Title: Real time Deep Learning Inversion of Seismic Full Waveform Data for CO_2 Saturation and Uncertainty in Geological Carbon Storage Monitoring

Short Running Title: DL FWI for estimating CO_2 saturation and uncertainty

List of Authors:

Evan Schankee Um, Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA, USA, 94720, esum@lbl.gov

David Alumbaugh, Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA, USA, 94720, dlalumbaugh@lbl.gov

Youzuo Lin, Earth and Environment Division, Los Alamos National Laboratory, Los Alamos, NM, USA, 87545, ylin@lanl.gov

Shihang Feng, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA, 87545, shihang@lanl.gov

Conflicts of Interest

The authors declare no conflicts of interest in preparing this manuscript.

Abstract

Deep learning inversion has recently drawn attention in geological carbon storage research due to its potential of imaging and monitoring carbon storage in real time, significantly improving efficiency and safety of carbon storage operations. We present a deep-learning full waveform inversion method that after the neural network has been trained can image CO_2 saturation and its uncertainty in real time. Our deep learning inversion method is based on the U-Net architecture with the neural network trained on pairs of synthetic seismic data and CO_2 saturation models. Accordingly, our training establishes a mapping relationship between seismic data and CO_2 saturation models and once fully trained directly estimates CO_2 saturation as a function of subsurface location. We further quantify uncertainties of CO_2 saturation estimates using the Monte Carlo dropout method and a bootstrap aggregating method. For this proof-of-concept study, the CO_2 training models and data are derived from the Kimberlina 1.2 model, a hypothetical 3D geological carbon storage model that is constructed based on various geological and hydrological data from the Southern San Joaquin Basin, California. We perform deep-learning inversion experiments using noise-free and noisy training and test data sets and compare the results. Our modeling experiments show that 1) the deep-learning inversion can estimate 2D distributions of CO_2 fairly well even in the presence of Gaussian random noise and 2) both CO_2 saturation imaging and uncertainty quantification can be done in real time. Our results suggest that the deep-learning inversion method can serve as a robust real-time monitoring tool for geological carbon storage and/or other time varying reservoir/aquifer properties that result from injection, extraction, and/or other subsurface transport phenomena.

Keywords: Full Waveform, Inversion, Monitoring

Data Availability Statement

The geophysical models and data used in this work will be available to the public via U.S. Department of Energy's NETL EDX system, https://edx.netl.doe.gov/.

Introduction

Carbon capture and geological carbon storage (GCS) are a viable technology for reducing CO_2 emission into the atmosphere (Metz et al., 2005; Benson and Cole, 2008; Davis et al., 2019; Ringrose, 2020). High concentrations of CO_2 are captured from large-scale industrial sources such as factories and power plants and then injected to deep depleted oil/gas reservoirs or saline aquifers for permanent storage. The large-scale injection of CO_2 into a deep reservoir over a long period changes geomechanical and hydrological states inside and around the reservoir (Rutqvist, 2012). For successful and safe GCS operations, therefore, it is imperative to improve subsurface monitoring methods for tracking CO_2 plumes, evaluating cap rock integrity and early detecting CO_2 leak from a reservoir.

Seismic methods have been widely used for monitoring CO_2 plume migration and detecting leakage. For example, 3D surveys are repeatedly performed at an interval of a few years to detect changes in migrated reflection events (e.g., travel times and amplitudes) associated with CO_2 saturation (Chadwick et al., 2010; Lumley, et al., 2010). Geophysical inversion (Tarantola, 2005; Sen and Stoffa, 2013; Aster et al., 2018) provides a quantitative way to construct a subsurface velocity model that explains recorded seismic data. For example, first-arrival travel time inversion is commonly applied to crosswell seismic data for constructing seismic velocity models (Lazaratos and Marion, 1997; Ajo-Franklin et al., 2013). Time-lapse analysis of these data types allows us to quantitatively monitor changes in seismic velocity related to CO_2 saturation. Seismic full waveform inversion (FWI) is one of the most advanced seismic tomography methods and has been recently applied to CO_2 monitoring problems (Arts et al., 2003; Zhang et al., 2012; Queißer et al., 2013; Dupuy et al., 2017; Egorov et al., 2017). Compared to partial waveform inversion such as first arrival travel time inversion, FWI minimizes misfits between recorded and predicted full seismograms in an iterative manner (Virieux and Operto, 2009; Fichtner 2010; Virieux et al., 2017), reconstructing a high-resolution velocity model. Time-lapse FWI is capable of monitoring subtle changes in reservoir properties associated with CO_2 migration and leakage (Raknes et al., 2015; Li et al., 2021).

