent with a test where the first three slices are utilized for training a neural network, while the last slice is used for testing the pre-trained neural network in inference.

The Overthrust geology model contains various structures, including the combination of thin layers, folds and faults. I selected a trap around 1400 meters in depth with continuous top layers overhead to simulate the CO_2 injection zone. The volume and shape of this zone are customized among different slices using sine and cosine functions. To add further complexity to the simulation, I included several thin horizontal layers on top, crossing curved layers to mimic seabeds. Under the design and description above, the modified example had 241×351 gridpoints in total, with a 10-meter spatial interval. There were 88 shots and 176 receivers located at 10 meters in depth, with 2000 time steps. The recorded time sampling rate was 0.0015 seconds. Four slices in this subsection are assumed to represent reservoir variation over time.

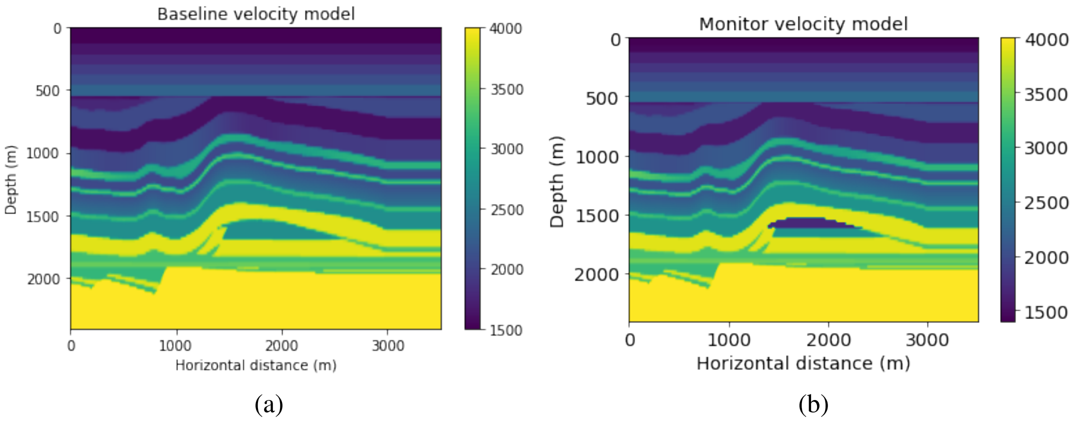
The primary aim was to assess the capability of the trained networks to generalize and predict the CO_2 injection migration in the unseen slice accurately. By training the neural networks on historical data (slices 1-3 in this case), they can be utilized to identify and forecast the potential migration of CO_2 injection over time. The neural networks can be deployed as a tool to provide insights and forecasts on how the CO_2 injection reservoir may change and evolve in the near future. The neural network's ability to recognize patterns and make predictions based on past data can be a valuable asset for anticipating future changes in the CO_2 injection migration pattern. As such, it can aid in the development of more effective strategies for CO_2 injection management and reduce the potential negative impacts and costs on the environment.

Slices 1-3

The volume of the pretended CO_2 injection area expands from slices 1 to 3. Figures 23-25 show the variation of the anomaly shape over time. The reservoir anomaly in the simulation starts as a small top within a trap and gradually expands to a larger volume with over 200 meters of thickness as time progresses. The shallow window is selected within the time range of 0-0.975 seconds, while the deep window extends from 0.450 to 1.350 seconds. It should be noted that the reflections of the reservoir change are observed after 1.5 seconds.

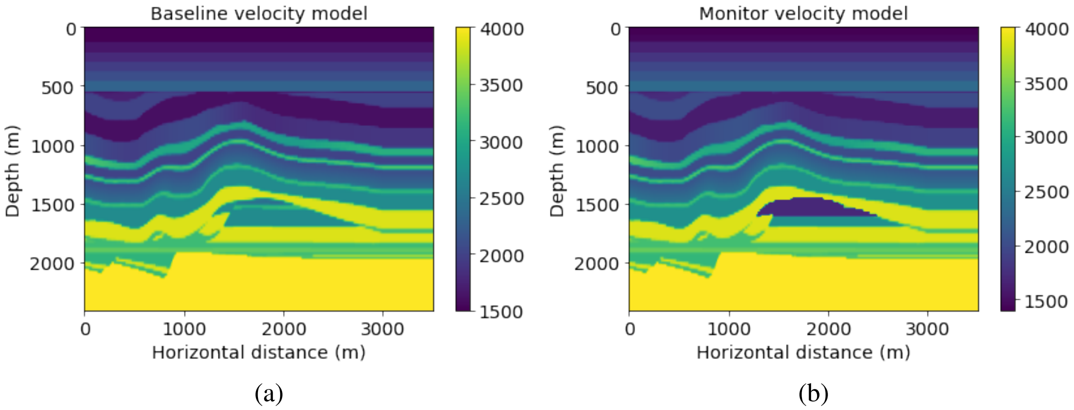
To make the neural network work on slices and predict data within a reasonable range, I propose a workflow with two separate training tasks: a) stacked LSTM for predicting shot records, and b) U-Net for forecasting migration images. However, in this example, the pre-trained stacked LSTM is used for the next sliced model to make enhancements until all the slices are processed. The first step is to let a neural network with stacked LSTM architecture to learn general information about the Overthrust model from slice 1. The input is 88 shot records of baseline data above the CO_2 injection area, and the corresponding output is two-set of predicted baseline data depending on the demand. This pre-trained network is then saved and used in slice 2 for fine-tuning. Next, the fine-tuned network is applied to slice 3 for further adjustment. The final saved network is treated as the one used for testing on slice 4. Similarly, the U-Net follows the same logic but outputs the estimation of the reflectivity difference. This example aims to assess the generalization ability of the proposed method to adapt as the subsurface structure changes either by time or spatial variations.

The following figures in this subsection present the results of the neural network ap-



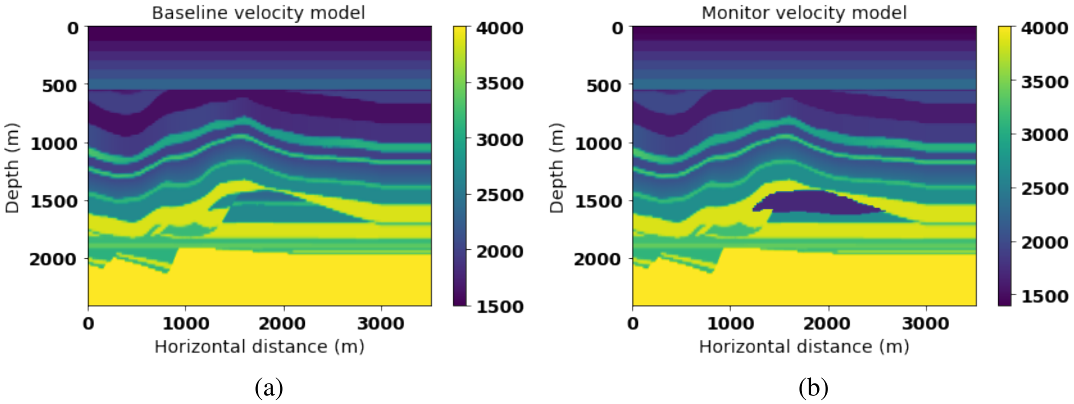
(a) (b)

FIG. 23: Velocity model for Overthrust slice 1 model (a) baseline and (b) monitor systems.



