

Lawrence Berkeley National Laboratory

LBL Publications

Title

3D Deep Learning Joint Inversion of Active Seismic Full Waveform and Passive Seismic Traveltime Data for Reservoir Imaging and Uncertainty Quantification

Permalink

<https://escholarship.org/uc/item/1805s6mv>

Journal

Geophysical Prospecting, 74(1)

ISSN

0016-8025

Authors

Um, Evan Schankee

Alumbaugh, David

Wang, Hanchen

et al.

Publication Date

2026

DOI

10.1111/1365-2478.70126

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

3D Deep Learning Joint Inversion of Active Seismic Full Waveform and Passive Seismic Traveltime Data for Reservoir Imaging and Uncertainty Quantification

The List of Authors:

Evan Schankee Um, Earth and Environmental Sciences, Lawrence Berkeley National Laboratory,
Berkeley, CA 94702, USA

David Alumbaugh, Earth and Environmental Sciences, Lawrence Berkeley National Laboratory,
Berkeley CA 94702, USA

Hanchen Wang, University of North Carolina-Chapel Hill, School of Data Science and Society, 211
Manning Drive, Chapel Hill, NC 27599, USA

Youzuo Lin, University of North Carolina-Chapel Hill, School of Data Science and Society, 211
Manning Drive, Chapel Hill, NC 27599, USA

ABSTRACT

We present deep learning (DL) networks for three-dimensional (3D) joint inversion of active seismic full waveform and passive seismic traveltime data to image reservoirs and their properties and quantify imaging uncertainties. Active seismic full-waveform data can provide high-resolution monitoring images but are collected only intermittently because of their high acquisition cost. In contrast, passive seismic data can be gathered at relatively low cost between regular active surveys, although their imaging quality can be compromised by factors such as low signal-to-noise ratios and limited ray coverage of the target. While these datasets are routinely acquired together at CO₂ storage sites, their combined inversion within a 3D DL framework has not been previously demonstrated. To our knowledge, this is the first study to address this gap, combining the strength of both data types. For efficient data storage and DL training with large 3D seismic datasets, we use a 3D data matrix in which a random number of passive seismic traveltime data are stored as parabolic envelopes using one-hot encoding, and a 3D full waveform data matrix in which multiple shot gathers are summed. Two network architectures are evaluated: a single-encoder U-Net for single data type inversion and a dual-encoder U-Net for joint inversion of active and passive seismic data. We also evaluate the single-encoder U-Net for joint inversion by concatenating full waveform data and traveltime data. We propose a systematic approach for selecting an optimal dropout rate that balances regularization during training and Monte Carlo dropout-based uncertainty quantification during prediction by

examining the correlation coefficient between standard deviation and prediction error, along with the training misfit, across a range of dropout rates. 3D DL inversion experiments include five different network configurations, with evaluations under ideal, noisy, and dropout-enabled conditions. Both model and data uncertainties are assessed, as well as their combined effects. Across all conditions, the networks consistently predict accurate CO₂ saturation models with low prediction errors such as structural similarity index of 0.993 and CO₂ difference of 1.1%. Uncertainty estimates show strong spatial correlation with prediction errors, confirming the effectiveness of the proposed dropout selection approach. The results demonstrate that our DL approach, utilizing compact data representations and appropriate uncertainty quantification, yields accurate subsurface images under various inversion conditions and provides valuable insights into the reliability of predictions.

INTRODUCTION

Physics-based seismic full waveform inversion (FWI) aims to find subsurface velocity and other geophysical property models that best explain the recorded seismic full waveform data by minimizing the misfit between the recorded and predicted data (Virieux and Operto, 2009; Fichtner, 2010; Virieux et al., 2017). FWI has been widely used in a range of applications such as energy resource exploration, development and monitoring of geological storage, carbon capture and storage, subsurface characterization for waste disposal, and earthquake hazard assessment and many others (e.g., Kallivokas et al., 2013; Queier and Singh, 2013; Hicks et al., 2016). Compared to other geophysical inversions such as diffusion-physics-based electromagnetic data inversion and potential-physics-based gravity, magnetic, and electrical data inversion, FWI is based on wave physics phenomena, utilizes full waveform information, and can therefore delineate subsurface structures at higher resolution.

Nonetheless, physics-based FWI is numerically challenging from multiple standpoints (Virieux and Operto, 2009; Shin and Cha, 2009; Petrov and Newman, 2014; Um et al. 2014; Lin et al., 2023; Zhu et al., 2022). For example, FWI requires a large amount of computational resources and time to repeatedly solve its governing equations (e.g. acoustic and elastic wave equations) throughout the iterative inversion process. The computational cost becomes significantly high when 3D FWI is considered. This enormous computational burden often leads to using partial waveform data, such as travel time data, by solving a simplified version of the governing equation, such as the eikonal equation. Furthermore, the objective function of FWI is highly non-linear and typically requires a good initial model, which is not usually available.

Deep-learning (DL) based full-waveform inversion (DL-FWI) has the potential to overcome several limitations of traditional physics-driven FWI. By training on large ensembles of realistic velocity models and their corresponding full waveform data, a DL network learns complex, nonlinear correlations between the waveform-data space and the model space and is therefore already informed about plausible subsurface structures when performing inversion (Araya-Polo et al., 2018; Yang and Ma, 2019; Wu and Lin, 2019; Zeng et al., 2021; Lin et al., 2023; Zhu et al., 2022; Dai et al., 2023; Um et al., 2023a). The training models and data serve as effective regularizers that guide the solution away from cycle skipping and other local minima that often hamper physics-based FWI, particularly when low-frequency data are absent. Exploiting the nonlinear relationships embedded in the training set, DL-FWI can recover high-wavenumber details that physics-only updates may miss, producing sharper images while delivering predictions in near real time and without the need for an accurate starting model. For example, using the 2D Kimberlina model set, Rojas-Gómez et al. (2020) systematically compared the ground-truth velocity models with those obtained from physics-based FWI and DL-based FWI. They showed that DL-based FWI (mean absolute error ≈ 0.01 ; structural similarity ≈ 0.99) characterizes small CO₂ leakage plumes at the Kimberlina sequestration site more accurately than traditional physics-based FWI (mean absolute error ≈ 0.07 ; structural similarity ≈ 0.97). Similar improvements were noted by Zeng et al. (2021) for the 3D Kimberlina models, and by Feng et al. (2022) using both synthetic and field data. For a detailed comparison, readers are referred to those papers and the references therein.

DL-FWI, however, introduces its own challenge. A wide range of realistic velocity models and corresponding synthetic full waveform data must be generated before training (Um et al., 2023a; Jin et al., 2024). Without such a wide coverage, the network may overfit to a narrow range of conditions and fail to generalize to unseen field data, producing unreliable images. Creating a wide range of training sets can be much more burdensome than performing a conventional physics-based FWI on the observations themselves, particularly in frontier areas with limited prior geological information. In contrast, DL-FWI becomes attractive for geological carbon-storage sites and other regulated facilities where baseline geological characterization information is already mandated (Alumbaugh et al., 2024a). That baseline information can then be used to build realistic, site-specific models, simulate their geophysical responses, and train DL-FWI. Once trained, the network can provide near-real-time updates of subsurface properties, enabling operators to manage CO₂ injection more efficiently and safely while continuously tracking plume evolution.

Multiple types of geophysical data are often acquired together at a CO₂ storage site, and recent research has increasingly focused on developing DL joint inversion methods to utilize these complementary datasets together. Early efforts demonstrated the value of this approach in 2D. For

instance, Um et al. (2023^b) showed that a 2D DL network integrating seismic, electromagnetic, and gravity data improves imaging accuracy, while Hu et al. (2023) developed a framework for the joint inversion of seismic traveltimes and DC resistivity data. More recently, Guo presents (2025) a 2D U-Net based joint inversion framework that autonomously learns correlations among subsurface properties, integrates multi-physics geophysical data without predefined relationships or extensive prior information, and is validated with synthetic electromagnetic, seismic, and gravity examples. Building on this foundation, the field has begun to advance into 3D joint inversion. For example, Celaya and Araya-Polo (2024) introduced a 3D DL framework that jointly inverts time-lapse surface gravity and seismic full waveform data to monitor CO₂ plume evolution. However, despite this progress, a critical gap remains: the joint inversion of active-source full waveform and passive-source traveltimes data, two of the most common seismic datasets in CO₂ monitoring, has not yet been explored within a 3D DL context.

In this paper, we present 3D DL-based inversion networks for single and joint data types (i.e. active seismic full waveform data vs. passive seismic travel time data vs. their combined use) and apply them to imaging CO₂ saturation and quantifying associated uncertainties. In this paper we introduce 3D deep-learning inversion networks for individual and joint data types (i.e. active seismic full-waveform, passive seismic traveltimes, and their combination) and apply them to image CO₂ saturation and quantify the associated uncertainty. The joint inversion of active and passive seismic data is motivated by their complementary strengths. Active full-waveform surveys provide high-resolution information on subsurface properties but are acquired only intermittently (for example, every five years) because of their high cost (Alumbaugh et al., 2024a). Passive seismic data can be collected more often and at much lower cost between active surveys, thereby supplementing them. However, passive data may not always effectively image the CO₂ plume. For example, there may be too few microseismic events, their signal-to-noise ratios may be low, ray paths might miss the plume, or event locations may not correlate with plume migration. Combining the two data types therefore offers a more reliable and cost-effective means of imaging and monitoring CO₂ storage.

The networks presented here are based on the 3D U-Net architecture (Ronneberger et al., 2015; Çiçek et al. 2016) and its modified version adapted for joint data types. We train and evaluate the networks using 3D seismic full waveform data and passive seismic travel time data, both simulated over velocity models (Alumbaugh et al., 2024^b) derived from realistic 3D brine reservoir models (Zhou and Birkholzer, 2011; Wainwright, et al, 2013). To our knowledge, this is the first study to jointly invert 3-D full-waveform and passive-seismic travel-time data within a DL framework, even though the two data sets are routinely acquired together at CO₂ storage sites. All training and inference are performed on graphics-processing units (GPUs); therefore, special attention is devoted to compressing the

seismic data so that memory requirements remain manageable without sacrificing subsurface resolution.

Beyond producing point estimates, we quantify the reliability of the inversion results by addressing both data-related (aleatoric) and model-related (epistemic) uncertainties (Hüllermeier and Waegeman, 2021). Aleatoric uncertainty is evaluated by injecting controlled levels of random noise into the input data, whereas epistemic uncertainty is estimated with Monte Carlo dropout applied during inference (Gal and Ghahramani, 2016; Oh and Byun, 2021; Zhu et al., 2022; Um et al., 2023a, 2023b). We also examine the combined influence of these two uncertainty sources and introduce a calibration procedure for selecting an appropriate dropout rate for geophysical imaging. Finally, we systematically compare single-data-type inversions (active-source only or passive-source only) with the proposed joint-data inversion across a range of noise levels and uncertainty settings.

3D INVERSION NETWORKS

Figure 1a illustrates a baseline 3D U-Net used for DL inversion for a single data type in this study. The encoder processes input seismic data with a shape of $(96, 96, 448, N_s)$, where the last dimension N_s corresponds to the number of shot gathers (the structure of seismic data will be described in detail in the next section). It consists of four hierarchical encoding blocks that progressively reduce spatial resolution while increasing feature depth. Each block begins with a convolutional unit comprising two 3D convolutional layers (kernel size $3 \times 3 \times 3$), which extract local spatial and temporal patterns from the input. These are followed by batch normalization, which improves training stability by normalizing the output of the convolutional layers, and ReLU activation, which introduces non-linearity and helps avoid vanishing gradients. The output of each convolutional block is passed through a 3D max pooling layer with a stride of 2 in all dimensions. Max pooling reduces the spatial resolution while retaining the most prominent features. Dropout is then applied to reduce overfitting by randomly deactivating some neurons during training. The encoder increases the number of feature channels from 16 to 128 across its four levels, ending in a bottleneck that applies two additional convolutional layers with 256 filters. As a side note, the network depth, number of channels, and kernel sizes were determined through iterative trial-and-error experiments using the training data and models described in the next section.

The decoder mirrors the encoder structure but performs upsampling to reconstruct the original input structure. Each decoder block starts with a 3D transposed convolution that doubles the spatial dimensions. This is followed by concatenation with the corresponding feature maps from the encoder

via skip connections, which help preserve fine-grained spatial information lost during downsampling. After concatenation, dropout is applied again, followed by a convolutional unit to refine the combined features. The number of filters decreases symmetrically from 128 to 16 across the decoder path. Then, a 3D convolution with a $1 \times 1 \times 1$ kernel produces a single-channel output volume representing the predicted model. Finally, a sigmoid activation function is applied to constrain the predicted model between 0 and 1, as required for predicting a CO₂ saturation model. For further information on the operations described above, including convolution, pooling, batch normalization, dropout, activation functions, and transposed convolution, the reader is referred to Goodfellow et al. (2016) and Chollet (2021). In short, this architecture effectively combines local feature extraction, multi-scale representation, and uncertainty quantification, making it particularly well-suited for seismic inversion tasks.