Despite the advantages in resolution that they provide over other geophysical techniques, seismic imaging methods have their own drawbacks. In general, the conventional seismic imaging methods for CO_2 monitoring require a significantly large amount of time and labor for processing 3D seismic data (Arogunmati and Harris, 2012). Running FWI on the data is further involved from a computational standpoint. For example, FWI is highly non-linear and often gets stuck within local minima (Shin and Cha, 2008; Shin and Cha, 2009). FWI also requires a significant amount of computer resource and time especially for 3D problems because it updates a velocity model in the course of inversion by repeatedly solving a full wave equation which requires fine temporal and/or spatial discretization (Fichtner 2010; Um et al. 2011; Petrov and Newman, 2012).

Statistical methods play an important role in GCS studies because they can provide probability distribution of CO_2 saturation given geophysical monitoring data and quantify uncertainties associated with CO_2 plume imaging (Grana et al., 2021). Applying traditional uncertainty quantification methods such as Markov Chain Monte Carlo methods to large-scale seismic imaging problems is often challenging because of significant computational cost and time (Hunziker et al., 2019; Zhao and Sen, 2021), and because of this, 3D seismic inversion methods tend to remain deterministic at this time. The large data-processing and computational cost mentioned above makes traditional seismic imaging methods less practical as a real-time monitoring and uncertainty analysis tool.

In recent years, deep learning (DL) inversion has drawn attention in seismic inversion as it overcomes some of the drawbacks of the traditional FWI mentioned above (Araya-Polo et al., 2018; Yang and Ma, 2018; Wu and Lin, 2019; Zhang and Alkhalifah, 2019; Colombo et al., 2020; Oh et al., 2020; Mosser et al., 2020; Zhang and Lin, 2020, Kaur et al., 2021; Kazei et al., 2021; Li and Yang, 2021). DL inversion requires training a neural network using realistic subsurface models and data. It is still computationally challenging to build a number of realistic velocity models and solve a full wave equation for each model. However, once the training process is completed, the trained network can be applied to recognize complex non-linear correlations between the models and the data in nearly real time. The network can instantaneously predict a velocity model from newly acquired and unprocessed seismic data. Recent advances in DL research also make it possible to perform a certain type of uncertainty quantification analysis without significant computational burdens as will be shown in this paper. More specifically, compared to the traditional FWI inversion, the DL inversion requires most of high computational costs in advance before new data are collected thus allowing for near-real time imaging since selecting and evaluating a candidate GCS site can require extensive flow and geophysical modeling work. The resulting models and data can be used for training a neural network for DL inversion. These characteristics makes DL inversion well suited for imaging and monitoring GCS in real or near real time.

DL inversion for monitoring GCS is relatively new but has been an active area of research in recent years. For example, Puzyrev (2019) trains a fully convolutional network (FCN) using CO_2 storage models with numerically computed time-lapse electromagnetic (EM) data and show that the deep learning EM inversion can track the movement of a CO_2 plume. Li et al. (2021) trains an FCN using time-lapse seismic difference data and reservoir velocity differences, showing that the trained network can monitor velocity changes in a reservoir due to CO_2 injection. Using FCN and long short-term memory networks, Zhou et al. (2019) show that DL inversion can establish a direct mapping relationship between seismic data and CO_2 leakage mass, providing an end-to-end detection approach.

Estimating the uncertainty associated with a CO_2 image is another active open GCS problem and deep neural networks can be used for the purpose. For example, Grana et al. (2020) develop an inversion approach of multiphysics data that combines geostatistical methods, stochastic optimization and DL (i.e., the deep convolutional auto-encoder) methods for predicting CO_2 saturation and its uncertainty. Kaur et al. (2020) derives CO_2 saturation and velocity models from a range of porosity and permeability values expected for a CO_2 reservoir and subsequently train deep neural networks (i.e., generative adversarial networks) on pairs of the CO_2 models and seismic data to predict CO_2 saturation and estimate its uncertainty. Tang et al. (2021) propose an efficient GCS forecasting workflow that updates reservoir properties including CO_2 saturation and predicts their associated uncertainties using an ensemble smoother with multiple data assimilation framework and deep neural networks (i.e., U-Net) as a surrogate model.