(a) (b)

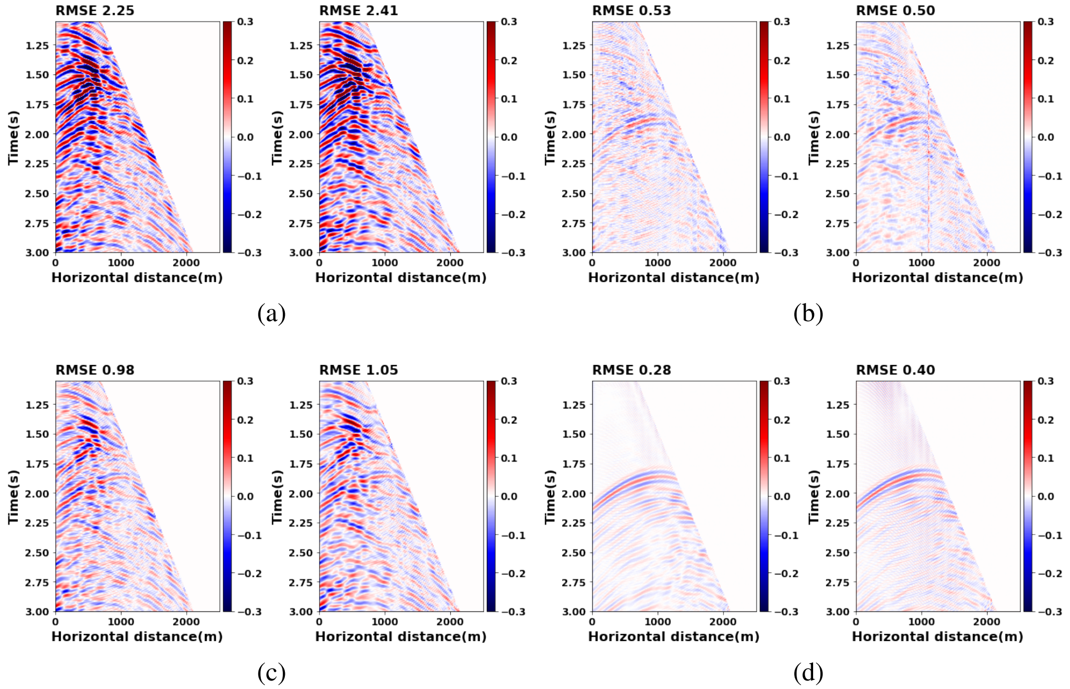
FIG. 24: Velocity model for Overthrust slice 2 model (a) baseline and (b) monitor systems.



(a) (b)

FIG. 25: Velocity model for Overthrust slice 3 model (a) baseline and (b) monitor systems.

proach using the stacked LSTM to recover reflections from the reservoir in each of the slices. Figure 26-28 display the data differences of exampled shot records. By predicting two sets of baseline shot records by shallow and deep windows, the neural network method can mitigate noise above 1.75 seconds and improve the amplitudes of the reflections. The performance of the neural network method is compared with the traditional matching filter approach, and the results show that the RMSE has dropped by an average of 43% with the neural network approach. This observation indicates that compared to using a matching filter, the neural network approach using LSTM can capture information and predict long time series with fewer artifacts.



(a) (b)

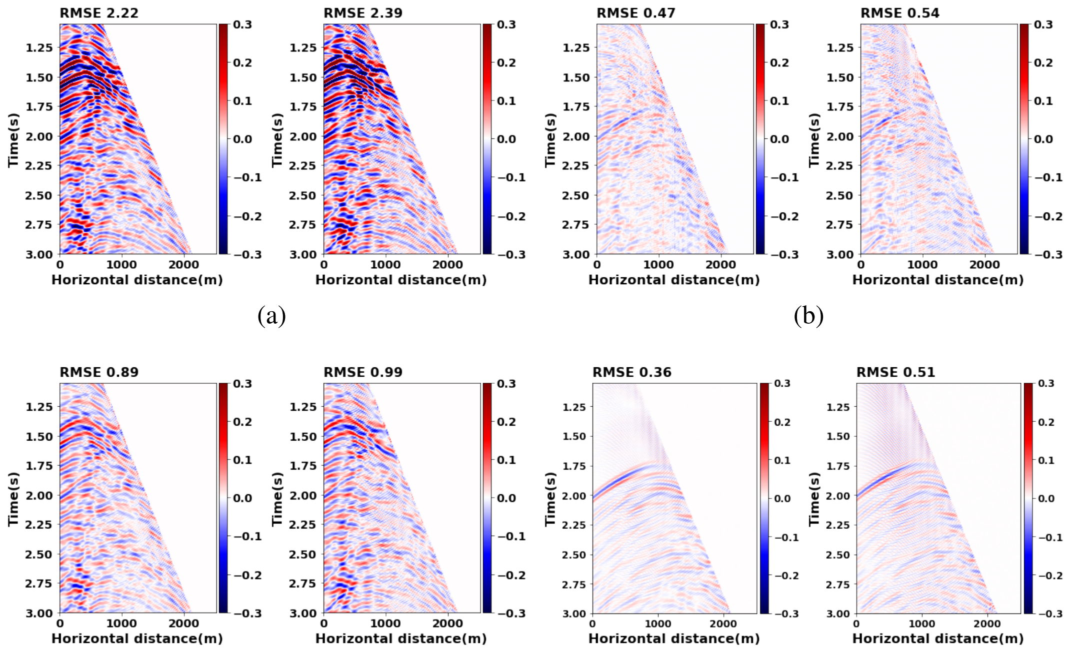
(c) (d)

FIG. 26: Data differences of slice 1 between (a) observed monitor and baseline, (b) observed monitor and predicted baseline using a deep window, (c) observed monitor and predicted monitor using a matching filter, and (d) the target difference excluding near-surface change in the observed monitor data.

The estimations of reflectivity differences are depicted in figures 29-31. Following the same observation shown before, the double-difference method works better in recovering the reflectivity differences caused by CO_2 injection, leading to increased amplitudes compared to the observation migration difference. Moreover, the output images of slices 1-3 predicted by the U-Net display cleaner output images with boosted resolution. It proves that fine-tuned neural networks are capable of handling different reservoir volumes and enhancing the quality of the output images.

Slice 4

The fine-tuned stacked LSTM and U-Net networks discussed in the previous subsection are evaluated for their generalization ability in this example. The CO_2 injection area with a curvature at the bottom, as shown in Figure 32, adds some complexity and challenges to the fine-tuned networks since shapes with large curvatures were not present in the previous slices. The training slices had slightly curved or horizontal layers. After fine-tuning the stacked LSTM network, the forecast baseline by a deep window is displayed in Figure 33b. It exhibits significant reflections from the CO_2 injection area at around 1.75 seconds, with less noise compared to the prediction after applying a matching filter (Figure



(a) (b)

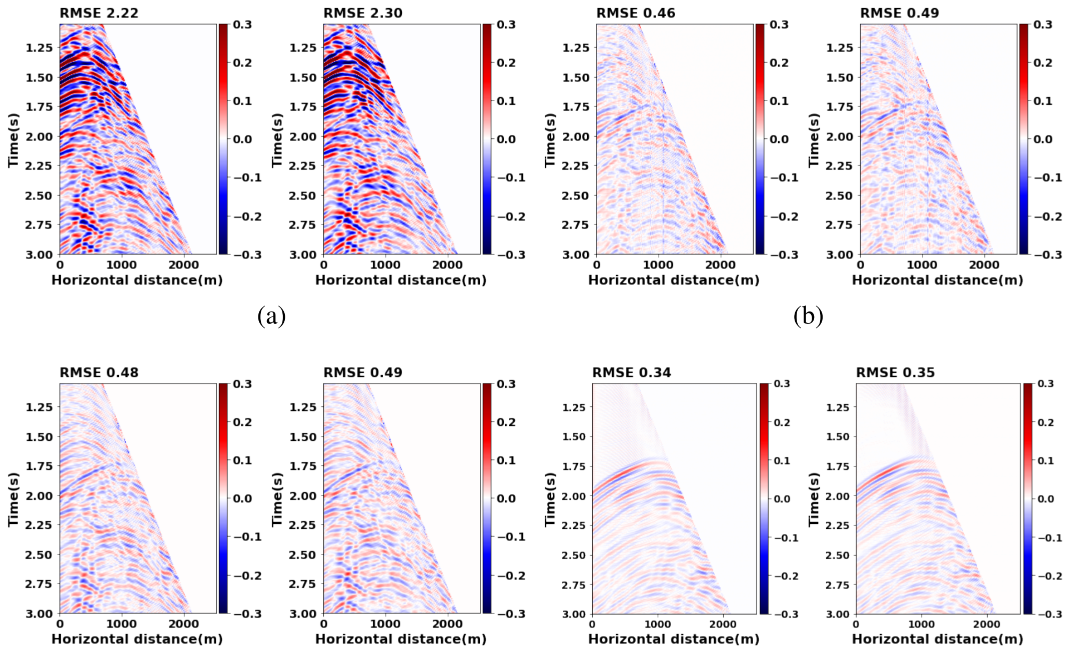
(c) (d)

FIG. 27: Data differences of slice 2 between (a) observed monitor and baseline, (b) observed monitor and predicted baseline using a deep window, (c) observed monitor and predicted monitor using a matching filter, and (d) the target difference excluding near-surface change in the observed monitor data.

33c). This observation demonstrates that the fine-tuned stacked LSTM network has a good generalization ability, and it can predict accurately in scenarios with different shapes and structures.