We extend the 3D U-Net architecture by introducing a dual-encoder framework (**Figure 1b**) designed to jointly process full-waveform data and traveltimes data. Each encoder path follows the same hierarchical structure as in the single-encoder model (**Figure 1a**), with four encoding blocks that progressively downsample the spatial dimensions while increasing feature depth from 16 to 128. The full-waveform and traveltimes data are passed through separate encoders, each consisting of convolutional blocks with two 3D convolutions (kernel size $3 \times 3 \times 3$), followed by batch normalization, ReLU activation, 3D max pooling, and dropout. The outputs from the deepest layers of both encoders are concatenated at the bottleneck and processed by a central convolutional block with 256 filters. The decoder path then symmetrically upsamples and fuses feature maps from both encoders via skip connections, using transposed convolutions, concatenation, dropout, and convolutional refinement blocks. This architecture aims to effectively extract information from both full-waveform and traveltimes data, enabling robust and accurate model predictions when both data types are available for the same label models.

Both networks have been implemented in Python using TensorFlow (Abadi et al., 2016) and Keras (Chollet, 2015). The single-encoder network and the double-encoder network have 5,658,225 and 7,999,809 trainable parameters, respectively. As a side note, the dropout layers in both networks are used for dual purposes: regularization and uncertainty quantification. A methodology for selecting an appropriate dropout rate for geophysical imaging and uncertainty quantification will be presented later.

KIMBERLINA MODELS AND DATA

To generate seismic training data, we use 3D seismic velocity models (Alumbaugh et al., 2024^b) derived from the Kimberlina 1.2 CO₂ injection reservoir model (Birkholzer et al., 2011; Wainwright et al., 2013). This model represents a hypothetical CO₂ storage site located approximately 30 km northwest of Bakersfield, California, in the Southern San Joaquin Basin. It includes 12 geologic formations, with CO₂ injected into the deep Vedder sandstone formation at a depth of about 2,750 meters. The Vedder formation is a permeable sandstone reservoir unit overlain by a low-permeability shale seal, making it a suitable target for long-term CO₂ storage. The version of the Kimberlina model used in this study includes 1,764 CO₂ saturation models (63 scenarios or sub-models $\times$ 28 time steps) and 1,764 corresponding V_p models.

Each model spans a spatial domain of $4 \times 4 \times 3.5$ kilometers in the x, y, and z directions, respectively, with a uniform grid spacing of 10 meters. The models were generated at 28 non-uniformly sampled time steps spanning from year 0 to year 150, specifically: 0, 1, 2, 5, 10, 15, 20, 25, 30, 35, 40, 45, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, and 150. During the first 50 years, CO₂ is injected at a constant rate of 5 million tons per year. Injection then ceases, and the model simulates CO₂ migration and redistribution during a 100-year post-injection phase. **Figure 2** illustrates the evolution of CO₂ plumes and the associated changes in P-wave velocity over time for one of the Kimberlina scenarios.

Using the 1,764 velocity models, we generated synthetic microseismic data. For each model, we assume 1,000 microseismic events induced by CO₂ injection. The event locations, shown in **Figure 3**, are distributed near the CO₂ plumes. Note that the induced microseismic events can also happen in regions beyond the immediate vicinity of the CO₂ plumes (Rutqvist et al., 2016). However, in this modeling study, we assume that all microseismic events are clustered within and around the CO₂ plumes to support an ideal imaging experiment. Due to the high computational cost associated with simulating wavefields for all 1,764 models and their corresponding 1,000 events per model, we chose to compute traveltimes using an eikonal solver (Alkhalifah and Fomel, 2001) rather than performing full-waveform simulations. In our traveltimes simulations, surface receivers are densely placed at 10-meter intervals in both the x- and y-directions, resulting in a traveltimes data cube of size (400, 400, 1000) for each model.

In a realistic scenario, the number of microseismic events occurring over a given period is not fixed but random. To accommodate this variability, we construct a new data cube of size (400, 400, N_t), where the first two indices represent surface receiver locations, and the third index corresponds to the time axis with a time interval of Δt . For example, if $N_t = 1000$ and $\Delta t = 3$ ms, the time axis spans from 0 to 3 seconds. The data cube is initially populated with zeros. Each layer of the original traveltimes

data cube, such as (400, 400, $i=1$) shown in **Figure 4a**, is then mapped onto the new data cube as a 3D parabolic envelope, where unity is used to indicate the presence of an event (i.e., using a one-hot encoding scheme). As this marking process is repeated for each layer of the traveltime data, additional parabolic envelopes are added to the new cube, as illustrated in **Figure 4b**. Regions where multiple envelopes overlap result in higher integer values, indicating intersections of arrival patterns.

The advantage of this data structure is clear: it allows for a fixed data size regardless of the number of microseismic events occurring during a given period. This consistency in data size is necessary when using the U-Net and its modified version shown in **Figure 1**. To introduce more realism into our DL inversion experiments, we randomly select between 600 and 800 microseismic events from the 1,000 simulated traveltime events for each model and construct its corresponding traveltime data cube. In other words, each microseismic training data contain a randomly varying number of microseismic events within that range.

To construct the proposed traveltime data cube, we initially set $\Delta t = 3$ ms and $N_t = 1000$. We then remove the first 280 and the last 272 time layers, as they contain only zeros. Because the traveltime data vary smoothly in space (see **Figure 4a**), we also downsample the surface receivers by selecting every 40 meters in both the x - and y -directions and discard the last five receivers in each direction. This yields a final data cube (**Figure 4c**) with dimensions of (96, 96, 448), which is about 2.5% of the original size (400, 400, 1000). As demonstrated in the DL inversion results presented later, this compact data representation retains sufficient information for accurate delineation of the CO₂ plume.

We downloaded seismic full waveform data for the velocity models described above from the US Department of Energy EDX (<https://edx.netl.doe.gov>). These data were generated by simulating vwavefields in the 3D models using a finite-difference acoustic wave equation solver (Moczo et al., 2007). Each model includes 25 shots. Each shot gather is represented as a 3D array of size (40, 40, 5001), where the first two indices correspond to geophone locations spaced at 100-meter intervals, and the third index represents the time axis with a temporal resolution of 1 millisecond. As a result, the time axis spans from 0 to 5 seconds. For more detailed information, readers are referred to Alumbaugh et al. (2024^b).

Based on the velocity structure of the models, the early-time data from 0 to 1 second and the late-time data after 4.5 seconds contain little information related to the CO₂ plumes and were therefore removed. The remaining traces were downsampled by selecting every eighth point, resulting in a data size of (40, 40, 448). Then, the waveform data were linearly interpolated to a size of (96, 96, 448) to match the spatial dimensions of the traveltime data cube. The reason for aligning the data dimensions will be explained in the next section. For training, we use the difference between two full waveform

datasets, one at year X and the other at year 0 (i.e. baseline data before the injection started), as illustrated in **Figure 5a**.

Since each model contains 25 seismic sources, the full waveform data for each CO₂ model has a size of (96, 96, 448, 25). In this study, we also generate an alternative data structure in which all shot gathers are summed along the last axis, resulting in a more compact representation with a size of (96, 96, 448, 1), as shown in **Figure 5b**. Although this summation may lead to some loss of information, it significantly reduces the data size by approximately 97% compared to the original (40, 40, 5001, 25) format. As we will show and discuss later, this compact representation can not only effectively delineate the CO₂ plume but also has potential for fast and efficient generation of seismic full waveform training data. Note that reducing the input channel from 25 to 1 does not significantly lower the training computational load; however, the smaller input data size can reduce overall training time by alleviating GPU memory bandwidth pressure. We will compare the training time costs of the (96, 96, 448, 1) and (96, 96, 448, 25) formats in the inversion experiment section.

As a final note, we train the inversion networks using pairs of (3D seismic data, 3D CO₂ saturation models) rather than (3D seismic data, 3D velocity models), because our primary interest lies in directly imaging CO₂ saturation (Um et al., 2023^a; Um et al., 2023^b) rather than geophysical proxy models such as velocity or velocity changes. The original velocity models, initially sized at (400, 400, 350), are first resampled to (96, 96, 350) and then extended to (96, 96, 448) by applying zero-padding to the lower portion of the depth axis to match the output dimensions required by the network.

TRAINING NETWORKS

Using the training data and model sets described in the previous section, we consider five different DL inversion experiments. First, we train the single-encoder U-Net (**Figure 1a**) using pairs of seismic full waveform data with size (96, 96, 448, 25) and the CO₂ label model. Second, we train the single-encoder U-Net using pairs of seismic full waveform data with size (96, 96, 448, 1) and the CO₂ label model. Third, we train the single-encoder U-Net using pairs of passive seismic traveltime data with size (96, 96, 448, 1) and the CO₂ label model. Fourth, we train the dual-encoder U-Net using pairs of full waveform data with size (96, 96, 448, 1) and passive seismic traveltime data with size (96, 96, 448, 1), along with the CO₂ label model.

In the previous section, we made both active seismic full waveform data and passive seismic traveltime data have the same size of (96, 96, 448, 1). In both datasets, the first, second, and third

indices represent the x, y, and time axes, respectively. Therefore, the two datasets are somewhat homogenized, although their third index has different time resolutions (Δt). Therefore, it is natural to examine whether concatenating the two datasets and training the single-encoder U-Net for the joint inversion would be feasible. Thus, in the fifth experiment, we concatenate the full waveform data and traveltime data, each with size (96, 96, 448, 1), into a combined training input of size (96, 96, 448, 2) and train the single-encoder U-Net using pairs of this concatenated data and the CO₂ label model. The first three DL inversions use a single data type, while the latter two involve joint data types. For easier reference, the inversion experiments are referred to as FWI25, FWI1, Traveltime1, JNT1, and JNT2. The five DL inversion experiments are summarized in **Table 1**.

Before training begins, the 63 data samples and models from year 20 are reserved as the test set and excluded from the training process. The remaining data and models are split into training and validation sets in a 9:1 ratio. The Adam optimizer (Kingma, 2014) is used with a learning rate of 5×10^{-4} , and training is configured to stop if the validation loss, measured by mean squared error (MSE) between the true and predicted CO₂ models, does not improve after 20 epochs. The learning rate is reduced by half if the validation loss plateaus for 8 epochs, helping to ensure convergence. All DL training experiments were performed using an NVIDIA A40 GPU. Training FWI1, Traveltime1, and JNT2 each took less than 24 hours to complete, while JNT1 required approximately 30 hours and FWI25 approximately 48 hours. The longer training time for FWI25 compared to FWI1 is primarily due to the larger file sizes and memory bandwidth demands associated with FWI25, rather than a significantly higher computational load.

Figure 6 presents the training and validation loss plots for the DL inversion experiments. First, note that FWI25 and FWI1 exhibit comparable final validation losses, indicating that the compact data representation formed by summing 25 differential shot gathers is effective. Second, JNT1 and JNT2 also show comparable final validation losses, suggesting that a single-encoder U-Net may handle 3D joint data types in certain circumstances such as when the two data types share underlying similarities. In contrast, Traveltime1 displays a significantly higher final validation loss compared to the other experiments. This is primarily due to the randomness in the number of microseismic events in each training dataset, as described earlier. When the network is trained using datasets containing the full set of 1000 seismic events, we confirmed that Traveltime1 also achieved a final validation loss comparable to the other four experiments. Nevertheless, the final validation loss of Traveltime1, as shown in **Figure 6c**, remains sufficiently low to ensure good imaging performance, as will be demonstrated later.

DROPOUT RATE

During the training processes shown in **Figure 6**, the dropout rate was set to 0.35. In this section, we describe how the dropout rate was selected for dual purposes: regularization and uncertainty quantification. The dropout method was originally developed as a regularization tool (Srivastava et al., 2014). It randomly drops out a subset of neurons during each training iteration, which helps reduce overfitting by ensuring the network learns more robust, generalized features rather than memorizing the training data. For regularization purposes, we aim to find a dropout rate that minimizes the validation loss, which serves as our misfit measure for convergence.

We identify a range in which dropout provides effective regularization without causing underfitting or non-convergence by gradually increasing the dropout rate and observing the smallest validation loss. **Figure 7a** shows the smallest validation loss achieved for FWI1 as the dropout rate increases from 0.05 to 0.5 in intervals of 0.05. As the dropout rate increases, the loss initially decreases due to improved generalization. However, beyond the optimal point (i.e., 0.25), the minimum validation loss begins to rise again, as excessive dropout perturbs the network too strongly.

Once the dropout method is used as a regularization technique during training, the dropout layers are typically disabled during prediction, and all neurons are used to produce the most accurate results. However, Gal and Ghahramani (2016) demonstrated that enabling dropout layers during prediction and performing multiple forward passes can approximate Bayesian inference and be used to quantify model uncertainty. This approach is known as the Monte Carlo (MC) dropout method. MC dropout offers several computational advantages that make it a practical and widely used method for uncertainty estimation. First, it introduces no additional parameters because dropout layers are already included in most neural networks, keeping memory usage and model size unchanged. Second, it enables uncertainty estimation using a single model by performing multiple stochastic forward passes, eliminating the need to store or manage multiple models as required in ensemble approaches as is required in ensemble methods like bootstrap aggregation (James et al., 2013).