In this paper, we present a DL-based full waveform inversion for monitoring GCS. Our DL inversion utilizes a U-Net architecture (Ronneberger et al., 2015). The U-Net was originally developed for biomedical image segmentation and has been also used in geophysical imaging research (Yang and Ma, 2018; Colombo et al., 2020; Oh et al., 2020). In this work, the neural network is trained on pairs of seismic full waveform data and CO_2 saturation models rather than pairs of synthetic seismic data and velocity models, thus directly imaging the distribution of CO_2 saturation inside a storage. We further quantify uncertainties associated with a CO_2 saturation image using the Monte Carlo dropout method (Gal and Ghahramani, 2016; Zhu et al., 2022) and the bootstrap aggregating (bagging) method (Friedman et al., 2016; James et al., 2021) and compare the two uncertainty quantification results.

Deep Neural Network

Based on the U-Net, we build a deep neural network for predicting a CO_2 saturation model (**Figure** 1). The network consists of a contraction section, a bottom section and an expansion section. The contraction section repeatedly applies a 2D convolution operation followed by batch normalization, rectified linear activation (ReLU) and max-pooling operations. The convolution operation consists of many convolution filters and extracts feature maps that highlight regions of the input layer that resembles each filter (James et al., 2021). In contrast, the max pooling operation down-samples the feature maps and reduces a large feature map to smaller summary map. The batch normalization normalizes the output maps using the mean and variance and improves the gradient propagation. Our choice of activation function is ReLU, which allows better gradient propagation because the function is not saturated when an input value is large.

After each max-pooling operation, the spatial information is reduced by half and the feature information is doubled, allowing the network to effectively learn complex structures. The contraction section repeats the same procedures four times. The number of repetitions for successful U-Net performance depends on the specific problem being addressed, and are determined through network design experiments. The bottom section consists of two convolution blocks without a max-pooling operation and connects the contraction section to the expansion path. Similar to the contraction section, the expansion section consists of four repetitions. After each inverse convolution, the spatial information is doubled and the feature information is reduced by half. As the network is symmetric, the expansion section can concatenate its feature maps with those from the contraction section, improving information flow between the two paths. For details on the operations mentioned above, the reader is referred to Goodfellow et al. (2016), Chollet (2017), Oh et al. (2020) and James et al. (2021).

To predict a CO_2 saturation model using seismic full waveform data, we make several changes to the original U-Net as described below. First, we use a sigmoid activation function (Chollet, 2017) in the final layer and constrain an output CO_2 saturation value between 0 and 1. Second, the channel of the input layer is also changed from 3 (i.e., red, green and blue in a color image) to N which is the number of common-shot gathers within the data. Third, we add a dropout layer (Srivastava et al., 2014; Goodfellow et al., 2016; James et al., 2021) in each block. The dropout method is a regularization technique for mitigating overfitting a neural network to training data. When the dropout layer is turned on during training, it randomly sets some output features of the layer to zero. In contrast, the dropout layers are disabled during prediction. Thus, the entire network takes a part in predicting a model to improve the accuracy. This approach reduces the co-adaptation between the network and training data, reducing overfitting. A higher dropout rate can better mitigate overfitting but may deteriorate the network accuracy. A proper dropout rate is selected by trial and error.

We also use the dropout method for quantifying uncertainty in a deep neural network using the Monte Carlo (MC) dropout method of Gal et al. (2016). They show a theoretical link between Gaussian processes and dropout, which makes it possible to extract the information without increasing either the computational cost or prediction accuracy. For this purpose, the neural network is trained in the same way as previously described except that during the prediction phase, the dropout layers are not disabled but rather are activated. This operation is used for approximating variational inference in Bayesian neural networks. In other words, the network generates a different predicted model at every DL inversion pass for the same test data because some parts of the trained network are randomly dropped during the prediction phase. After a sufficient number of DL inversion runs, we have a distribution on CO_2 saturation at each cell of a predicted model and can calculate a mean CO_2 saturation and its standard deviation.