The outcome obtained from the double-difference method (Figure 34c) exhibits increased amplitude in the reservoir change area, which closely aligns with the target expectation shown in Figure 34e, in contrast to the observed difference (Figure 34b). The U-Net predictions of the reflectivity difference are shown in Figure 34d, demonstrating that the fine-tuned U-Net can handle the curved shape of the CO_2 injection area and provide a clean and enhanced image with refined resolution. These results prove that the proposed fine-tuned neural networks can cope with reservoir volumes with different shapes and structures, providing high-quality predictions with enhanced image resolution. This observation signifies the potential of neural networks to improve the accuracy of reservoir monitoring and management in CO_2 injection applications.

This test is done under the assumption that the weathering layer variation is slight for baseline and monitoring observations. Another assumption is that the noise level is relatively small compared with field data. It is important to note that these assumptions were made for the purpose of this experiment, and in future work, it is necessary to adjust the workflow to work with real field datasets. This can be done by incorporating techniques to handle the variability in weather layer and noise levels commonly observed in field data.



(a) (b)

(c) (d)

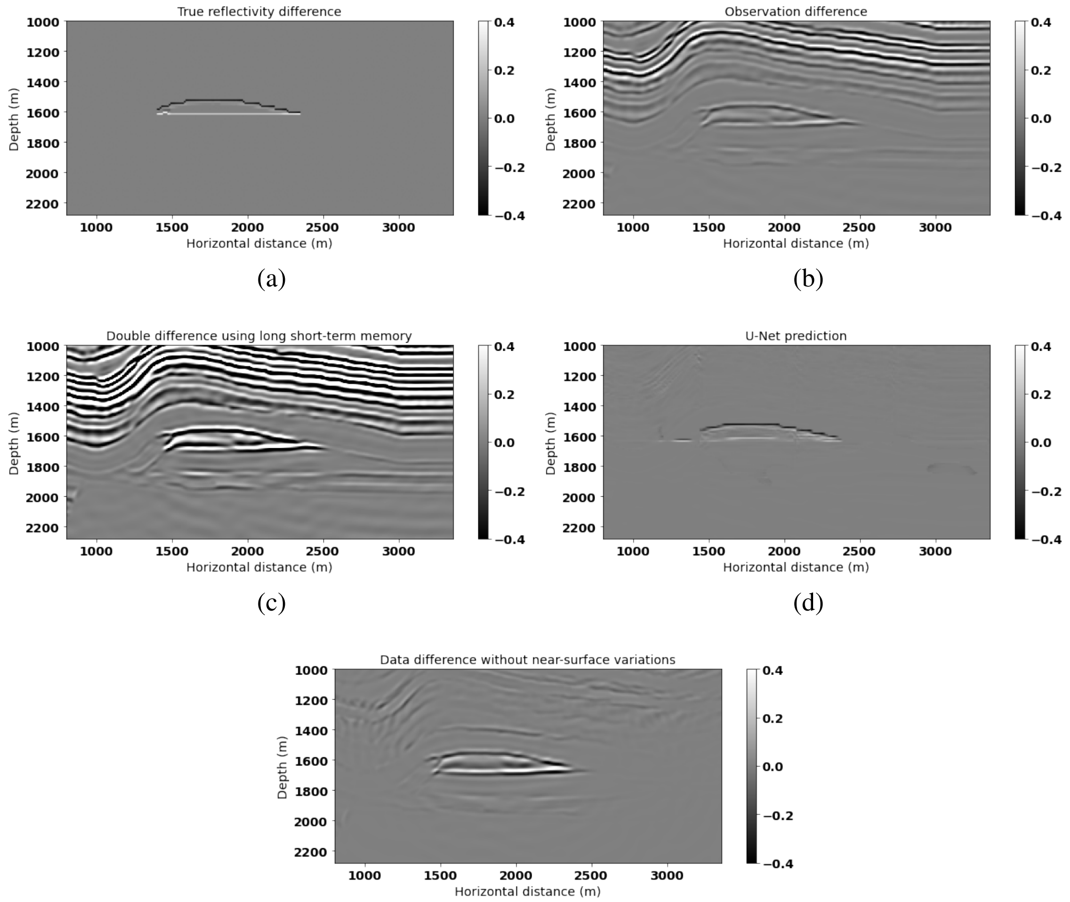
FIG. 28: Data differences of slice 3 between (a) observed monitor and baseline, (b) observed monitor and predicted baseline using a deep window, (c) observed monitor and predicted monitor using a matching filter, and (d) the target difference excluding near-surface change in the observed monitor data.

DAS VSP CaMI FRS CO_2 injection project

DAS VSP data and the geology of CaMI FRS project

In this section, I will apply the proposed method to data from a Rayleigh scattering-based distributed acoustic sensing (DAS) system. DAS data are a continuous, real-time measurement using fibre optic cables to measure axial strain. The optical fibre is the sensing element. This system can detect and cover large distances, for example, an entire length of the well, and offer an opportunity to deploy some acquisition geometries for which conventional geophone sensors cannot be used. The benefit of using the DAS system in time-lapse monitoring is that measurements can be recorded along the fibre with a very dense spatial sampling. Combining DAS and vertical seismic profiling (VSP) using densely sampled accelerometers, can monitor the growth of reservoir variations. This measurement is cost-effective and provides high resolution and easy repeatability near the wellbore. The deployment of DAS fibre cable in vertical wells is optimal for P-wave VSP surveys to monitor reservoir changes because DAS system fibre is mostly sensitive to longitudinal strain (Pevzner et al., 2022; Eaid et al., 2022; Wang and Lawton, 2022b).

The data for this experiment comes from the Containment and Monitoring Institute (CaMI) of Carbon Management Canada (CMC). CaMI has conducted a few DAS VSP surveys for the present study at the Field Research Station (FRS), in collaboration with the University of Calgary, in Newell County, Alberta, Canada (Figure 35). Time-lapse

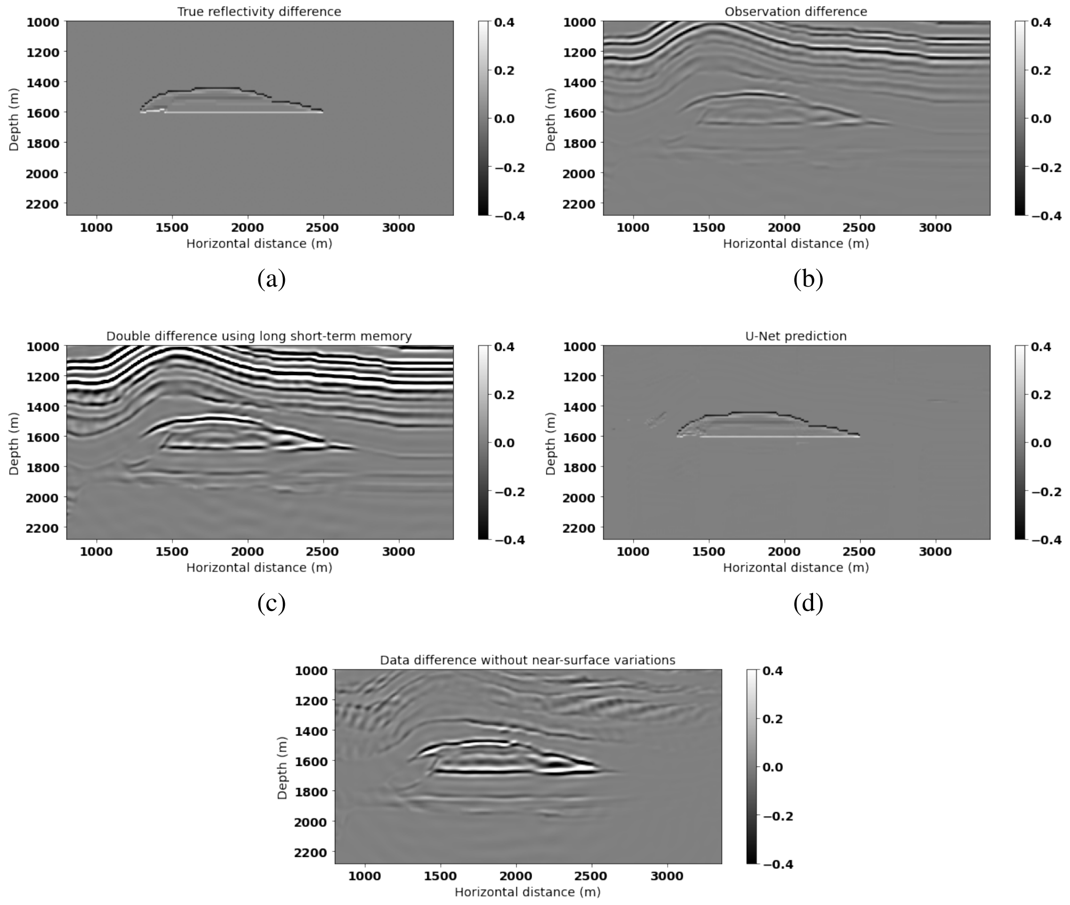


(a) (b) (c) (d) (e)

FIG. 29: Overthrust slice 1 (a) true reflectivity difference. Migration differences are generated by (b) the difference between the observed monitor and baseline, (c) the double-difference method, and (d) U-Net prediction. (e) Target migration difference without near-surface change.