In this study, we use the MC dropout method for measuring the network uncertainty. First, we perform a sufficient number of predictions with dropout activated and then calculate their mean and standard deviation, which serve as the final prediction and its uncertainty measure. In the context of MC dropout, selecting the dropout rate that yields the smallest validation error is not directly relevant to the objective of uncertainty quantification. If the dropout rate is too low, the resulting approximate posterior behaves like a deterministic network. In contrast, if the dropout rate is too high, it not only affects the sampling process but also alters the posterior itself. Therefore, it is more appropriate to choose the dropout rate at which the spatial distribution of the standard deviation is most correlated

with that of the prediction error. To determine this, we train the network using dropout rates ranging from 0.05 to 0.5 in intervals of 0.05. For each trained model, we perform 1,000 forward passes with dropout enabled and compute the mean and standard deviation of the predictions. The prediction error is defined as the absolute difference between the true CO₂ label model and the mean CO₂ prediction. The standard deviation and prediction error are both represented as 3D matrices of shape (96, 96, 448). We flatten these matrices and calculate the correlation coefficient between them, as shown in **Figure 7**.

As the dropout rate increases, the correlation coefficient peaks around a dropout rate 0.35. At several selected dropout rates, we present a direct comparison (**Figure 8**) between the true CO₂ label model, the predicted mean CO₂ model, the absolute difference between the imaged and true model, and the standard deviation of the predicted CO₂ model. Compared to the cases with a dropout rate that is too low (0.05; **Figure 8a**) or too high (0.5; **Figure 8c**), the case with a dropout rate of 0.35 (**Figure 8b**) shows the highest similarity between the spatial distribution of prediction error and that of the standard deviation. Based on this observation, we conclude that a dropout rate in the range of 0.3 to 0.4 is suitable for both regularization and uncertainty quantification. We repeat this correlation analysis for the other four networks and observe a similar result. Therefore, a dropout rate of 0.35 is selected as the final setting for this study.

DL INVERSION EXPERIMENTS

First, we predict a CO₂ saturation model using five different trained networks with the dropout layers inactivated (inversion experiment group 1). The test data used in this experiment are noise-free. **Figure 9** shows a 3D comparison between the true CO₂ model and the five predicted CO₂ saturation models. Overall, all five predicted models appear very close to the true model. For a more detailed comparison, **Figure 10** presents cross-sectional views of the 3D models along with their relative differences, defined as the difference between the true and predicted models divided by the true model. All five inversion results successfully recover the three CO₂ plume layers, the lateral extent of the plume, and saturation distribution. High relative differences are mainly observed near the boundaries of the CO₂ plume.

FWI1 (**Figure 10b**) performs as well as FWI25 (**Figure 10a**), indicating that the compact data representation formed by summing 25 differential shot gathers does not degrade the inversion result. This suggests that, during training dataset generation, it is feasible to simulate 25 different seismic sources simultaneously rather than simulating each source individually at 25 different positions. By

combining all sources into a single forward modeling step (Um et al., 2012), this approach can significantly reduce the computational cost of generating full waveform training data. Traveltime1 (**Figure 10c**) also produces results comparable to FWI1 and FWI25, further supporting the effectiveness of the traveltime data structure proposed in this work. JNT2 (**Figure 10e**) yields results similar to JNT1 (**Figure 10d**), although slightly higher relative differences are observed in regions with high CO₂ saturation.

Table 2 presents quantitative metrics comparing the true model with the predictions from each network. Both the mean squared error (MSE) and root mean squared error (RMSE) are low, and the structural similarity index (SSI) (Wang et al., 2004) is close to 1, indicating that all five deep learning inversion results closely match the true model. Another way to evaluate the inversion performance is to compare the total amount of CO₂ predicted with that of the true model. Assuming that all CO₂ is stored in the Vedder sand formation, with a known porosity of 0.264 (Alumbaugh et al., 2024^b), we calculate the relative difference in total CO₂ volume between the true and predicted model. As shown in the sixth column of **Table 2**, the DL inversion results predict the total CO₂ amount within a few percent error. Among the five models, JNT1 shows the smallest relative difference.

Second, for each network, we perform 1,000 forward passes with dropout activated using noise-free test data and compute the mean and standard deviation of the predictions (inversion experiment group 2). As mentioned earlier, once fully trained, the networks can generate predictions in near real time; for example, 1,000 predictions can be completed in just a few minutes. **Figure 11** compares the true model with the mean predicted CO₂ models, the relative differences from the true model, and the standard deviations. Compared to **Figure 10**, the relative differences are slightly elevated in regions with high CO₂ concentrations. The standard deviation distribution closely aligns with the relative difference, indicating that the MC dropout rate is appropriately selected for analyzing model uncertainty. Overall, the standard deviation remains very small, implying that the trained network can predict CO₂ saturation with high confidence when the test data fall well within the distribution of the training data. The highest value is observed in the region of high CO₂ saturation in the Traveltime1 (**Figure 11c**), but it is still less than 0.01. **Table 2** shows that the quantitative metrics for this group are comparable to those from inversion experiment group 1.

Third, for each network, we perform 1,000 forward passes with dropout inactivated using noisy test data (inversion experiment group 3). In each forward pass, the full waveform data are contaminated with Gaussian noise with a mean of 0 and a standard deviation equal to 20% of the amplitude of each data point. The traveltime data are contaminated with Gaussian noise with a mean of 0 and a standard deviation of 0.003 seconds, corresponding to the time resolution used in constructing the traveltime

data structure. We then compute the mean and standard deviation of the predictions. This experiment is designed to evaluate the effect of data uncertainty on the DL inversion. **Figure 12** compares the true model with the mean predicted CO₂ models, the relative differences from the true model, and the standard deviations. Overall, all five DL inversion networks successfully recover the CO₂ plume. The performance metrics shown in **Table 2** do not indicate significant degradation. However, visual inspection and comparison of **Figure 12** with **Figures 10** and **11** reveal that both the relative differences and standard deviations are further elevated in this group, reflecting the added uncertainty from noisy input data.

Finally, for each network, we perform 1,000 forward passes with dropout activated using the noisy test data generated in the same manner as in the previous experiment (inversion experiment group 4). **Figure 13** compares the true model with the mean predicted CO₂ models, the relative differences, and the standard deviations. Both relative differences and standard deviations are slightly but further elevated, as this inversion experiment introduces both model and data uncertainty into the inversion. The performance metrics shown in **Table 2** indicate that this inversion experiment group yields the lowest overall performance compared to the other three groups. Nonetheless, the degradation remains fairly small, and the inversion results remain robust and well within acceptable accuracy.

For all five inversion experiments, JNT1 and JNT2 produce comparable results. This suggests that a smaller single-encoder network can be used for joint data by concatenating them into a single, larger dataset if the joint data are contextually similar and share the same format. This approach can improve overall computational efficiency. However, it may not be effective when data types differ significantly in nature or size, such as large seismic full waveform data in the time domain and sparse electromagnetic data in the frequency domain. Additionally, geophysical datasets may not be centralized and concatenated when they are confidential and cannot be shared. We encountered such cases when collecting legacy data from multiple data owners to construct training datasets for a potential CO₂ storage site. In these situations, each data owner must train a shared inversion network locally, sharing a locally trained network or its weights rather than the entire raw data. This approach, known as federated learning (Li et al., 2020), avoids data centralization. For such scenarios, we believe a dual/multi-encoder approach would be more suitable, and we are currently exploring federated learning for multiphysics data using a multi-encoder network model.

Finally, we discuss how the imaging accuracy of DL-FWI compares with that of physics-based FWI, drawing on published examples rather than a strict point-by-point comparison. A direct point-by-point comparison is impossible because DL-FWI estimates a CO₂ saturation model, whereas traditional physics-based FWI yields seismic-velocity models. Instead, we use the structural similarity index

(SSI) as a common benchmark. Our DL inversions reach SSI values as high as 0.993 (Table 2), whereas physics-based FWI applied to comparable Kimberlina data sets typically achieves about 0.97 (Rojas-Gómez et al., 2020; Lin et al., 2023). In these studies, visual inspection also shows that DL-based images resolve finer plume details. This enhanced resolution is attributed to the models learned during training, which encode high-wavenumber geological patterns often missed by physics-only inversion. For a more detailed comparison between the two approaches, readers are referred to Lin et al. (2023).

Computational cost provides additional context. Generating the 44,100 synthetic full waveform shot gathers used for training (1,764 models $\times$ 25 shots) in this study required roughly 23,000 hours using a typical HPC node (e.g. a single node with two Haswell processors), and it subsequently took about 30 GPU-hours on an NVIDIA A40 to train each network. By contrast, a traditional physics-based FWI run on the same survey would need fewer than 10,000 to 20,000 forward simulations (25 shots $\times$ 3~4 simulations per inversion iteration $\times$ 100-200 inversion iterations) when solved with a nonlinear conjugate-gradient algorithm (Nocedal and Wright, 2006). Thus, the one-time expense of DL-FWI is higher than that of traditional physics-based FWI. Once trained, however, the DL inversion can predict a saturation model and its uncertainty in seconds and can be incrementally updated as new data are acquired over the lifetime of the geological storage. This trade-off, namely higher initial cost for near-real-time, repeatable predictions, positions DL-FWI as a valuable complement to traditional FWI methods for long-term monitoring of CO₂ storage sites.

CONCLUSIONS

We have proposed 3D DL inversion networks and demonstrated their performance for CO₂ saturation imaging and uncertainty quantification using both active seismic full waveform data and passive seismic traveltimes data. For example, the inversion experiments demonstrate the structural similarity index up to 0.993 and CO₂ difference up to 1%. The evaluation of single-encoder and dual-encoder U-Net architectures, as well as the single-encoder U-Net with concatenated joint data, all proved capable of producing accurate CO₂ saturation models under both ideal and noisy conditions. Furthermore, the systematic approach for selecting an optimal dropout rate effectively balanced regularization and MC dropout-based uncertainty quantification, as supported by the strong spatial correlation between prediction errors and uncertainty estimates.

The proposed compact data representations can enable efficient training data generation, compact data storage, and fast DL training and prediction, compared to using the originally large 3D seismic

1 datasets. For example, the compact representation formed by summing 25 differential shot gathers
2 suggests that 25 seismic sources can be simulated simultaneously in a single forward modeling step,
3 rather than simulating each one individually during the training data generation phase. This can
4 significantly reduce both storage and computational costs for 3D seismic full waveform data.
5 Similarly, the proposed compact representation for passive seismic traveltime data allows the data to
6 be stored in a fixed data structure, regardless of the number of microseismic events that occur during a
7 given period.

8 The inversion results across the five different network configurations and uncertainty scenarios
9 consistently demonstrated accurate plume imaging. Standard deviation analysis provided valuable
10 insights into the reliability of predictions, supported by relative difference analysis. The single-
11 encoder U-Net using summed full waveform shot gather data achieved inversion results comparable
12 to those obtained with much larger unsummed full waveform data. The single-encoder network
13 trained on concatenated joint data also performed comparably to the dual-encoder network. This
14 research highlights the potential of 3D DL geophysical inversion to overcome the computational
15 challenges associated with traditional physics-based 3D seismic inversion methods, while delivering
16 crucial uncertainty information for improved reservoir management and risk assessment.

17 The work presented here relies solely on numerical modeling data from the Kimberlina 1.2
18 benchmark to evaluate network performance under controlled conditions and perform rigorous
19 uncertainty analysis. Although the tests confirm the value of the compact data representation, joint-
20 data training, and dropout-based uncertainty calibration, there are several important factors that our
21 numerical modeling analysis cannot address. They include intrinsic uncertainties in rock physics
22 models, noise in field data that numerical modeling analysis cannot fully address, acquisition
23 imperfections and others. Our current and future efforts therefore focus on (1) applying the networks
24 to time-lapse seismic and microseismic data from active CO₂ sites, (2) broadening synthetic scenarios
25 to include variable event counts and geometries, and (3) implementing federated learning so
26 proprietary data can be used without centralization. These steps will more rigorously test and refine
27 the methodology for real-world deployment.

28 29 **ACKNOWLEDGMENTS**

30 This work was supported by (1) Project the Science-informed Machine learning to Accelerate Real
31 Time decision making for Carbon Storage (SMART-CS) Initiative from the U.S. Department of
32 Energy's Office of Fossil Energy and Carbon Management's Carbon Storage Research program

1 through the National Energy Technology Laboratory and (2) Project Physics-Informed and Energy-
2 Aware Federated Learning of Neural Multi-Operator Learners as Scientific Foundation Models
3 (FedNeMO) from Advanced Scientific Computing Research of DOE's Office of Science. This
4 research used the Lawrencium computational cluster resource provided by the IT Division at the
5 Lawrence Berkeley National Laboratory (Supported by the Director, Office of Science, Office of
6 Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-
7 05CH11231). The authors thank the editor, Professor Felix Herrmann of the Georgia Institute of
8 Technology, and anonymous reviewers for their constructive feedback.