Last, we implement a 'bagging' method (Friedman et al., 2016; James et al., 2021) by adding a bootstrap layer to the input layer of **Figure** 1. This algorithm is an ensemble-learning technique and was originally developed for improving predictions by reducing variance. The bagging method is particularly useful in a relatively small training data set environment such as those found in geophysical applications. The method can also be used for quantifying the uncertainty associated with a given estimator, which in this paper is the neural network. The bootstrap creates distinct pairs of training models and data by repeatedly sampling pairs from the original set with replacement. Sampling with replacement means that some pairs of models and data can be drawn more than once in a bootstrap data set, while some are never drawn. The bagging method repeats creating a distinct bootstrapped training data set and training a neural network with the bootstrapped data. During each training phase, the pairs of models and data that are not drawn from the original training set are used for validating the current results. Finally, each trained neural network independently predicts a CO_2 model. Using the ensemble of models, we calculate a mean CO_2 saturation and its standard deviation.

As described above, both methods provide uncertainty estimation but work differently. For the MC dropout method, we train only one full neural network and then repeatedly predict a CO_2 model for the same test data using a different 'sub-network' out of the parent network. The MC dropout method estimates uncertainty related to the neural network model error, and this method does not require much in the way of additional computational costs and complexity. In contrast, the bagging method uses bootstrapping as a sampling method and create many unique 'sub-data sets' out of the original data sets. Thus, we separately train neural networks using the unique sub-data sets. Finally, we perform prediction in those networks in parallel. Therefore, compared to the MC dropout method, the bagging method requires much larger computational cost and is often difficult to scale up to very large DL inversion problems.

Kimberlina Velocity Models

One major challenge for DL geophysical inversion is to create a large number of realistic subsurface models (i.e., labels) and simulate their geophysical responses (i.e., data) for training purposes. For realistic DL inversion and uncertainty quantification experiments in this paper, we derive training models and data from the 3D Kimberlina 1.2 CO_2 reservoir model (Birkholzer et al., 2011; Zhou and Birkholzer, 2011, Wainwright et al., 2013). The Kimberlina reservoir model was built to simulate a potential commercial-scale GCS in the Southern San Joaquin Basin of California, 30 km northwest of Bakersfield, CA, USA. For realistic modeling of CO_2 movement and leakage, the models are constructed based on various geologic and hydrogeologic data obtained from many oil wells in the region. TOUGH2 (Pruess et al., 1999), a modeling software for nonisothermal flows of multicomponent, multiphase fluids in porous and fractured media was used for simulating CO_2 injection and flow in the Kimberlina model.

The Kimberlina flow model we employed for training consists of a single case from 300 different realizations that were run. Note that a single realization outputs a 3D volume of simulated hydrologic properties at 35 different time steps from 0 to 200 years. The simulations were run with a single CO_2 injection well operating for 50 years at an injection rate of 5 Mt per year, and a post-injection period of 150 years. The hydrological and petrophysical properties of the 35 reservoir time-stamps are converted to P wave velocity models (Alumbaugh et al., 2021) using a conversion workflow described in Wang et al.

(2018) and Yang et al. (2019) and petrophysical relationships (Mavko et al., 2020). As a proof of concept and to make the computational requirements tractable, we performed these initial DL inversion experiments using 2D models and data. This involved slicing the 3D models at 100m intervals in the *x* direction to create 53 2D velocity and CO_2 saturation models at each time step. Thus, the one Kimberlina realization produces 1855 pairs of 2D velocity and CO_2 models. **Figure** 2 shows several pairs of velocity and CO_2 models over time.

Each 2D velocity model is 6 km in the horizontal direction and 3.5 km in the vertical direction, with a grid size of 10m by 10m. 6 sources are located on the surface with a spacing of 1 km (i.e., 0, 1, 2, 3, 4 and 5 km), and traces are recorded by 600 receivers spaced at an interval of 10 m. We chose a Ricker wavelet with a peak frequency of 25 Hz as the source function to generate seismic waves due to its empirical success in processing seismic fields (Fichtner, 1993). The seismic data were simulated using the finite difference approximation (Alford et al, 1974) to the acoustic wave equation. An example of common shot gathers for a single 2D velocity model is given in Figure 3. The total recording time is 5 s with a timestep of 0.0005 s. In the training of the network, the common shot gathers and label CO_2 models are down-sampled to (512, 256) due to the memory and computational cost issues. The threshold level of down-sampling depends on the nature of the data and label models and should be carefully chosen by trial and error. Without further data processing, the down-sampled common shot gathers are directly fed into the neural network for training and prediction. This simple data processing further helps us to quickly prepare the input data for real-time imaging as will be demonstrated later.