DAS VSP is one of the key technologies at the FRS (Wang and Lawton, 2022b). There is one CO_2 injection well and two deep observation wells (OBS1 and OBS2) with a maximum depth of 348 m and 334 m separately at the FRS shown in Figure 36. A southwest–northeast (SW–NE) oriented horizontal trench is set close to OBS2, where DAS fibre cables with a continuous loop are located. Figure 36 also shows the shots location of the one-shot walkaround and three walkaway lines survey. The baseline and monitor data is extracted from Line 13. The fibre cables are from the top to the maximum depth of 348 m in OBS1 and 334 m in OBS2. The seismic acquisition system repeatability for Line 13 is high because of the DAS fibre cables permanent installation and the same source type (Kolkman-Quinn and Lawton, 2022; Wang and Lawton, 2022b).

A small volume of CO_2 (several tens of tonnes per year) is injected into the shallow subsurface, the Upper Cretaceous Basal Belly River Sandstone Formation (BBRS), at a depth of approximately 300 m, to simulate an unplanned CO_2 migration in a deeper and larger scale (Isaac and Lawton, 2016; Lawton et al., 2019). The goal of that project seeks



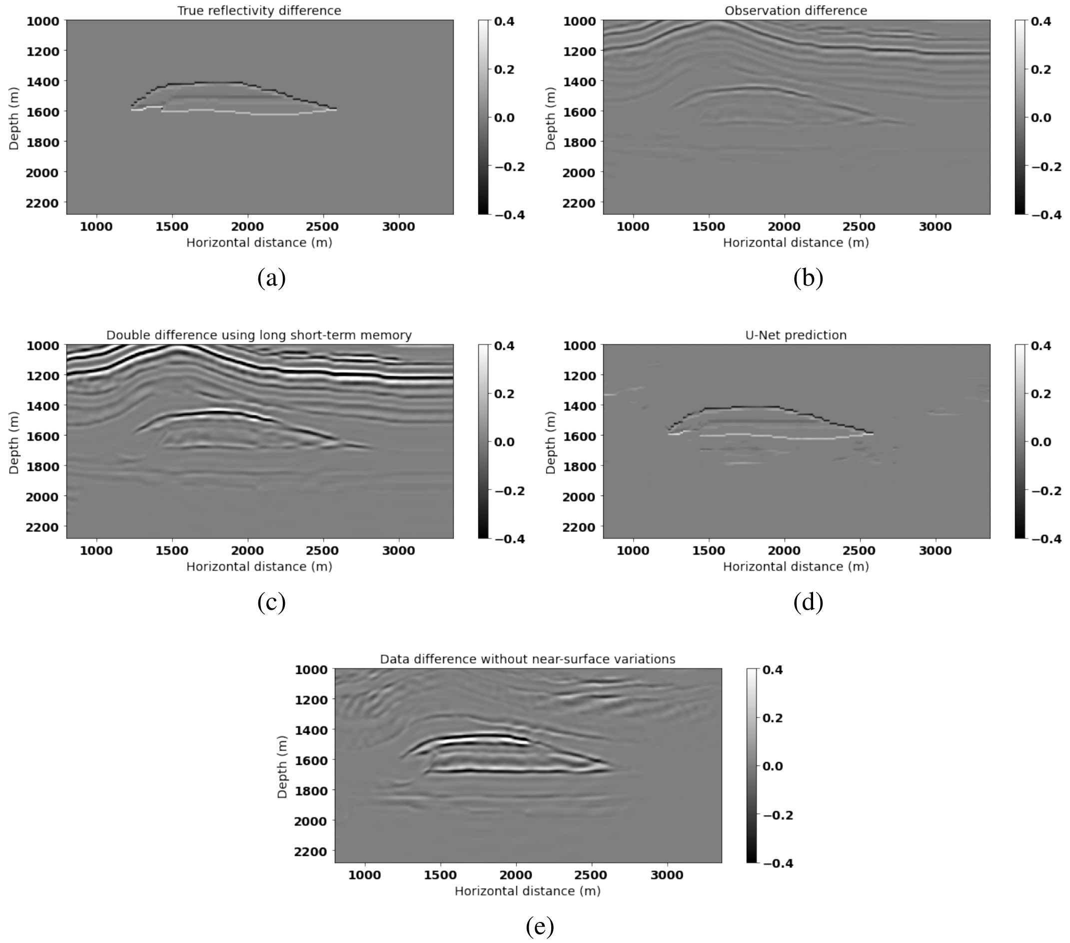
(a) (b) (c) (d) (e)

FIG. 30: Overthrust slice 2 (a) true reflectivity difference. Migration differences are generated by (b) the difference between the observed monitor and baseline, (c) the double-difference method, and (d) U-Net prediction. (e) Target migration difference without near-surface change.

for the technologies to track the growth of CO_2 plume injection and reduce the risk of potential leakage (Macquet et al., 2022). This subsection will develop a robust neural network to locate reservoir variations and can be used for further time-lapse monitoring of CO_2 injection migration.

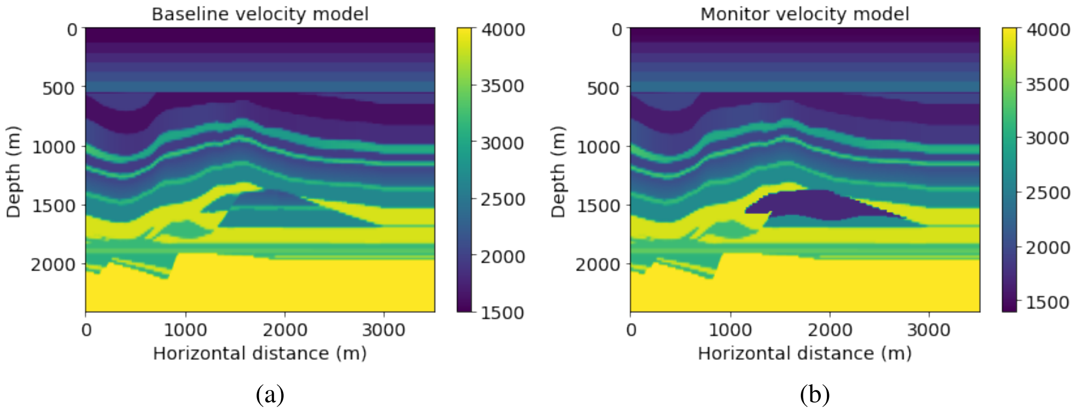
The first CO_2 gas plume injection was done in May 2017. A small volume of CO_2 of up to 20 tonnes per year has been injected into the target Basal Belly River Sandstone (BBRS) formation with a 7 m thickness. This target formation has fine- to medium-grained sandstone, composed mainly of poorly to well-sorted, angular to subangular quartz grain. The caprock is the Upper Cretaceous Foremost Formation sealing for CO_2 storage. It is a 152-m thick clayey sandstone with interbedded coal layers (Macquet et al., 2022). Then, the monitor data was acquired in March 2021.