9

REFERENCES

- Abadi, M., P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, et al., 2016, TensorFlow: A system for large-scale machine learning: *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, **12**, 265–283.
- Alkhalifah, T., and S. Fomel, 2001, Implementing the fast marching eikonal solver: Spherical versus Cartesian coordinates: *Geophysical Prospecting*, **49**, 165–178.
- Alumbaugh, D. L., J. Correa, P. Jordan, B. Petras, S. Chundur, and W. Abriel, 2024^a, An assessment of the role of geophysics in future US geologic carbon storage projects: *The Leading Edge*, **43**, 72–83.
- Alumbaugh, D., E. Gasperikova, D. Crandall, M. Commer, S. Feng, W. Harbert, Y. Li, Y. Lin, and S. Samarasinghe, 2024^b, The Kimberlina synthetic multiphysics dataset for CO₂ monitoring investigations: *Geoscience Data Journal*, **11**, 216–234.
- Araya-Polo, M., J. Jennings, A. Adler, and T. Dahlke, 2018, Deep-learning tomography: *The Leading Edge*, **37**, 58–66.
- Brun, P., 2020, Energy transition fueling offshore wind market growth: *Offshore Magazine*, **80**, no. 1, 26–27.
- Celaya, A. and Araya-Polo, M., 2024, July. Deep learning-based joint inversion of time-lapse surface gravity and seismic data for monitoring of 3D CO₂ plumes. In Fourth International Meeting for Applied Geoscience & Energy (pp. 1119-1123). Society of Exploration Geophysicists and American Association of Petroleum Geologists.
- Çiçek, Ö., A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, 2016, 3D U-Net: Learning dense volumetric segmentation from sparse annotation: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, **19**, 424–432.
- Chollet, F., 2015, Keras: <https://keras.io>, accessed 28 October 2021.

- 1 Chollet, F., 2021, *Deep learning with Python*: Simon and Schuster.
- 2 Dai, Q., Y. H. Lee, H. H. Sun, G. Ow, M. L. M. Yusof, and A. C. Yucel, 2023, 3DInvNet: A deep
3 learning-based 3D ground-penetrating radar data inversion: *IEEE Transactions on Geoscience and*
4 *Remote Sensing*, **61**, 1–16.
- 5 Feng, S., Lin, Y. and Wohlberg, B., 2021. Multiscale data-driven seismic full-waveform inversion
6 with field data study. *IEEE transactions on geoscience and remote sensing*, 60, pp.1-14.
- 7 Fichtner, A., 2010, *Full seismic waveform modelling and inversion*: Springer.
- 8 Gal, Y., and Z. Ghahramani, 2016, Dropout as a Bayesian approximation: Representing model
9 uncertainty in deep learning: *Proceedings of the 33rd International Conference on Machine Learning*
10 *(ICML)*, PMLR, 1050–1059.
- 11 Goodfellow, I., Y. Bengio, and A. Courville, 2016, *Deep learning*: MIT Press.
- 12 Guo, R., Zhou, H., Wei, X., Lin, Z., Li, M., Eldar, Y.C., Yang, F., Xu, S. and Abubakar, A., 2025.
13 Deep joint inversion of multiple geophysical data with U-net reparameterization. *Geophysics*, 90(3),
14 pp.WA61-WA75.
- 15 Hicks, E., H. Hoeber, M. Houbiers, S. P. Lescoffit, A. Ratcliffe, and V. Vinje, 2016, Time-lapse
16 full-waveform inversion as a reservoir-monitoring tool—A North Sea case study: *The Leading Edge*,
17 **35**, 850–858.
- 18 Hu, Y., Wei, X., Wu, X., Sun, J., Chen, J., Huang, Y. and Chen, J., 2023. A deep learning-enhanced
19 framework for multiphysics joint inversion. *Geophysics*, 88(1), pp.K13-K26.
- 20 Hüllermeier, E., and W. Waegeman, 2021, Aleatoric and epistemic uncertainty in machine learning:
21 An introduction to concepts and methods: *Machine Learning*, **110**, 457–506.
- 22 James, G., D. Witten, T. Hastie, and R. Tibshirani, 2013, *An introduction to statistical learning*:
23 Springer.

Jin, P., Y. Feng, S. Feng, H. Wang, Y. Chen, B. Consolvo, Z. Liu, and Y. Lin, 2024, An empirical study of large-scale data-driven full waveform inversion: *Scientific Reports*, **14**, 20034.

Kallivokas, L. F., A. Fathi, S. Kucukcoban, K. H. Stokoe II, J. Bielak, and O. Ghattas, 2013, Site characterization using full waveform inversion: *Soil Dynamics and Earthquake Engineering*, **47**, 62–82.

Kingma, D.P., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Li, T., Sahu, A.K., Talwalkar, A. and Smith, V., 2020. Federated learning: Challenges, methods, and future directions. *IEEE signal processing magazine*, 37(3), pp.50-60.

Lin, Y., Theiler, J. and Wohlberg, B., 2023. Physics-guided data-driven seismic inversion: Recent progress and future opportunities in full-waveform inversion. *IEEE Signal Processing Magazine*, 40(1), pp.115-133.

Moczo, P., J. O. Robertsson, and L. Eisner, 2007, The finite-difference time-domain method for modeling of seismic wave propagation: *Advances in Geophysics*, **48**, 421–516.

Nocedal, J. and Wright, S.J., 2006. Numerical optimization. New York, NY: Springer New York.

Oh, S., and J. Byun, 2021, Bayesian uncertainty estimation for deep learning inversion of electromagnetic data: *IEEE Geoscience and Remote Sensing Letters*, **19**, 1–5.

Petrov, P. V., and G. A. Newman, 2014, Three-dimensional inverse modelling of damped elastic wave propagation in the Fourier domain: *Geophysical Journal International*, **198**, 1599–1617.

Queiße, M., and S. C. Singh, 2013, Full waveform inversion in the time lapse mode applied to CO₂ storage at Sleipner: *Geophysical Prospecting*, **61**, 537–555.

Rojas-Gómez, R., Yang, J., Lin, Y., Theiler, J. and Wohlberg, B., 2020. Physics-consistent data-driven waveform inversion with adaptive data augmentation. *IEEE Geoscience and Remote Sensing Letters*, 19, pp.1-5.

Ronneberger, O., P. Fischer, and T. Brox, 2015, U-net: Convolutional networks for biomedical image segmentation: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, **18**, 234–241.

Rutqvist, J., Rinaldi, A.P., Cappa, F., Jeanne, P., Mazzoldi, A., Urpi, L., Guglielmi, Y. and Vilarrasa, V., 2016. Fault activation and induced seismicity in geological carbon storage—Lessons learned from recent modeling studies. *Journal of Rock Mechanics and Geotechnical Engineering*, **8**, 789–804.

Shin, C., and Y. Ho Cha, 2009, Waveform inversion in the Laplace–Fourier domain: *Geophysical Journal International*, **177**, 1067–1079.

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

Um, E.S., Harris, J.M. and Alumbaugh, D.L., 2012. An iterative finite element time-domain method for simulating three-dimensional electromagnetic diffusion in earth, *Geophysical Journal International*, **190**, 871–886.

Um, E. S., D. Alumbaugh, Y. Lin, and S. Feng, 2023^a, Real-time deep-learning inversion of seismic full waveform data for CO₂ saturation and uncertainty in geological carbon storage monitoring: *Geophysical Prospecting*, **72**, 199–212.

Um, E. S., D. Alumbaugh, M. Commer, S. Feng, E. Gasperikova, Y. Li, Y. Lin, and S. Samarasinghe, 2023^b, Deep learning multiphysics network for imaging CO₂ saturation and estimating uncertainty in geological carbon storage: *Geophysical Prospecting*, **72**, 183–198.

Um, E. S., M. Commer, and G. A. Newman, 2014, A strategy for coupled 3D imaging of large-scale seismic and electromagnetic data sets: Application to subsalt imaging: *Geophysics*, **79**, ID1–ID13.

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

Virieux, J., A. Asnaashari, R. Brossier, L. Métivier, A. Ribodetti, and W. Zhou, 2017, An introduction to full waveform inversion: *Encyclopedia of Exploration Geophysics*, SEG, R1–1.

Wainwright, H. M., S. Finsterle, Q. Zhou, and J. T. Birkholzer, 2013, Modeling the performance of large-scale CO₂ storage systems: A comparison of different sensitivity analysis methods: *International Journal of Greenhouse Gas Control*, **17**, 189–205.

Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, *13*(4), 600-612.

Wu, Y., and Y. Lin, 2019, InversionNet: An efficient and accurate data-driven full waveform inversion: *IEEE Transactions on Computational Imaging*, **6**, 419–433.

Yang, F., and J. Ma, 2019, Deep-learning inversion: A next-generation seismic velocity model building method: *Geophysics*, **84**, R583–R599.

Zeng, Q., Feng, S., Wohlberg, B. and Lin, Y., 2021. InversionNet3D: Efficient and scalable learning for 3-D full-waveform inversion. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, pp.1-16.

Zhou, Q., and J. T. Birkholzer, 2011, On scale and magnitude of pressure build-up induced by large-scale geologic storage of CO₂: *Greenhouse Gases: Science and Technology*, **1**, 11–20.

Zhu, W., K. Xu, E. Darve, B. Biondi, and G. C. Beroza, 2022, Integrating deep neural networks with full-waveform inversion: Reparameterization, regularization, and uncertainty quantification: *Geophysics*, **87**, R93–R109.

TABLES

Experiment	Type	Data Pairs
FWI25	FWI only	Active seismic full waveform data (96, 96, 448, 25) + CO ₂ label model
FWI1	FWI only	Active seismic full waveform data (96, 96, 448, 1) + CO ₂ label model
Traveltime 1	Travel time Inversion only	Passive seismic traveltime data (96, 96, 448, 1) + CO ₂ label model
JNT1	FWI + Traveltime	Active seismic full waveform data (96, 96, 448, 1) + Passive seismic traveltime data (96, 96, 448, 1) + CO ₂ label model
JNT2	FWI + Traveltime	Concatenated active seismic full waveform data and passive seismic traveltime data (96, 96, 448, 2) + CO ₂ label model

Table 1: Summary of DL Inversion Experiments

1

Group ID	Network ID	MSE	RMSE	SSI	CO ₂ Difference (%)
1	FWI25	1.31E-04	1.15E-02	9.89E-01	5.80
1	FWI1	7.85E-05	8.86E-03	9.93E-01	5.40
1	Traveltimel	2.14E-04	1.46E-02	9.86E-01	4.13
1	JNT1	8.03E-05	8.96E-03	9.93E-01	1.10
1	JNT2	1.56E-04	1.25E-02	9.89E-01	-2.14
2	FWI25	1.32E-04	1.15E-02	9.89E-01	4.77
2	FWI1	7.78E-05	8.82E-03	9.93E-01	3.78
2	Traveltimel	1.83E-04	1.35E-02	9.87E-01	4.55
2	JNT1	7.49E-05	8.66E-03	9.93E-01	1.64
2	JNT2	1.35E-04	1.16E-02	9.89E-01	-5.53
3	FWI25	1.31E-04	1.15E-02	9.89E-01	3.50
3	FWI1	1.12E-04	1.06E-02	9.92E-01	-4.07
3	Traveltimel	1.82E-04	1.35E-02	9.87E-01	3.53
3	JNT1	1.16E-04	1.08E-02	9.91E-01	-8.77
3	JNT2	2.04E-04	1.43E-02	9.87E-01	-5.80
4	FWI25	2.54E-04	1.59E-02	9.89E-01	-5.77
4	FWI1	1.14E-04	1.07E-02	9.91E-01	-4.89
4	Traveltimel	2.39E-04	1.55E-02	9.85E-01	-6.53
4	JNT1	1.27E-04	1.13E-02	9.90E-01	-8.93
4	JNT2	2.51E-04	1.58E-02	9.85E-01	-7.53

2

3 **Table 2.** Performance comparison between the five different inversion networks under the four
4 different inversion experiment groups. MSE, RMSE and SSI represent mean squared error, root mean
5 squared error and structural similarity index, respectively.

6

1 **FIGURE CAPTIONS**

2 **Figure 1.** (a) The 3D U-Net architecture used for single-physics inversion and experimental joint
3 inversion. (b) The modified 3D U-Net architecture for joint inversion. Blue lines indicate skip
4 connections.

5 **Figure 2.** Evolution of CO₂ plumes and the corresponding changes in V_p over time. (Top) CO₂
6 saturation models. (Bottom) Changes in the V_p models due to CO₂ injection.

7 **Figure 3.** The locations of microseismic events in the Kimberlina model space. The color chart
8 represents the depth of the events. The black dots on the surface of the earth represent geophone
9 locations.

10 **Figure 4.** The 3D Kimberlina traveltimes data. (a) The traveltimes data from a single microseismic
11 event whose location is denoted as a red star. (b) The 3D data cube containing 1000 microseismic
12 traveltimes envelopes. (c) The final traveltimes data cube whose time axis spans from 0.72 seconds to
13 2.18 seconds.

14 **Figure 5.** Construction of seismic full waveform input data. (a) Differential waveform data obtained
15 by subtracting a shot gather at year 0 from the corresponding gather at year 20. (b) A compact
16 representation formed by summing the differential data from all 25 shot gathers.

17 **Figure 6.** The training and validation loss plots for the five DL inversion experiments.

18 **Figure 7.** The smallest validation loss of FWI1 and the correlation coefficient between standard
19 deviation and prediction error as a function of dropout rate.

20 **Figure 8.** Comparison of true CO₂ model, predicted mean CO₂ model from FWI1, prediction error,
21 and standard deviation at different dropout rates.

22 **Figure 9.** 3D views of the true CO₂ model and five different DL inversion results. The dropout layers
23 are inactivated and noise-free test data are used.