As a final note, the governing physics of CO_2 flow and seismic wave propagation is not explicitly implemented inside the DL inversion network (Figure 1). Instead, the governing physics is implicitly embedded into the training seismic data and their corresponding CO_2 label models because they are generated using the realistic flow and seismic modeling algorithms described above. In other words, the network is trained on the sets of training data and label models that honor the governing physics and thus is indirectly informed about the physics. As a side note, one can further integrate the governing physics into a neural network by embedding a partial differential equation of the governing physics and its boundary and/or initial conditions into the loss function of a deep neural network. For details, the reader is referred to Raissi et al. (2019), Waheed et al. (2021) and Karniadakis et al. (2021). This physics-informed network is required for inverting seismic data from the real world. In contrast, the traditional FWI honors the physics of wave propagation by solving a wave equation at each inversion iteration. This is one major difference between traditional geophysical inversion and our DL inversion.

Deep Neural Network Training

For implementation, the U-Net algorithm has been written in Python using Keras library that includes various neural-network building blocks such as layers, activation functions and optimization tools. During the training phase, the U-Net solves an optimization problem for minimizing a misfit between training and predicted CO_2 saturation models. This is another interesting difference between DL and traditional inversion because traditional inversion minimizes a misfit between predicted and recorded 'data'. The misfit between the two models is measured using a loss function. A proper loss function is different for a given problem and is chosen by trial and error. After testing several loss functions, we choose to use a mean squared error (MSE) for our problem:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(m^{training} - m^{predicted})^2$$

where $m^{training}$ and $m^{predicted}$ are a training and predicted CO_2 saturation model, respectively.

Before training the neural network, the pairs of data and models (i.e., labels) are split into three groups: test, training and validation sets. The data and models for year 20 are selected and reserved for the test sets as a sufficient amount of CO_2 is observed inside the reservoir after 20 years of injection (Figure 2). These data and models are used only for evaluating the performance of the neural network 'after' training is completed. For the MC dropout method, 80% of the remaining data and models are randomly selected for training the neural network and 20% for validating the neural network ‘during’ training. The validation data set provides an initially unbiased evaluation of a model fit on the training set while tuning the neural networks. The evaluation becomes biased over iterations because the validation data set is gradually incorporated into the neural network. **Figure** 4a shows the training and validation losses history using the Adam optimizer (Kingma and Ba, 2015) with a batch size of 32, a learning rate of 1×10^{-4} and the dropout rate of 0.5. Note, the learning rate is a settable parameter inside an optimizer that determines a step size when the optimizer attempts to move toward the minimum of a misfit function. Both learning and dropout rate are determined by trial and error. Once fully trained, the neural network repeatedly predicts a CO_2 saturation model using the same test data, and we generate 400 predicted models for each test data. The predicted models are different from each other because the dropout layers are activated during prediction. To present the final results we calculate a mean CO_2 value and a standard deviation within each cell of the 2D section.

For the bagging method, a training set is randomly sampled from the original training set with replacement. The size of the bootstrapped training set is the same as that of the original set. With this resampling, some pairs of seismic data and models are drawn more than once and some are never drawn. On average, each bootstrapped training set makes use of about two-thirds of the original training set (Friedman et al., 2016, James et al., 2021). The remaining roughly one-third of the training set are referred to as the out-of-bag (OOB) set. The OOB data set is then used for validation during training. Because each neural network is trained on its own unique bootstrapped training set, the number of epochs required for the convergence varies for each training as shown in **Figure** 4b.

Training individual neural networks for the bagging method are embarrassingly parallel. Thus, several networks can be trained simultaneously when a GPU cluster is available. In our case, it took approximately 80 hours to fully train 400 neural networks for the bagging method on eight NVIDIA Tesla P100-PCIE GPUs. Nonetheless, the total computational cost is significantly more expensive than that for the MC dropout method that requires training only one neural network. Therefore, it appears less attractive to use the bagging method for large-scale DL inversion problems. However, the bagging results are used here as a benchmark to evaluate the uncertainty estimated by the MC dropout method. As a side note, once fully trained, both the MC dropout and bagging methods can make a single prediction of CO_2 saturation in a few seconds. The bagging uncertainty estimation which is based on 400 trained neural nets and resulting predictions can be completed in a few minutes.