The DAS data were processed following a sequence similar to VSP processing. According to Wang and Lawton (2022b), downgoing and upgoing wavefields from straight fi-



(a) (b) (c) (d) (e)

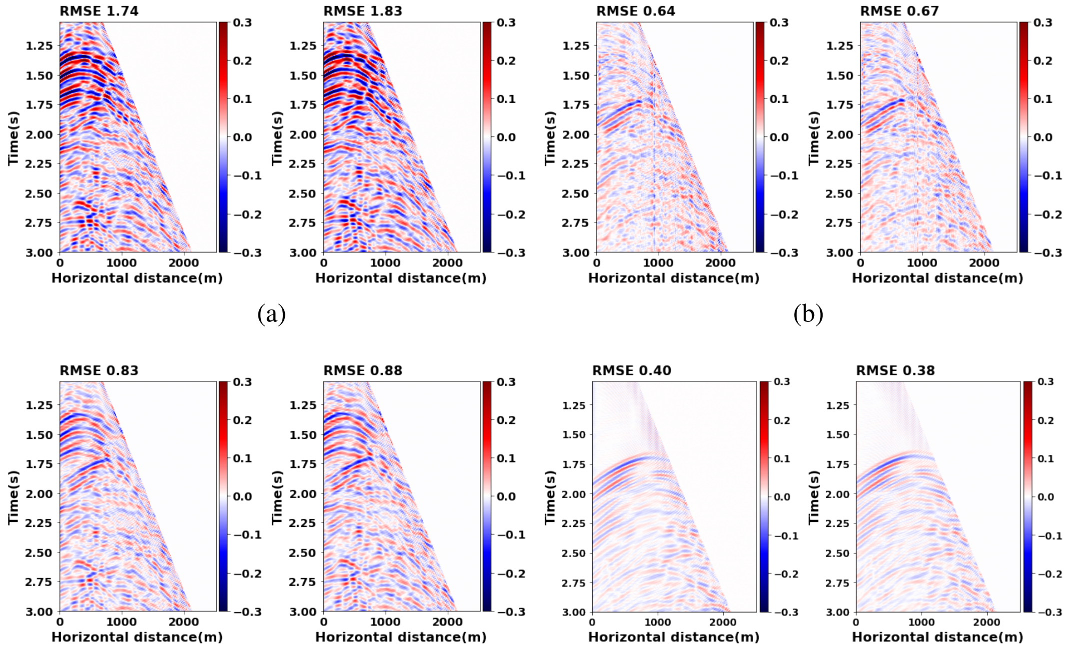
FIG. 31: Overthrust slice 3 (a) true reflectivity difference. Migration differences are generated by (b) the difference between the observed monitor and baseline, (c) the double-difference method, and (d) U-Net prediction. (e) Target migration difference without near-surface change.



(a) (b)

FIG. 32: Velocity model for Overthrust slice 4 (a) baseline and (b) monitor systems.

bre in OBS2 were separated through a median filter followed by a frequency-wavenumber



(a) (b)

(c) (d)

FIG. 33: Data differences of slice 4 between (a) observed monitor and baseline, (b) observed monitor and predicted baseline using a deep window, (c) observed monitor and predicted monitor using a matching filter, and (d) the target difference excluding near-surface change in the observed monitor data.

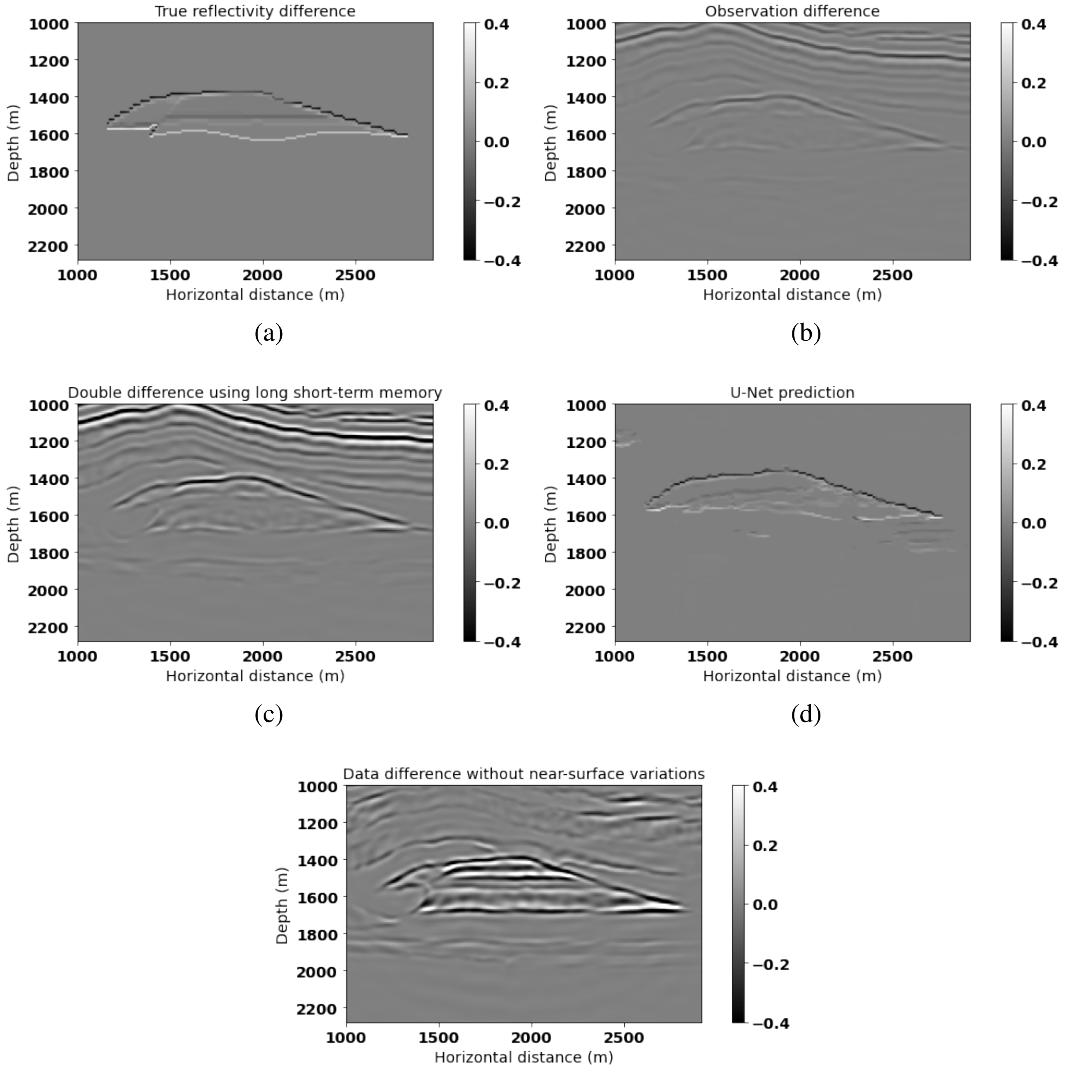
(*F*-*K*) filter (Figure 37). In this part, I apply a stacked CNN-Bi-LSTM network and the double-difference method to the processed stacked VSP data to improve CO_2 detection in the injection area.

Result

The target formation for the CaMI FRS CO_2 injection project is the Basal Belly River Sandstone (BBRS), which is located at around 260 ms in DAS VSP stacked data (Figure 38b). By applying the proposed method, significant improvements in the amplitude of the BBRS CO_2 injection zone in the DAS VSP stacks can be observed in Figure 38f. Within the BBRS injection area, reductions in the stacked data are noticeable, which can be attributed to increased CO_2 saturation leading to a decrease in seismic data amplitude.

This time-lapse response prediction aligns closely with the locations of CaMI's injection well (solid line) and southwestern observation well (dashed line). The predicted CO_2 migration area indicates that the reservoir has migrated along the target formation over the past four years, remaining within a reasonable range. Notably, the concentration zone of CO_2 is concentrated near the observation well, specifically between CDP 50 and 60, at around 260 ms.

For a comprehensive comparison, we present trace comparisons at DAS VSP CDP 50 (Figure 39a) and CDP 70 (Figure 39a) to illustrate the effectiveness of the proposed method.



(a)

(b)

(c)

(d)

(e)

FIG. 34: Overthrust slice 4 (a) true reflectivity difference. Migration differences are generated by (b) the difference between the observed monitor and baseline, (c) the double-difference method, and (d) U-Net prediction. (e) Target migration difference without near-surface change.

The blue line represents the predicted data difference, while the dashed line corresponds to the observed difference.

In Figure 39, a decrease in the amplitude of the stacked data is observed after the injection of CO_2 at a depth of 260 meters. This reduction in amplitude aligns with the expected behaviour, as the density decreases in the area where CO_2 is concentrated. The predicted data difference can capture this change. Additionally, apart from the CO_2 injection zone, a decrease in the predicted data difference is observed at around 220 and 230 meters. This indicates that the proposed method effectively mitigates the noise between baseline and monitor data, emphasizing the reservoir anomaly.