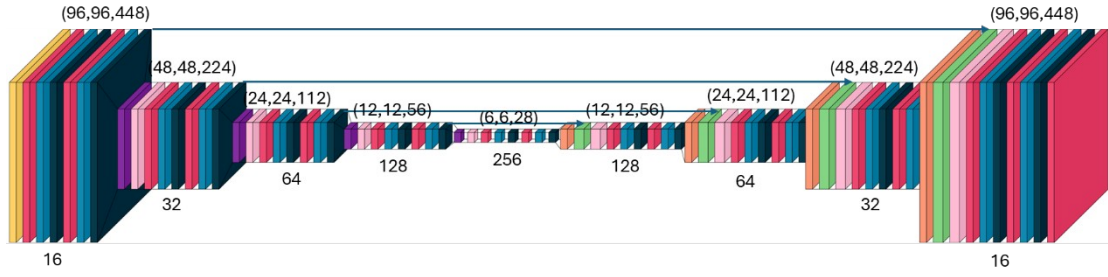
Figure 10. Cross-sectional views of true CO₂ model, predicted CO₂ model and their relative difference (%). The dropout layers are disabled and noise-free test data are used (inversion experiment group 1).

Figure 11. Cross-sectional views of true CO₂ model, predicted CO₂ model, their relative difference (%) and standard deviation. The dropout layers are activated during prediction and noise-free test data are used (inversion experiment group 2).

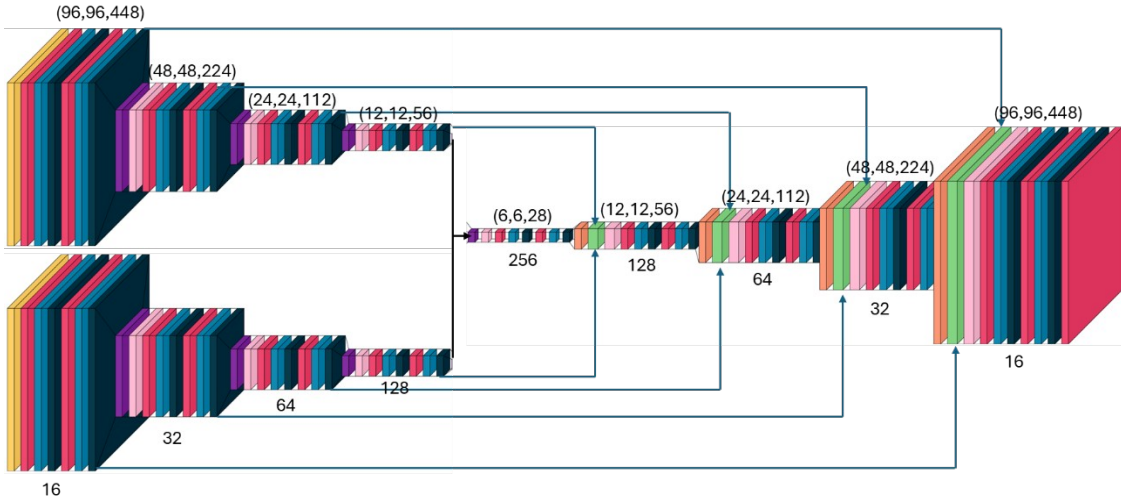
Figure 12. Cross-sectional views of true CO₂ model, predicted CO₂ model, their relative difference (%) and standard deviation. The dropout layers are inactivated during prediction and noisy test data are used (inversion experiment group 3).

Figure 13. Cross-sectional views of true CO₂ model, predicted CO₂ model, their relative difference (%) and standard deviation. The dropout layers are activated during prediction and noisy test data are used (inversion experiment group 4).

1 FIGURES



(a)



(b)

Figure 1. (a) The 3D U-Net architecture used for single-physics inversion and experimental joint inversion. (b) The modified 3D U-Net architecture for joint inversion. Blue lines indicate skip connections.

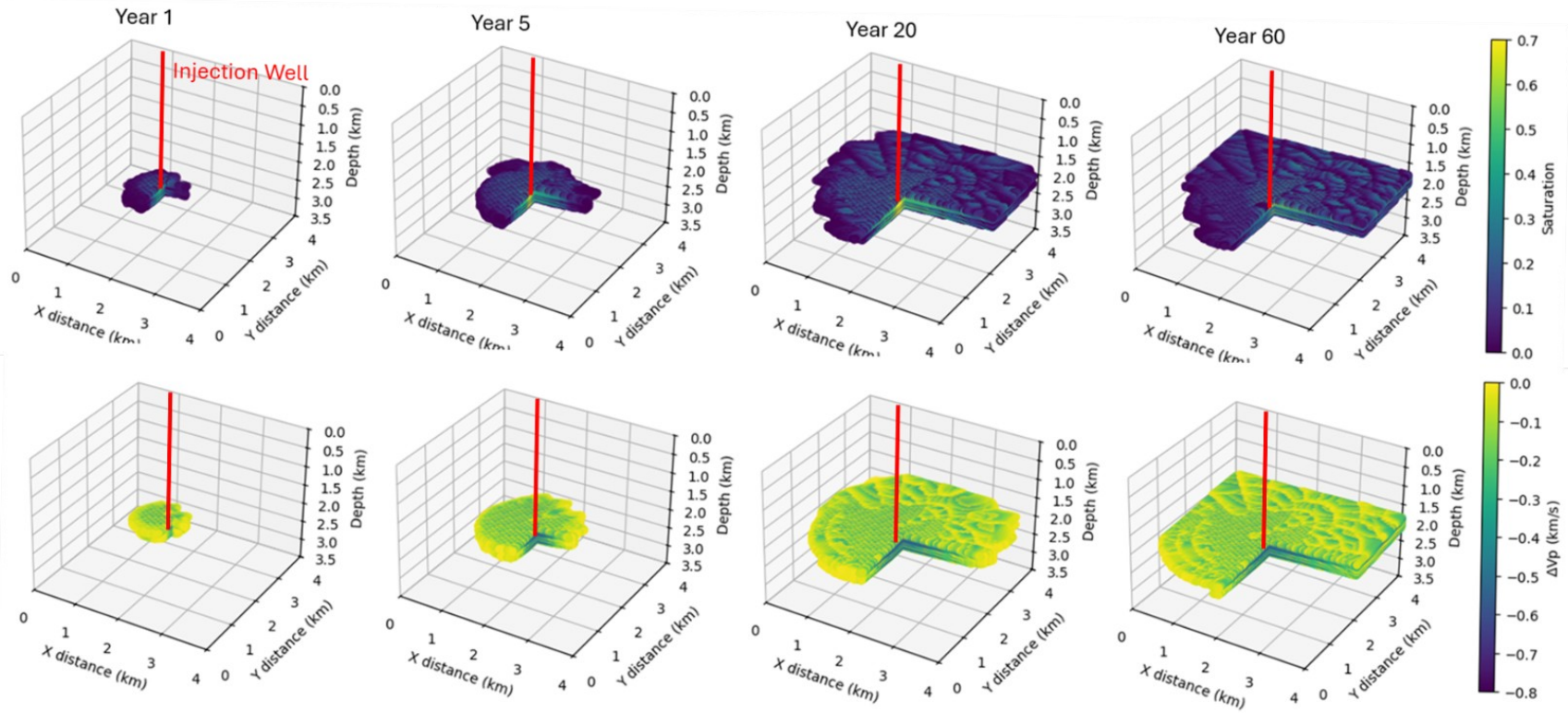


Figure 2. Evolution of CO₂ plumes and the corresponding changes in V_p over time. (Top) CO₂ saturation models. (Bottom) Changes in the V_p models due to CO₂ injection.

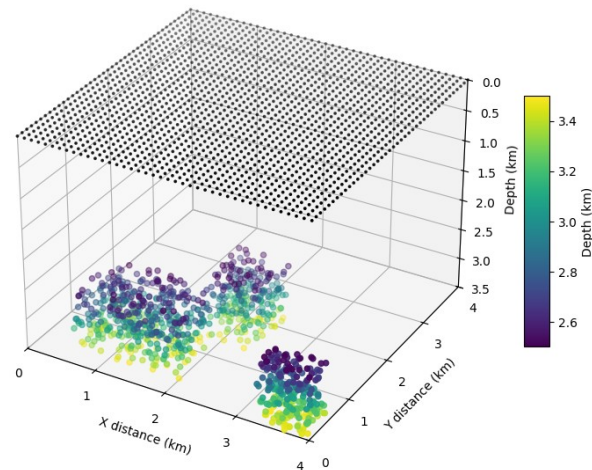


Figure 3. The locations of microseismic events in the Kimberlina model space. The color chart represents the depth of the events. The black dots on the surface of the earth represent geophone locations.

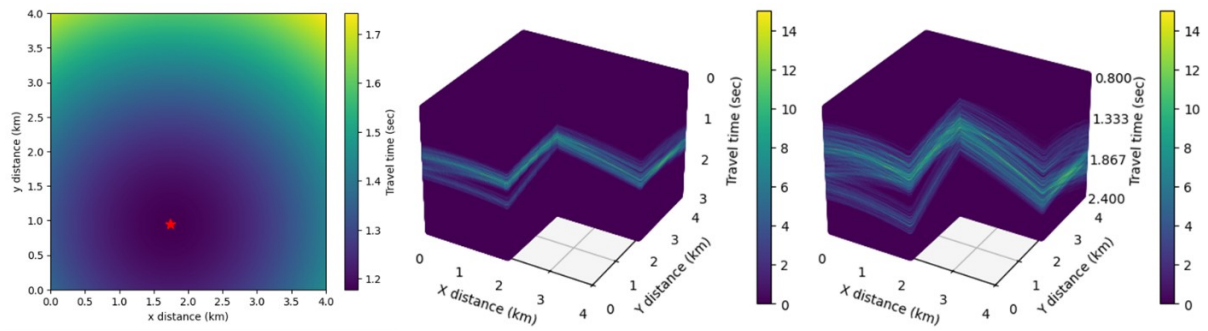
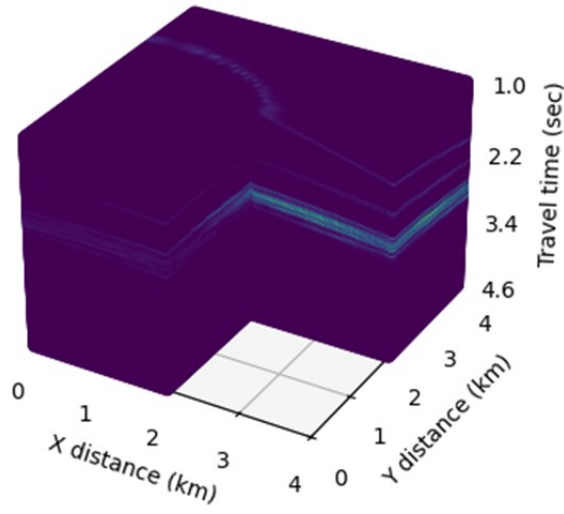
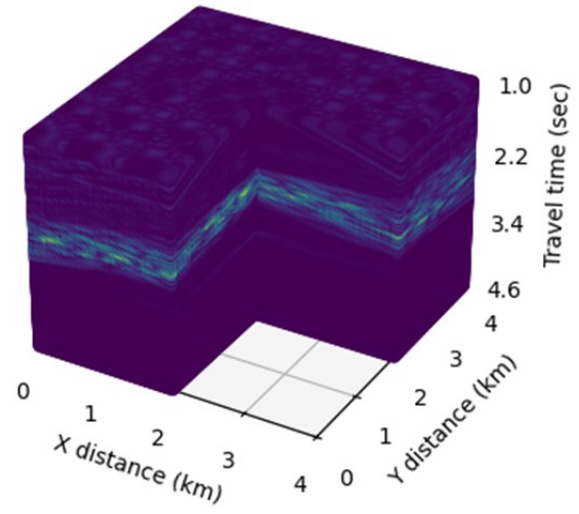


Figure 4. The 3D Kimberlina traveltimes data. (a) The traveltimes data from a single microseismic event whose location is denoted as a red star. (b) The 3D data cube containing 1000 microseismic traveltimes envelopes. (c) The final traveltimes data cube whose time axis spans from 0.72 seconds to 2.18 seconds.