For comparison purposes, we implement and train a single U-Net without an uncertainty quantification method. This is the simplest form of the U-Net for predicting a CO_2 model when seismic data are provided. **Figure** 4c shows its training and validation-loss history. The neural network is trained using the same stopping criterion and dropout rate used earlier. The dropout layers are activated for regularization during training but are deactivated during prediction in order to improve the prediction accuracy. The predicted CO_2 saturation models are deterministic and do not provide an uncertainty estimate.

Estimating CO_2 Saturation and Uncertainty

To evaluate the performance of the DL inversion methods, we predict CO_2 saturation using the MC dropout, bagging and single U-Net method and compare imaging results to the true Kimberlina model. In the first set of DL inversion examples, the neural networks are trained on a noise-free training data set and predicts CO_2 saturation models using the noise-free test data set. **Figure** 5 compares the true and predicted CO_2 saturation models at three different *y-z* sections. For the single U-Net method (the 2nd column of **Figure** 5), the predicted models show deterministic CO_2 saturation estimates after a single inversion run. For the MC dropout and bagging method (the 3rd and 4th column of **Figure** 5), the predicted models show the mean values of 400 predicted CO_2 saturation models.

The first y-z section (x=-1km) is extracted along the edge of the injected 3D CO_2 plume while the other two sections (x= 0 and 1km) were extracted from the central part of the injected volume. As expected since the imaging algorithms are trained using the same set of CO_2 saturation models, the three different neural networks show consistent imaging results. For example, the thickness of the CO_2 plume is consistent as is the vertical resolution of the three different layers of CO_2 that represent the three different high porosity layers within the reservoir and are recovered fairly well. At the edge of the plume (the 1st row of **Figure** 5), the single U-Net and the bagging method best resolve the three layers but the MC dropout method less clearly images the 2nd layer of CO_2, at least using the color scales applied here. For more detailed observation of the recovered CO_2 images, **Figure** 6 compares 1D CO_2 saturation profiles between the true model and the three recovered models at selected locations shown in **Figure** 5e-h. In general, the recovered CO2 saturation profiles are impressively close to the true profile.

Figure 7 shows the uncertainty estimates associated with the predicted CO_2 saturation models in terms of the standard deviation. Both uncertainty quantification methods generate standard deviation maps that are comparable to each other though the uncertainties using the bagging method are slightly higher than those produced by the MC dropout method. Thus, at least for case of noise free input data, the MC dropout method is a computationally efficient alternative to the bagging. The non-zero standard deviation values are found inside and around the recovered CO_2 plume rather than randomly scattered on the cross section which is a product of the neural networks being trained with a set of reservoir models that have CO_2 concentrations only within a given reservoir zone. The standard deviation values are roughly bounded between 0.04 and 0.08, which is approximately 10-15% of the estimated CO_2 saturation.

To evaluate the effects of data uncertainty due to noise on the DL inversion methods, we perform extra training and prediction experiments. First, the seismic test data sets at 20 years after injection start are contaminated with Gaussian noise with zero mean and standard deviation of 10%. This level of assumed noise is commonly found in both traditional and DL inversion work (Commer and Newman, 2008; Yang and Ma, 2018; Puzyrev, 2019; Colombo et al., 2020; Um et al., 2020). The noisy test data are then input into the networks that are trained using the noise-free training data sets. **Figures** 8 and 9 display the resulting predicted mean CO_2 saturation distributions which are comparable to those predicted in the absence of noise (**Figures** 5 and 6), demonstrating the robustness of these deep-learning imaging techniques. The corresponding standard deviation maps in **Figure** 10 are somewhat different from the corresponding noise-free examples shown in **Figure** 7. For example, the distribution of non-zero standard deviation values has a slightly higher magnitude and greater lateral spread. In general, the modeling results show that the presence of noise in the data set does not necessarily meaningfully deteriorate the overall prediction performance of the network.