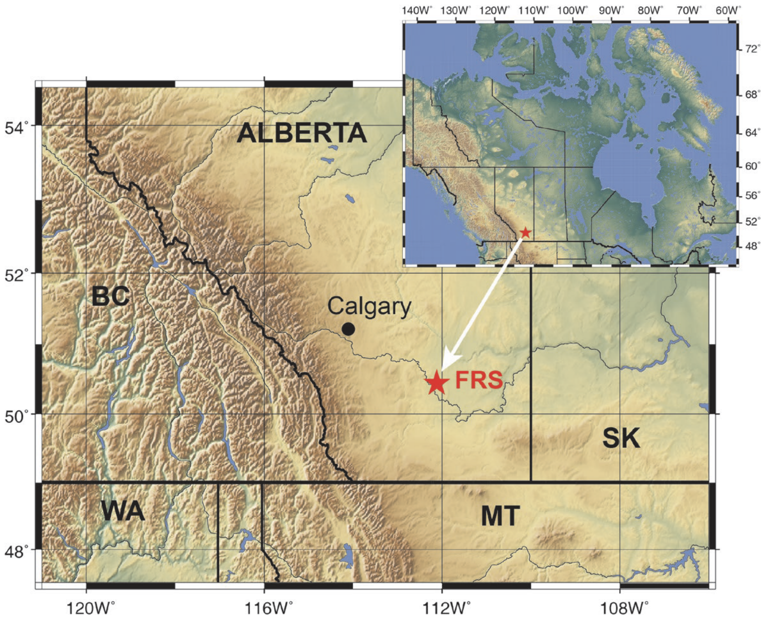


FIG. 35: CaMI FRS DAS data acquisition location map in southern Alberta, Canada (Macquet et al., 2022).

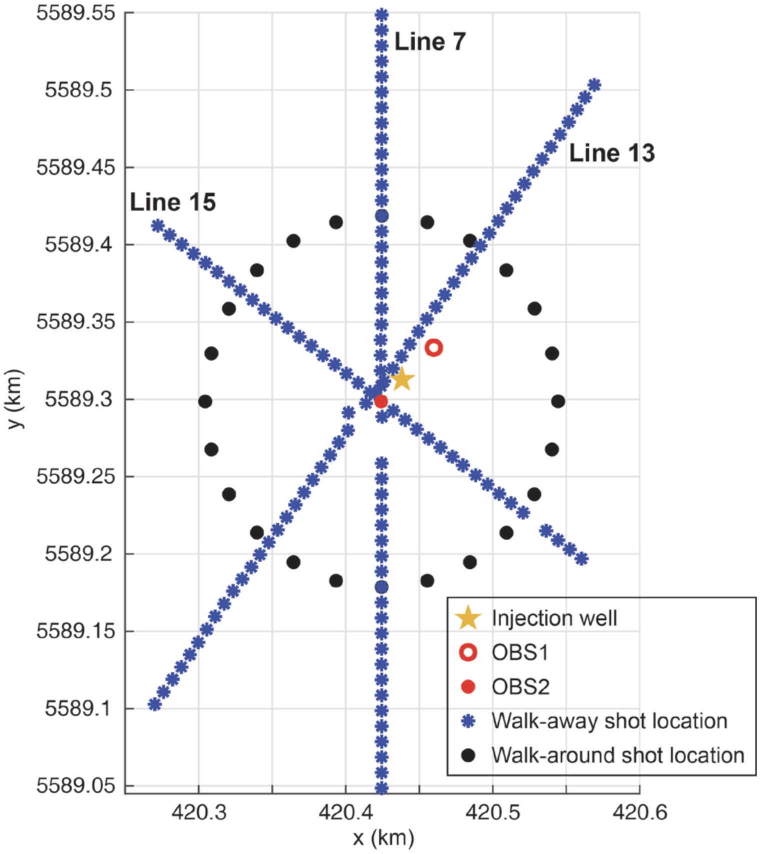


FIG. 36: VSP layout map of the survey acquired in March 2021 (Macquet et al., 2022; Wang and Lawton, 2022a).

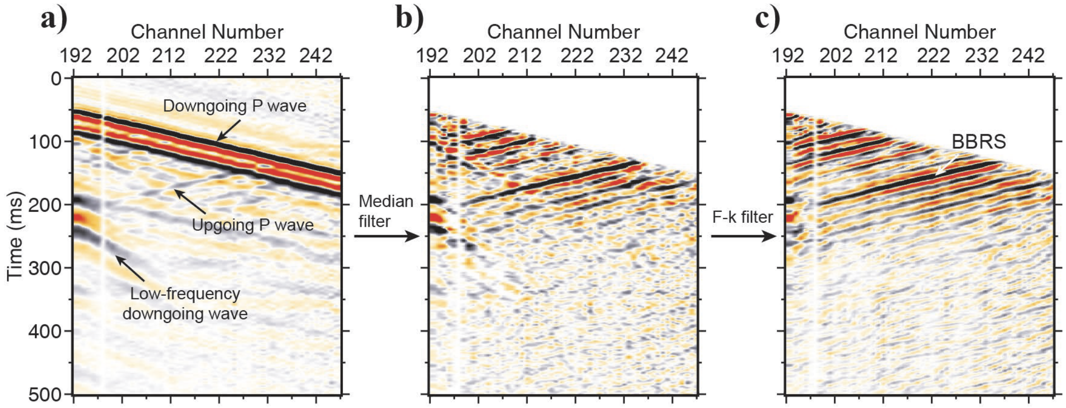


FIG. 37: Wavefield separation: (a) raw DAS data from straight fibre in OBS2, and isolated upgoing wavefield after (b) median filter and then (c) *F-K* filter. BBRS reflection is indicated in plot (c) (Wang and Lawton, 2022a).

This method serves as a valuable tool for tracking future CO_2 migration, ensuring that any potential leakage to other formations can be detected and prevented effectively.

DISCUSSION

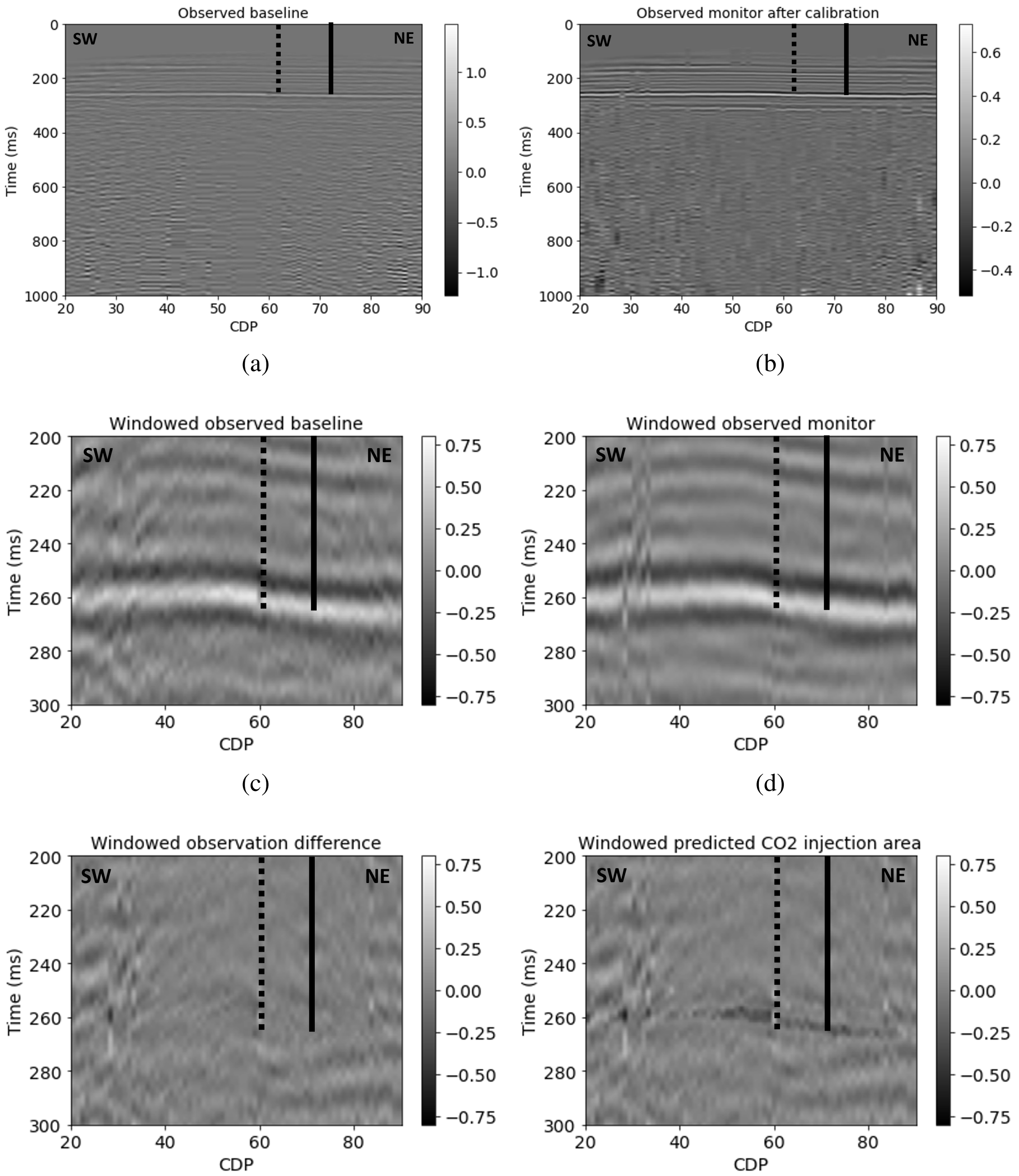
In addition to using clean traces in synthetic examples, I also examined the performance of the model when different levels of noise were added to the input traces. To be precise, I tested two levels of noise: 6% and 18%.