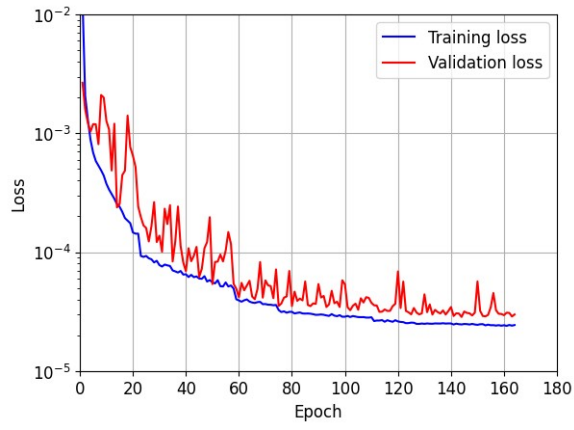


(a)

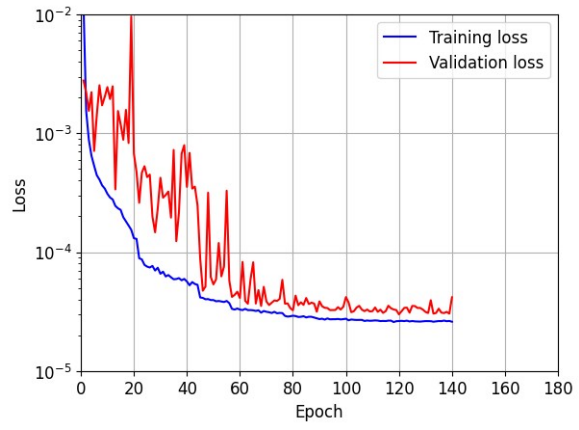


(b)

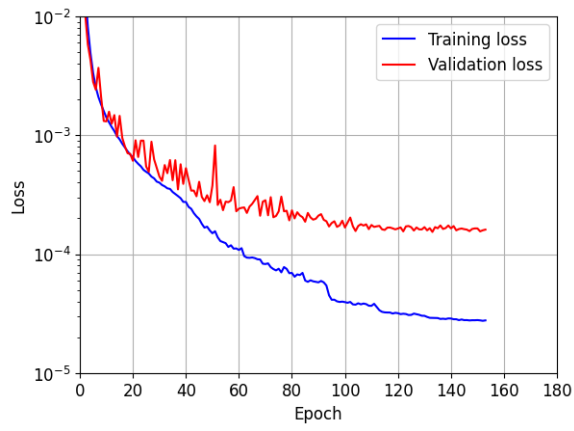
Figure 5. Construction of seismic full waveform input data. (a) Differential waveform data obtained by subtracting a shot gather at year 0 from the corresponding gather at year 20. (b) A compact representation formed by summing the differential data from all 25 shot gathers.



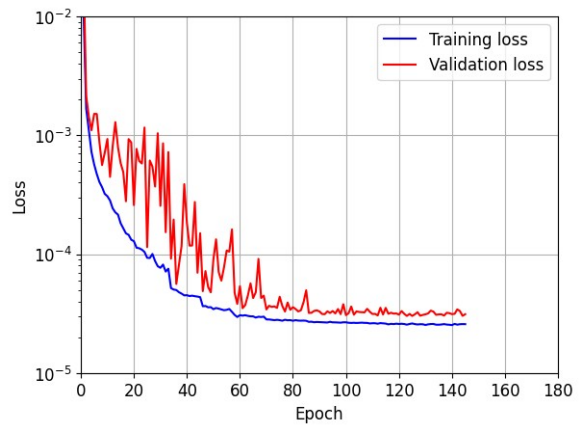
(a) FWI25



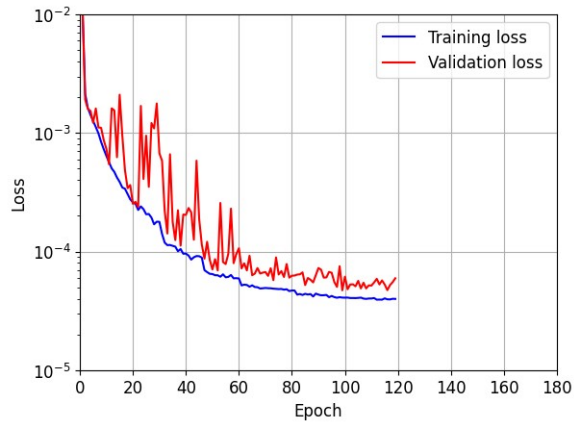
(b) FWI1



(c) Traveltime1



(d) JNT1



(e) JNT2

Figure 6. The training and validation loss plots for the five DL inversion experiments.

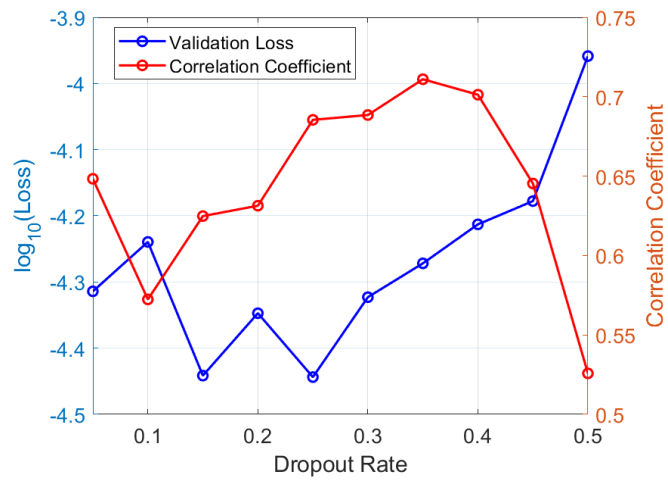
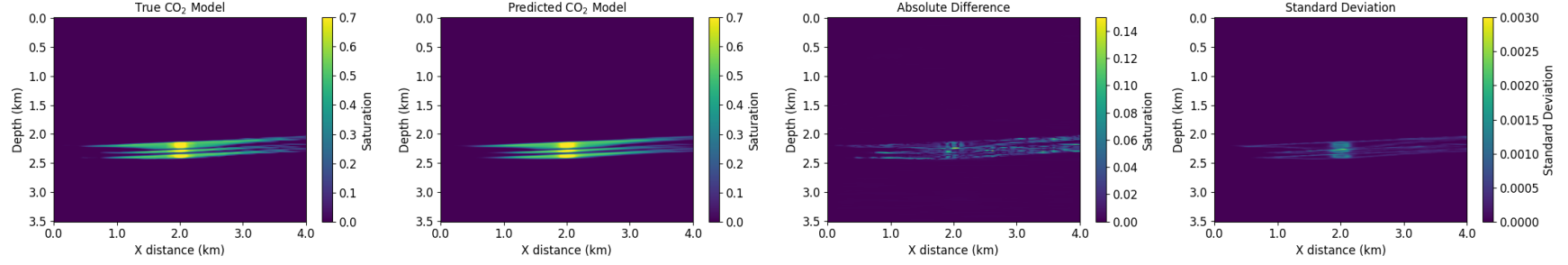
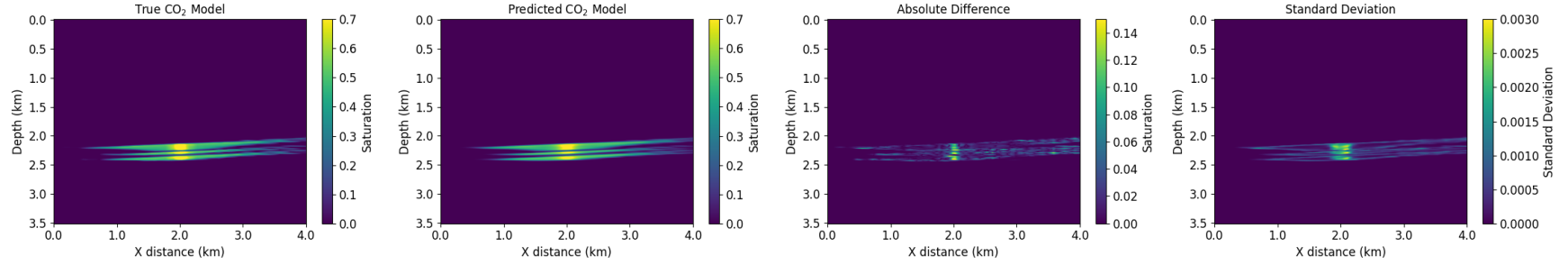


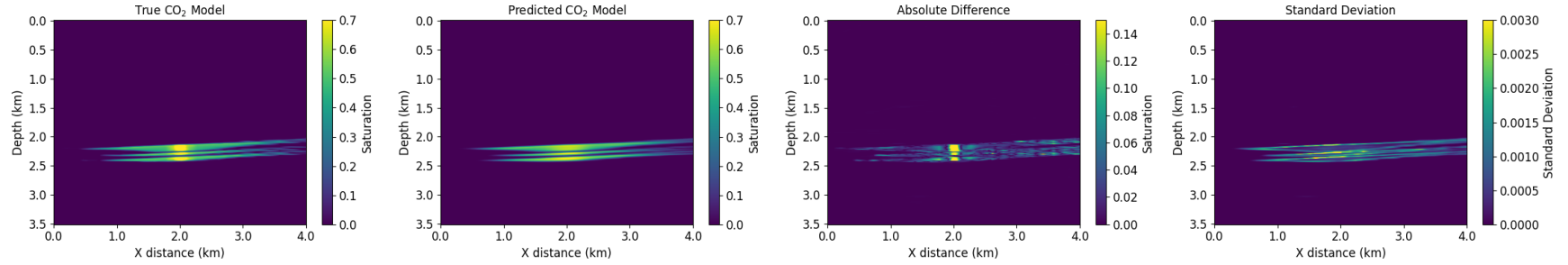
Figure 7. The smallest validation loss of FWI1 and the correlation coefficient between standard deviation and prediction error as a function of dropout rate.



(a) Dropout rate=0.05

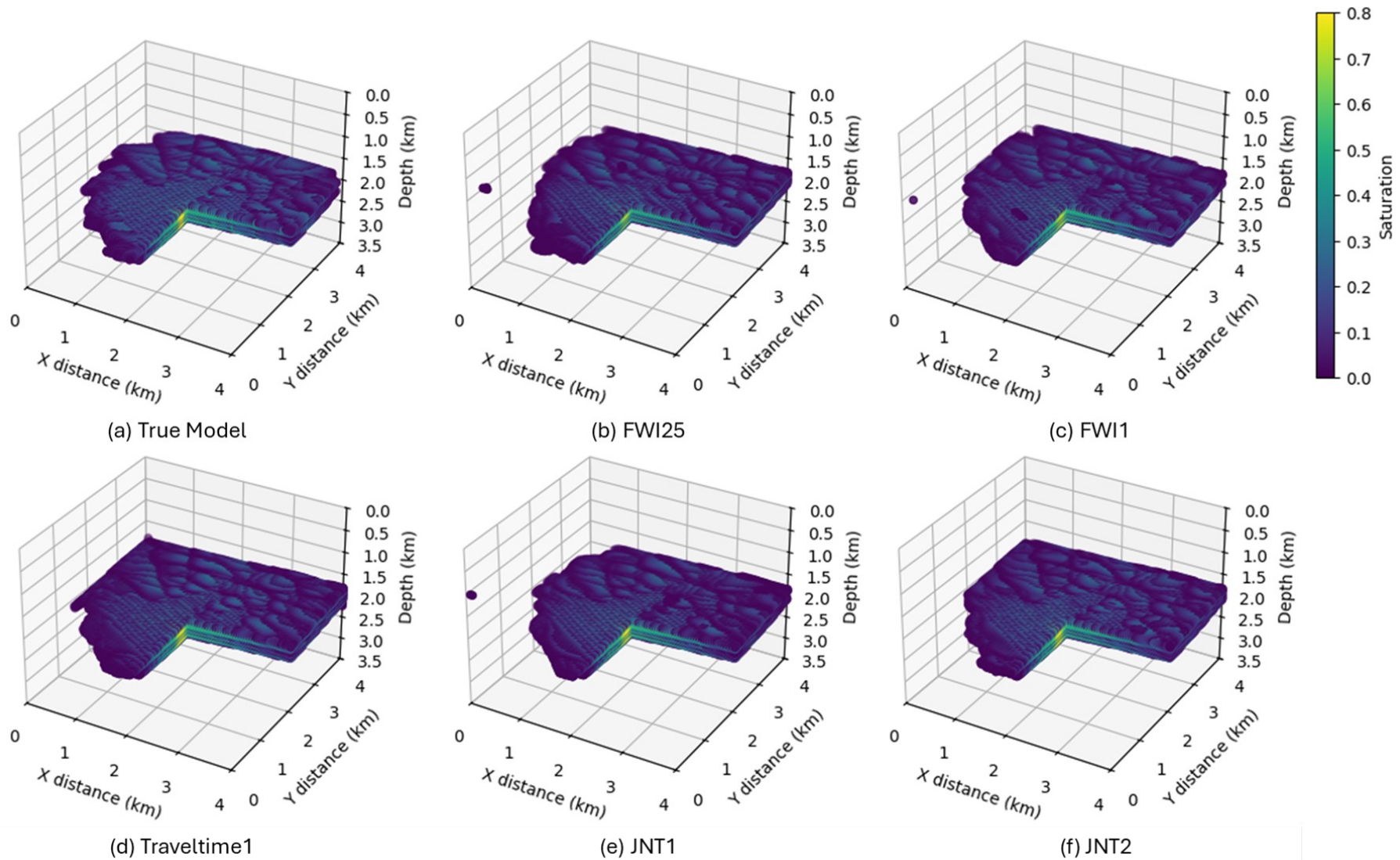


(b) Dropout rate=0.35

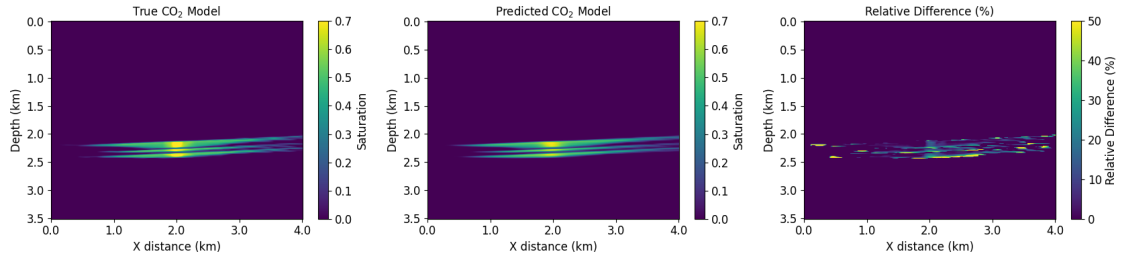


(c) Dropout rate=0.5

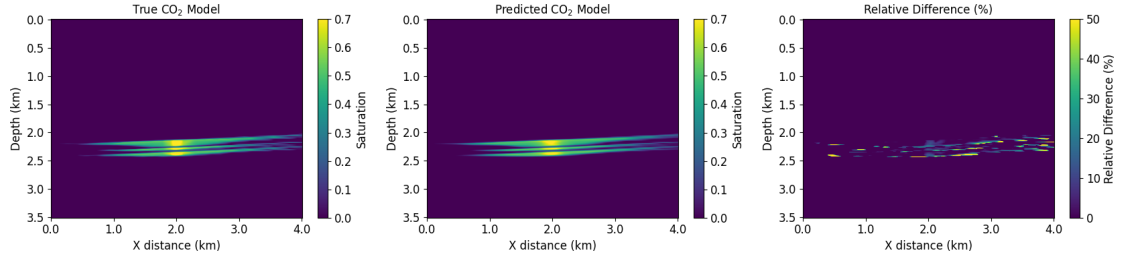
Figure 8. Comparison of true CO₂ model, predicted mean CO₂ model from FWI1, prediction error, and standard deviation at different dropout rates.