Conclusions and Discussion

In this paper, we have presented a DL-based FWI method for monitoring GCS in real time. Our DL inversion technique is based on the U-Net architecture and here has been trained on pairs of synthetic seismic data and subsurface CO_2 saturation models. Accordingly, the DL inversion is capable of directly imaging CO_2 saturation. Using the MC dropout and bagging methods, the DL imaging quantifies uncertainties associated with CO_2 saturation images in terms of standard deviation. Our numerical modeling experiments demonstrate that the DL inversion can provide realistic estimates of CO_2 model fairly well even when the data sets are contaminated with random noise. For the given problem, the two uncertainty quantification results are comparable to each other. Our DL inversion experiments indicates a possibility that the proposed DL inversion method can be used to continuously monitor CO_2 saturation and provide estimates of uncertainty in near-real time when seismic sensors are permanently installed at a GCS site and their data are continuously fed into the DL inversion.

The 2D examples presented here serve merely as a proof of concept of the DL imaging technique. These results have been generated using vertical profiles of hydrologic properties extracted from 3D volumes of saturation at 35 different simulation times after the start of injection for one realization of the Kimberlina 1.2 model. To provide the required number of training data sets for full 3D imaging as we envision it will require employing multiple 3D reservoir simulations where the underlying reservoir models are different-statistical realizations of a base reservoir model. Another level of simplicity present in this study is that we have assumed that we exactly know the rock/petrophysical transformations between the flow-model parameters and the geophysical property models that are used to generate the synthetic training data. In reality these transformations are not exactly known and can be stochastic in nature requiring the analysis of rock physics measurements made on core samples/and or petrophysical analysis of well log data to estimate. This will introduce increased levels of uncertainty that we will be exploring in future studies. In terms of moving this imaging technology to an operating field site we also recognize that real geophysical field data always contain levels of complication and sources of noise that are not easily duplicated in synthetic data sets such as those we have employed here. Our future plans involve applying this imaging technology on field data to test its robustness in real-world situations.

Acknowledgements

This work was completed as part of the Science-informed Machine learning to Accelerate Real Time decision making for Carbon Storage (SMART-CS) Initiative (edx.netl.doe.gov/SMART). Support for this initiative was provided by the U.S. Department of Energy's (DOE) Office of Fossil Energy's Carbon Storage Research program through the National Energy Technology Laboratory (NETL). The authors wish to acknowledge Mark McKoy (NETL, Carbon Storage Technology Manager), Darin Damiani (DOE Office of Fossil Energy, Carbon Storage Program Manager), and Mark Ackiewicz (DOE Office of Fossil Energy, Director, Division of Carbon Capture and Storage Research and Development), for programmatic guidance, direction and support. The numerical modeling experiments presented in this work were carried out using high performance computing facilities provided by the Geosciences Measurement Facility (GMF) at Lawrence Berkeley National Laboratory and NETL. The authors would like to thank Dr. Yuanyuan Li and two anonymous reviewers and editor Drs. Waheed and Moser for their constructive feedback.

1

Figure Captions

Figure 1. The U-Net architecture used for the DL inversion

Figure 2. Cross-sectional (y-z plane) views of the Kimberlina CO_2 saturation models (left column) and their corresponding P-wave velocity models (right column).

Figure 3. Examples of common shot gathers with a 2D velocity model.

Figure 4. Training and validation loss comparison for Training the U-Net. (a) The MC dropout method. (b) The bagging method. (c) The plain single U-Net method.

Figure 5. Comparison of the Kimberlina CO_2 saturation model (the 1st column) and the mean values of CO_2 as recovered from the three different DL methodologies (the 2nd to 4th columns). The networks are trained and tested on noise-free seismic data. The recovered CO_2 saturation values are compared to the true values along the black broken lines as shown in **Figure** 6.

Figure 6. 1D CO_2 saturation profile comparison of the four cases along the three black broken lines as shown in **Figure** 5e-h.

Figure 7. Cross-sectional views of standard deviation of CO_2 saturation from the MC dropout method (the left column) and the bagging method (the right column).

Figure 8. Comparison of the Kimberlina CO_2 saturation model (the 1st column) and the mean values of CO_2 as recovered from the three different DL methodologies (the 2nd to 4th columns). The networks are trained on the noise-free seismic data, and the test data are contaminated with 10% Gaussian noise. The recovered CO_2 saturation values are compared to the true values along the black broken lines as shown in **Figure** 9.