When 6% noise was added to the baseline data, the predicted results showed minor differences compared to using the clean traces, although the training accuracy was slightly higher. On the other hand, adding 18% noise resulted in poorer performance on the monitor data prediction. During training, the training and validation loss showed fluctuations, indicating unstable training. Although the training loss reached low values at some points, the overall instability was a concern.

These findings suggest that while the proposed method is capable of handling some levels of noise in the input data, higher levels of noise may adversely affect its performance. In future work, it would be important to investigate ways to improve the model's robustness to noise.

CONCLUSION

In this chapter, I proposed a novel approach that utilizes stacked LSTM, CNN-Bi-LSTM, and U-Net neural networks, and the double-difference method to learn changes due to the near-surface and shallow depth signal information, including multiple reflections, above the reservoir changes. By using shallow and deep windows, the method can be used in time-lapse monitoring to predict unwanted noise patterns and/or data distortions in the baseline data. The noise is then subtracted from observed monitor data using the double-difference method. Numerical examples show that this method can improve amplitudes of the reservoir change with artifact suppression, ultimately mitigating noise and improving the migration image. Additionally, fine-tuned stacked LSTM and U-Net can be



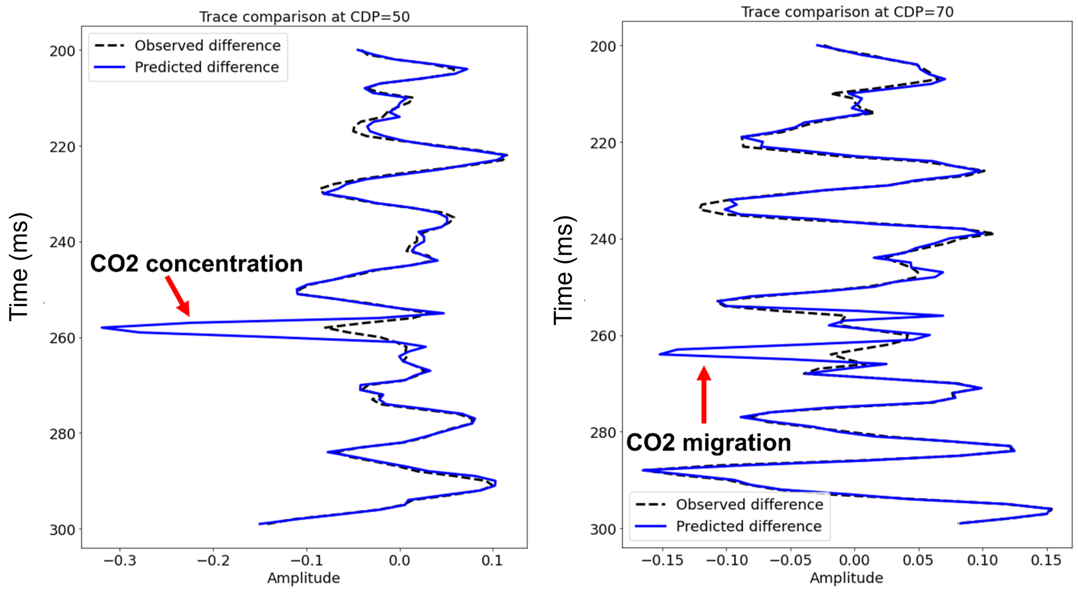
(a) (b)

(c) (d)

(e) (f)

FIG. 38: DAS VSP stacked data from CaMI FRS. (a) Observed baseline (b) Observed monitor after calibration. (c) Windowed observed baseline. (d) Windowed observed monitor. (e) Windowed observation difference. (f) Windowed predicted CO_2 injection area.

generalized to geology structures with similar patterns in the pre-training process. This implies that the pre-trained models could be reused for reservoirs with similar characteristics, where true velocity changes are unknown making it impossible to re-train the network. Moreover, this method also works for real DAS data that redirect the possible area for CO_2 injection migration. Future work can focus on extending the method to a broader range of complex geology models for both training and testing and exploring the transferability of pre-trained models to different geological structures and evaluating their performance.



(a) (b)

FIG. 39: Trace comparison for DAS VSP data CDP (a) 50 and (b) 70.

By testing this neural network architecture and workflow in various scenarios, the model can be further generalized to other subsurface imaging situations. Additionally, exploring the use of alternative neural network architectures and incorporating additional sources of information, such as well data, can provide further improvements to this approach.

ACKNOWLEDGEMENT

The sponsors of CREWES are gratefully thanked for continued support. This work was funded by CREWES industrial sponsors, NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 543578-19.

One of the authors of this report was supported by China Scholarship Council (CSC).

REFERENCES

Alali, A., Kazei, V., Sun, B., and Alkhalifah, T., 2022, Time-lapse data matching using a recurrent neural network approach: Geophysics, **87**, No. 5, 1–83.

Aleardi, M., 2022, Elastic properties and litho-fluid facies estimation from pre-stack seismic data through bidirectional long short-term memory networks: Geophysical Prospecting, **70**, No. 3, 558–577.

Alfarraj, M., and AlRegib, G., 2019, Semisupervised sequence modeling for elastic impedance inversion: Interpretation, **7**, No. 3, SE237–SE249.

Aminadeh, F., 1997, Seg/eaga 3-d salt and overthrust models: SEG/EAGE 3-D Modeling Serise No. 1, SEG.

Arts, R., Eiken, O., Chadwick, A., Zweigel, P., Van der Meer, L., and Zinszner, B., 2003, Monitoring of co2 injected at sleipner using time lapse seismic data, *in* Greenhouse Gas Control Technologies-6th International Conference, Elsevier, 347–352.

Asnaashari, A., Brossier, R., Garambois, S., Audebert, F., Thore, P., and Virieux, J., 2015, Time-lapse seismic imaging using regularized full-waveform inversion with a prior model: which strategy?: Geophysical prospecting, **63**, No. 1, 78–98.

Ayeni, G., and Biondi, B., 2010, Target-oriented joint least-squares migration/inversion of time-lapse seismic data sets: Geophysics, **75**, No. 3, R61–R73.

Bergmann, P., Kashubin, A., Ivandic, M., Lüth, S., and Juhlin, C., 2014, Time-lapse difference static correction using prestack crosscorrelations: 4d seismic image enhancement case from ketzin: Geophysics, **79**, No. 6, B243–B252.

Calderón-Macías, C., Sen, M. K., and Stoffa, P. L., 2000, Artificial neural networks for parameter estimation in geophysics [link]: Geophysical prospecting, **48**, No. 1, 21–47.

Chadwick, A., Williams, G., Delepine, N., Clochard, V., Labat, K., Sturton, S., Buddensiek, M.-L., Dillen, M., Nickel, M., Lima, A. L. et al., 2010, Quantitative analysis of time-lapse seismic monitoring data at the sleipner co 2 storage operation: The Leading Edge, **29**, No. 2, 170–177.

Cui, Z., Ke, R., Pu, Z., and Wang, Y., 2018, Deep bidirectional and unidirectional lstm recurrent neural network for network-wide traffic speed prediction: arXiv preprint arXiv:1801.02143.

Das, V., Pollack, A., Wollner, U., and Mukerji, T., 2019, Convolutional neural network for seismic impedance inversion: Geophysics, **84**, No. 6, R869–R880.

Deng, C., Feng, S., Wang, H., Zhang, X., Jin, P., Feng, Y., Zeng, Q., Chen, Y., and Lin, Y., 2021, OpenFWI: large-scale multi-structural benchmark datasets for seismic full waveform inversion: arXiv preprint arXiv:2111.02926.

Du, G., Wang, Z., Gao, B., Mumtaz, S., Abualnaja, K. M., and Du, C., 2020, A convolution bidirectional long short-term memory neural network for driver emotion recognition: IEEE Transactions on Intelligent Transportation Systems.

Eaid, M., Keating, S., and Innanen, K., 2022, Combined elastic fwi of accelerometer and das vsp data from a co2 sequestration test site in newell county, alberta, *in* SEG/AAPG International Meeting for Applied Geoscience & Energy, OnePetro.