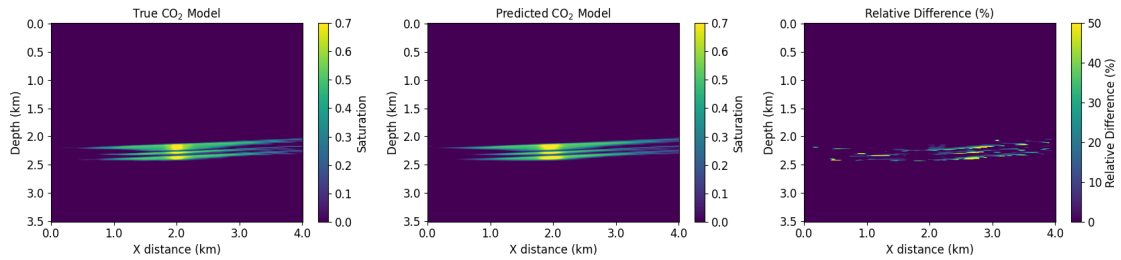
1
2 **Figure 9.** 3D views of the true CO₂ model and five different DL inversion results. The dropout layers are inactivated and noise-free test data are used.



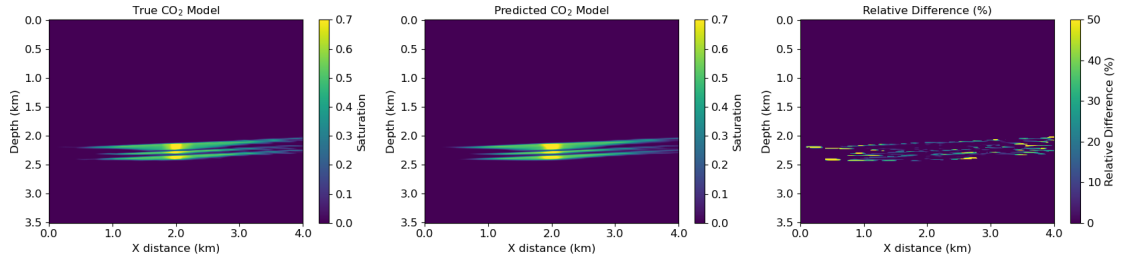
(a) FWI25



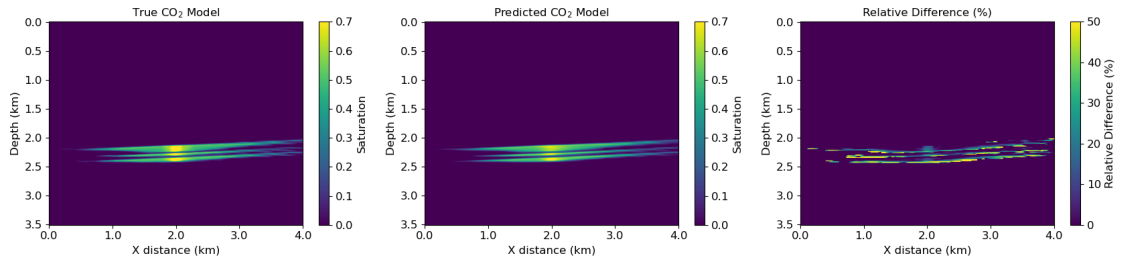
(b) FWI1



(c) Traveltime1

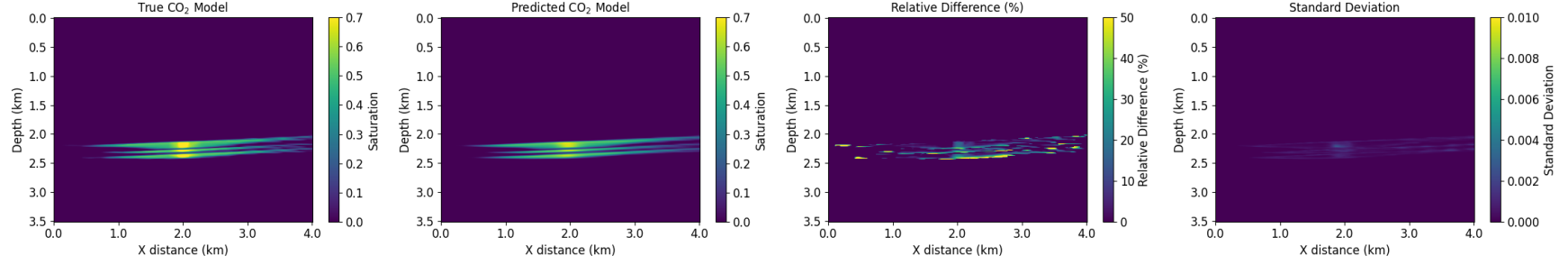


(d) JNT1

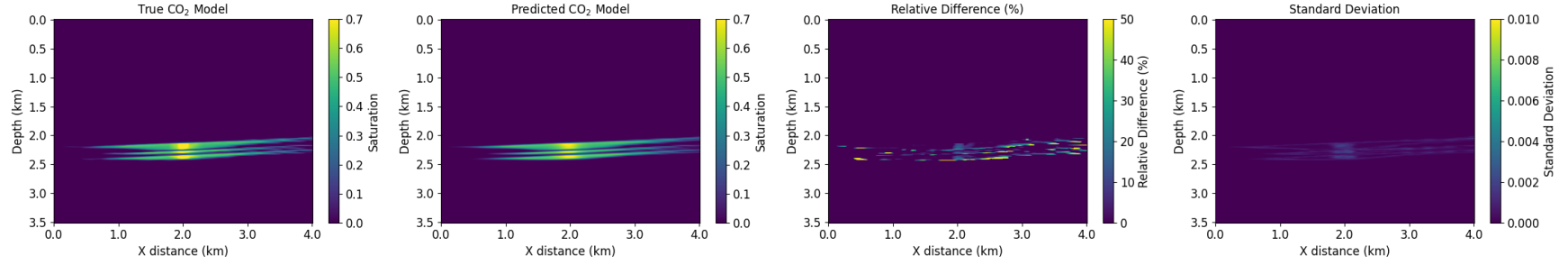


(e) JNT2

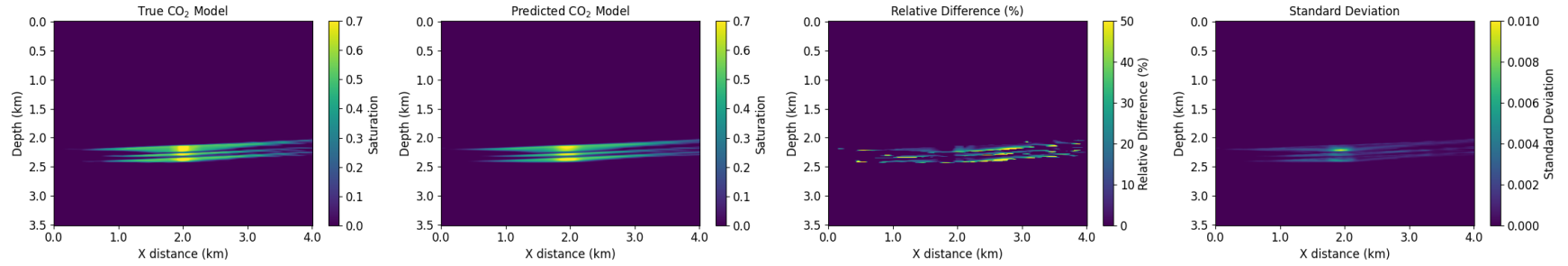
Figure 10. Cross-sectional views of true CO₂ model, predicted CO₂ model and their relative difference (%). The dropout layers are disabled and noise-free test data are used (inversion experiment group 1).



(a) FWI25

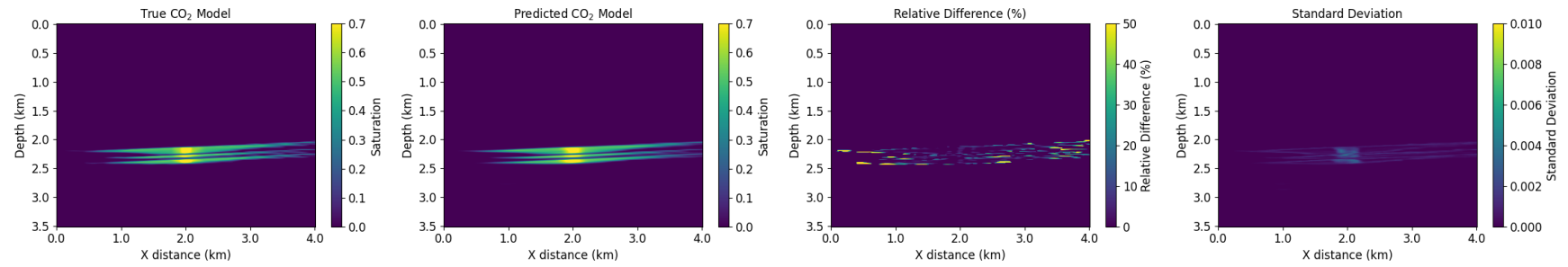


(b) FWI1

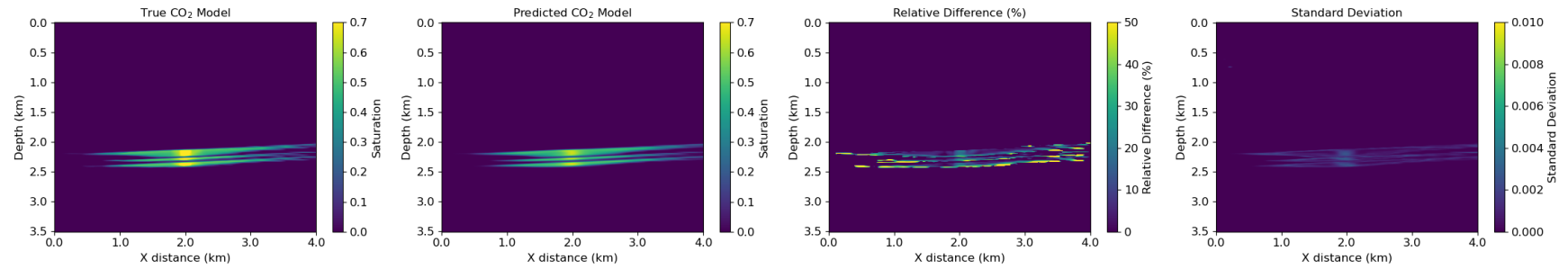


(c) Traveltime1

Figure 11. Cross-sectional views of true CO₂ model, predicted CO₂ model, their relative difference (%) and standard deviation. The dropout layers are activated during prediction and noise-free test data are used (inversion experiment group 2).