Fu, X., and Innanen, K. A., 2022, Time-lapse seismic imaging using shot gathers with non-repeatable source wavelets: Geophysics, **88**, No. 1, 1–108.

Graves, A., Jaitly, N., and Mohamed, A.-r., 2013a, Hybrid speech recognition with deep bidirectional lstm, *in* 2013 IEEE workshop on automatic speech recognition and understanding, IEEE, 273–278.

Graves, A., Mohamed, A.-r., and Hinton, G., 2013b, Speech recognition with deep recurrent neural networks, *in* 2013 IEEE international conference on acoustics, speech and signal processing, Ieee, 6645–6649.

Graves, A., and Schmidhuber, J., 2005, Framewise phoneme classification with bidirectional lstm and other neural network architectures: Neural networks, **18**, No. 5-6, 602–610.

Guo, R., Zhang, J., Liu, D., Zhang, Y., and Zhang, D., 2019, Application of bi-directional long short-term memory recurrent neural network for seismic impedance inversion, *in* 81st EAGE Conference and Exhibition 2019, vol. 2019, European Association of Geoscientists & Engineers, 1–5.

Henley, D. C., and Lawton, D. C., 2021, Time-lapse detection using raypath interferometry: Geophysics, **86**, No. 2, Q27–Q36.

Hochreiter, S., and Schmidhuber, J., 1997, Long short-term memory: Neural computation, **9**, No. 8, 1735–1780.

Hussein, M., Stewart, R. R., Sacrey, D., Johnston, D. H., and Wu, J., 2021, Unsupervised machine learning for time-lapse seismic studies and reservoir monitoring: Interpretation, **9**, No. 3, T791–T807.

Isaac, H. J., and Lawton, D. C., 2016, Brooks revisited: CREWES Research Report, **28**, No. 33.

Isaac, J. H., and Lawton, D. C., 2006, A case history of time-lapse 3d seismic surveys at cold lake, alberta, canada: Geophysics, **71**, No. 4, B93–B99.

Isaac, J. H., and Lawton, D. C., 2014, A case history of experimental time-lapse 3c 2d seismic reflection data for reservoir monitoring at cold lake, alberta, canada: Interpretation, **2**, No. 2, SE47–SE54.

Jordan, M., 1986, Serial order: a parallel distributed processing approach. technical report, june 1985-march 1986, Tech. rep., California Univ., San Diego, La Jolla (USA). Inst. for Cognitive Science.

Kolkman-Quinn, B. J., and Lawton, D. C., 2022, Detection threshold of a shallow co2 plume with vsp data from the cami field research station, *in* SEG/AAPG International Meeting for Applied Geoscience & Energy, OnePetro.

Koster, K., Gabriels, P., Hartung, M., Verbeek, J., Deinum, G., and Staples, R., 2000, Time-lapse seismic surveys in the north sea and their business impact: The Leading Edge, **19**, No. 3, 286–293.

Lawton, D. C., Dongas, J., Osadetz, K., Saeedfar, A., and Macquet, M., 2019, Development and analysis of a geostatic model for shallow co2 injection at the field research station, southern alberta, canada: Geophysics and Geosequestration. Cambridge University Press, Cambridge, **280**, 296.

Li, D., Peng, S., Guo, Y., Lu, Y., and Cui, X., 2021, Co2 storage monitoring based on time-lapse seismic data via deep learning: International Journal of Greenhouse Gas Control, **108**, 103,336.

Li, Y., and Alkhalifah, T., 2022, Target-oriented time-lapse elastic full-waveform inversion constrained by deep learning-based prior model: IEEE Transactions on Geoscience and Remote Sensing, **60**, 1–12.

Lumley, D. E., 2001, Time-lapse seismic reservoir monitoring: Geophysics, **66**, No. 1, 50–53.

Macquet, M., Lawton, D., Osadetz, K., Maidment, G., Bertram, M., Hall, K., Kolkman-Quinn, B., Parra, J. M., Race, F., Savard, G. et al., 2022, Overview of carbon management canada's pilot-scale co2 injection site for developing and testing monitoring technologies for carbon capture and storage, and methane detection: CSEG Recorder, **47**, 1–27.

Moya, A., and Irikura, K., 2010, Inversion of a velocity model using artificial neural networks: Computers & geosciences, **36**, No. 12, 1474–1483.

Pennington, W. D., 2000, "do no harm!"–seismic petrophysical aspects of time-lapse monitoring, *in* 2000 SEG Annual Meeting, OnePetro.

Pevzner, R., Glubokovskikh, S., Isaenkov, R., Shashkin, P., Tertyshnikov, K., Yavuz, S., Gurevich, B., Correa, J., Wood, T., and Freifeld, B., 2022, Monitoring subsurface changes by tracking direct-wave amplitudes and traveltimes in continuous distributed acoustic sensor vsp datamonitoring by traveltime and amplitude: Geophysics, **87**, No. 1, A1–A6.

Pham, N., and Wu, X., 2019, Missing sonic log prediction using convolutional long short-term memory, *in* SEG International Exposition and Annual Meeting, OnePetro.

Rickett, J., and Lumley, D., 2001, Cross-equalization data processing for time-lapse seismic reservoir monitoring: A case study from the gulf of mexico: Geophysics, **66**, No. 4, 1015–1025.

Ronneberger, O., Fischer, P., and Brox, T., 2015, U-net: Convolutional networks for biomedical image segmentation, *in* International Conference on Medical image computing and computer-assisted intervention, Springer, 234–241.

Roy, P., Zhu, X., and Fei, W., 2020, Machine learning assisted seismic inversion, *in* SEG International Exposition and Annual Meeting, OnePetro.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J., 1985, Learning internal representations by error propagation, Tech. rep., California Univ San Diego La Jolla Inst for Cognitive Science.

Schuster, M., and Paliwal, K. K., 1997, Bidirectional recurrent neural networks: IEEE transactions on Signal Processing, **45**, No. 11, 2673–2681.

Versteeg, R., 1994, The marmousi experience: Velocity model determination on a synthetic complex data set: The Leading Edge, **13**, No. 9, 927–936.

Wang, Y., and Lawton, D. C., 2022a, Acoustic and exact elastic impedance variations during co2 injection at the cami.frs: CREWES Research Report, **34**, No. 52.

Wang, Y., and Lawton, D. C., 2022b, Time-lapse attenuation-attribute variations during co2 injection using das vsp data from the cami.frs: CREWES Research Report, **34**, No. 53.

Wang, Y., and Morozov, I. B., 2020, Time-lapse acoustic impedance variations during CO2 injection in Weyburn oilfield, canada: Geophysics, **85**, No. 1, M1–M13.

Wang, Z., 1997, Feasibility of time-lapse seismic reservoir monitoring: The physical basis: The Leading Edge, **16**, No. 9, 1327–1330.

Wapenaar, K., and Van Ijsseldijk, J., 2021, Employing internal multiples in time-lapse seismic monitoring, using the Marchenko method, *in* 82nd EAGE Annual Conference & Exhibition, vol. 2021, European Association of Geoscientists & Engineers, 1–5.

Watanabe, T., Shimizu, S., Asakawa, E., and Matsuoka, T., 2004, Differential waveform tomography for time-lapse crosswell seismic data with application to gas hydrate production monitoring, *in* SEG Technical Program Expanded Abstracts 2004, Society of Exploration Geophysicists, 2323–2326.

Yoon, D., Yeeh, Z., and Byun, J., 2020, Seismic data reconstruction using deep bidirectional long short-term memory with skip connections: IEEE Geoscience and Remote Sensing Letters, **18**, No. 7, 1298–1302.

Yuan, C., Zhang, X., Jia, X., and Zhang, J., 2020, Time-lapse velocity imaging via deep learning: Geophysical Journal International, **220**, No. 2, 1228–1241.

Zhang, R., Ghosh, R., Sen, M. K., and Srinivasan, S., 2013, Time-lapse surface seismic inversion with thin bed resolution for monitoring co2 sequestration: A case study from cranfield, mississippi: International Journal of Greenhouse Gas Control, **18**, 430–438.

Zhong, Z., Sun, A. Y., and Wu, X., 2020, Inversion of time-lapse seismic reservoir monitoring data using cyclegan: a deep learning-based approach for estimating dynamic reservoir property changes: Journal of Geophysical Research: Solid Earth, **125**, No. 3, e2019JB018,408.

Zhou, R., Huang, L., and Rutledge, J., 2010, Microseismic event location for monitoring co 2 injection using double-difference tomography: The Leading Edge, **29**, No. 2, 208–214.