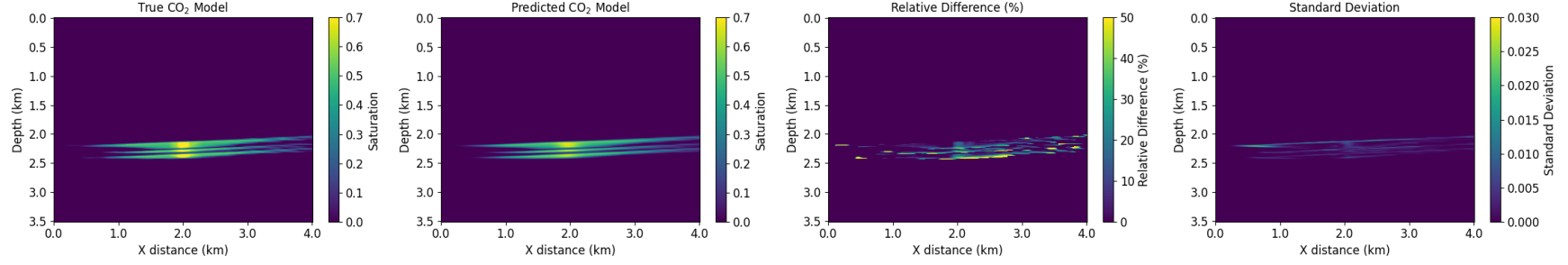


(d) JNT1

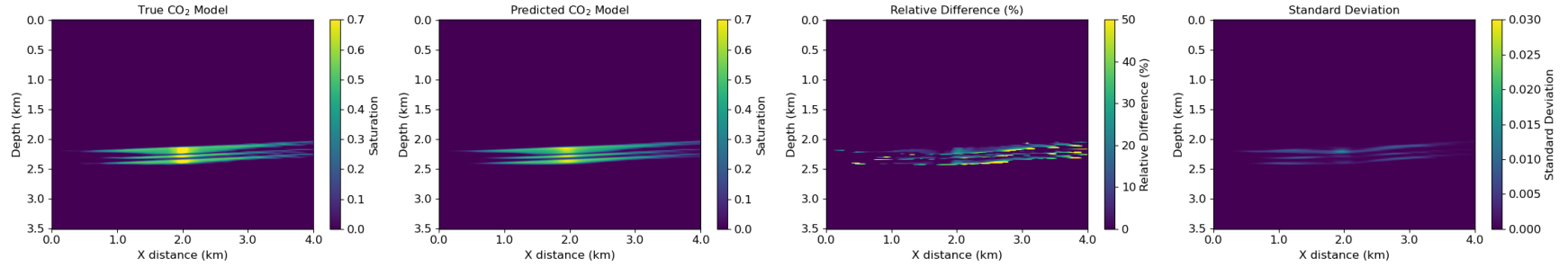


(e) JNT2

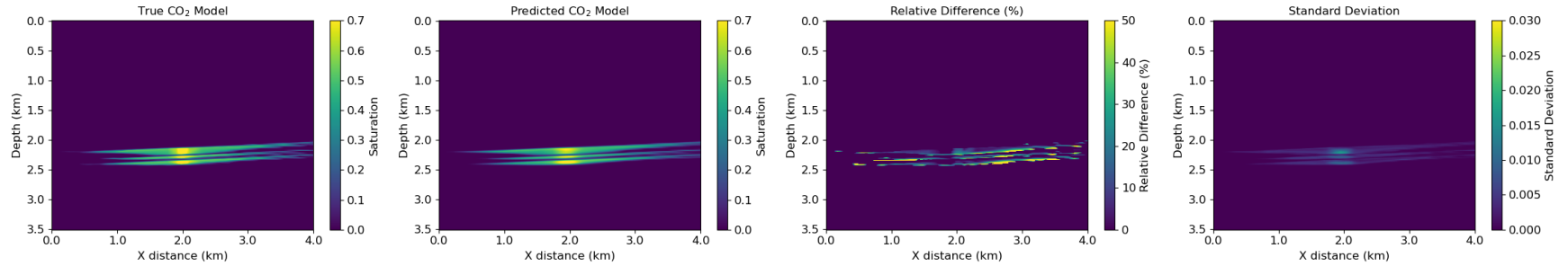
Figure 11. Continued.



(a) FWI25

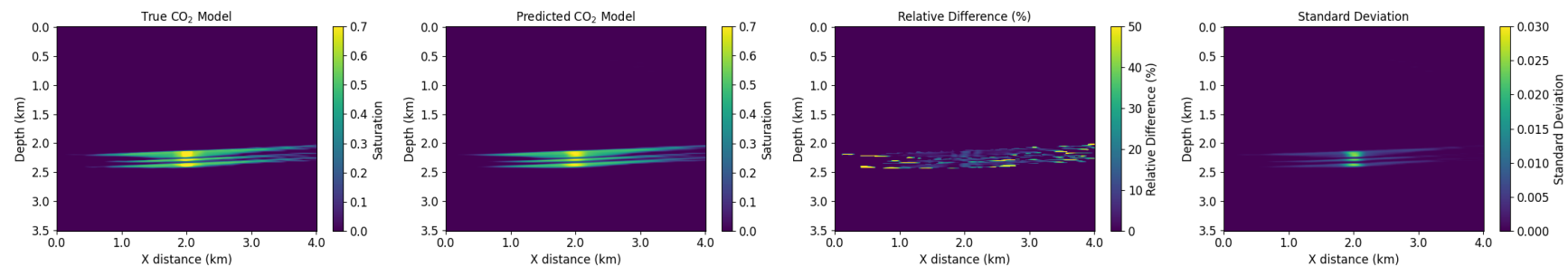


(b) FWI1

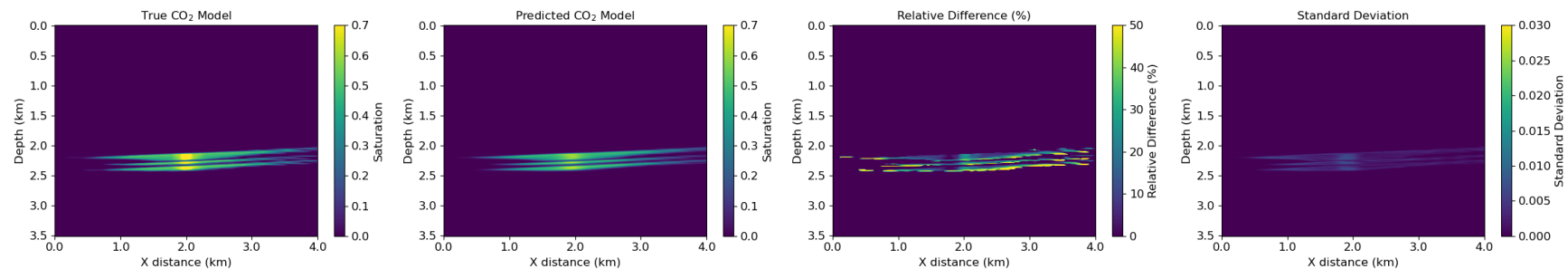


(c) Traveltime1

Figure 12. Cross-sectional views of true CO₂ model, predicted CO₂ model, their relative difference (%) and standard deviation. The dropout layers are inactivated during prediction and noisy test data are used (inversion experiment group 3).

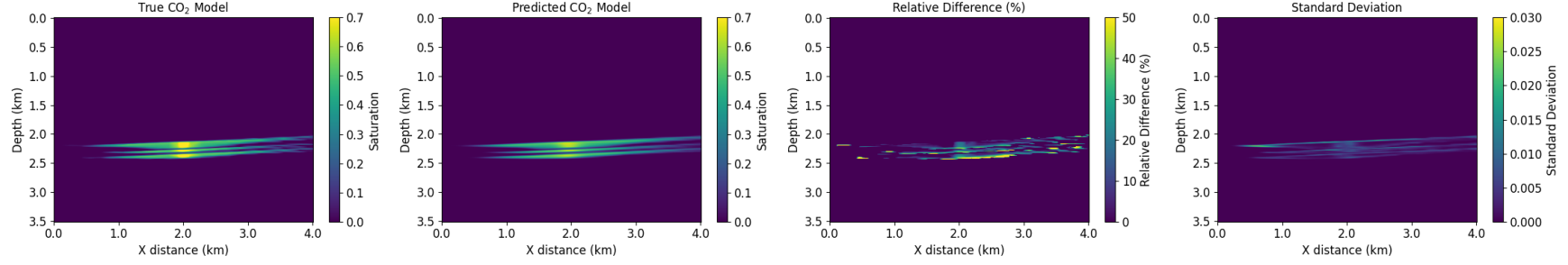


(d) JNT1

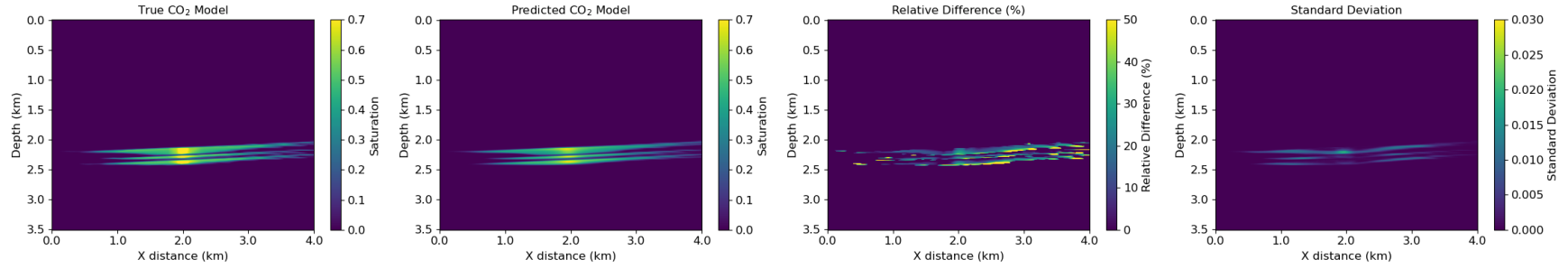


(e) JNT2

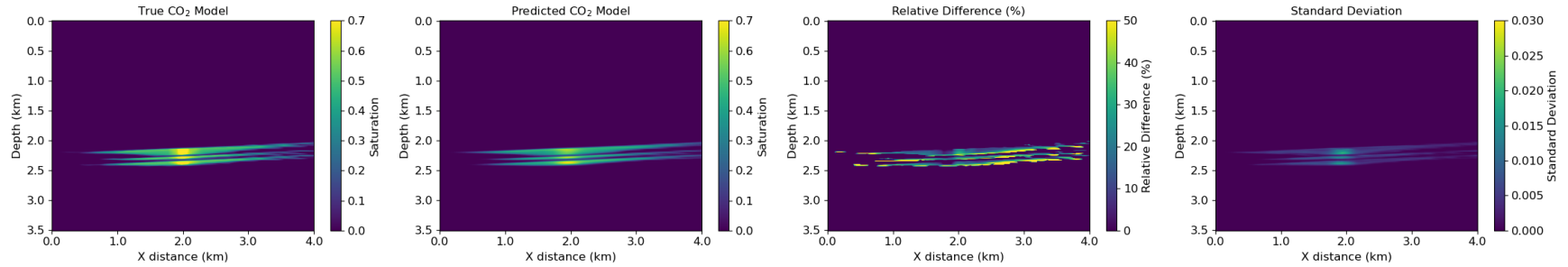
Figure 12. Continued.



(a) FWI25

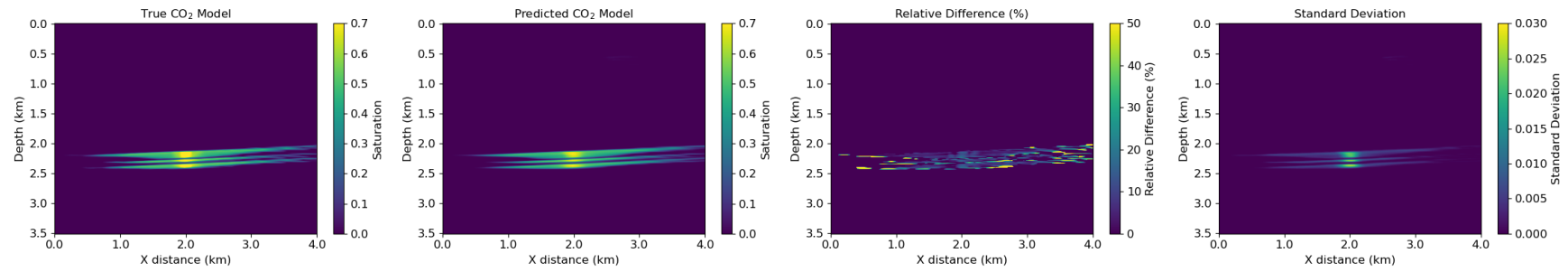


(b) FWI1

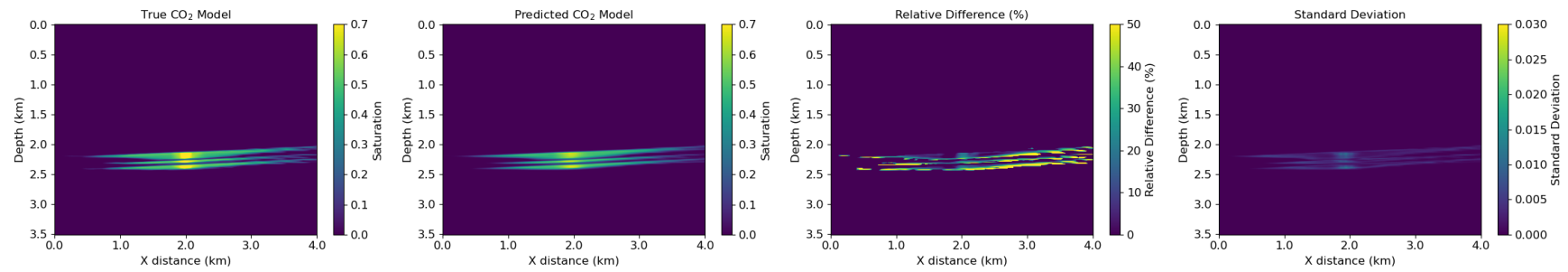


(c) Traveltime1

Figure 13. Cross-sectional views of true CO₂ model, predicted CO₂ model, their relative difference (%) and standard deviation. The dropout layers are activated during prediction and noisy test data are used (inversion experiment group 4).



(d) JNT1



(e) JNT2

Figure 13. Continued.