Figure 9. 1D CO_2 saturation profile comparison of the four cases along the three black broken lines as shown in **Figure** 8e-h.

Figure 10. Cross-sectional views of standard deviation of CO_2 saturation from the MC dropout method (the left column) and the bagging method (the right column). The networks are trained on the noise-free seismic data, and the test data are contaminated with 10% Gaussian noise.

References

Alumbaugh, D., Commer, M., Crandall, D., Gasperikova, E., Harbert, W., Li, Y., Lin, Y., Samarasinghe, S., and Yang, X. (2021) Development of a Multi-Scale Synthetic Data Set for the Testing of Subsurface CO_2 Storage Monitoring Strategies. AGU Meeting.

Ajo-Franklin, J.B., Peterson, J., Doetsch, J. and Daley, T.M. (2013) High-resolution characterization of a CO_2 plume using crosswell seismic tomography: Cranfield, MS, USA. *International Journal of Greenhouse Gas Control*, *18*, 497-509.

Alford, R.M., Kelly, K.R. and Boore, D.M. (1974) Accuracy of finite-difference modeling of the acoustic wave equation. *Geophysics*, *39*(6), 834-842.

Araya-Polo, M., Jennings, J., Adler, A. and Dahlke, T. (2018) Deep-learning tomography. *The Leading Edge*, *37*(1), 58-66.

Arogunmati, A. and Harris, J.M. (2012) An effective crosswell seismic traveltime-estimation approach for quasi-continuous reservoir monitoring. *Geophysics*, *77*(2), M17-M26.

Arts, R., Eiken, O., Chadwick, A., Zweigel, P., Van der Meer, L. and Zinszner, B. (2003) Monitoring of CO_2 injected at Sleipner using time lapse seismic data. *Greenhouse Gas Control Technologies-6th International Conference*, 347-352.

Aster, R.C., Borchers, B. and Thurber, C.H. (2018) Parameter estimation and inverse problems. Elsevier.

Benson, S.M. and Cole, D.R. (2008) CO_2 sequestration in deep sedimentary formations. *Elements*, *4*(5), 325-331.

Chadwick, A., Williams, G., Delepine, N., Clochard, V., Labat, K., Sturton, S., Buddensiek, M.L., Dillen, M., Nickel, M., Lima, A.L. and Arts, R. (2010) Quantitative analysis of time-lapse seismic monitoring data at the Sleipner CO 2 storage operation. *The Leading Edge*, *29*(2), 170-177.

Chollet, F. (2017) *Deep learning with Python*. Simon and Schuster.

Chollet, F. (2018) Keras: The python deep learning library. Astrophysics Source Code Library.

Colombo, D., Li, W., Sandoval-Curiel, E. and McNeice, G.W. (2020) Deep-learning electromagnetic monitoring coupled to fluid flow simulators. *Geophysics*, *85*(4), WA1-WA12.

Commer, M. and Newman, G.A. (2008) New advances in three-dimensional controlled-source electromagnetic inversion. *Geophysical Journal International*, *172*(2), 513-535.

Davis, T.L., Landrø, M. and Wilson, M. eds. (2019) Geophysics and Geosequestration. Cambridge University Press.

Dupuy, B., Romdhane, A., Eliasson, P., Querendez, E., Yan, H., Torres, V.A. and Ghaderi, A. (2017) Quantitative seismic characterization of CO_2 at the Sleipner storage site. North Sea. Interpretation, 5(4), SS23-SS42.

Egorov, A., Pevzner, R., Bóna, A., Glubokovskikh, S., Puzyrev, V., Tertyshnikov, K. and Gurevich, B. (2017) Time-lapse full waveform inversion of vertical seismic profile data: Workflow and application to the CO2CRC Otway project. *Geophysical Research Letters*, *44*(14), 7211-7218.

Friedman, J., Hastie, T. and Tibshirani, R. (2001) *The elements of statistical learning*. Springer.

Gal, Y. and Ghahramani, Z. (2016) Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*. PMLR, 1050-1059.

Goodfellow, I., Bengio, Y. and Courville, A. (2016) *Deep learning*. MIT press.

Grana, D., Liu, M. and Ayani, M. (2020) Prediction of CO_2 Saturation Spatial Distribution Using Geostatistical Inversion of Time-Lapse Geophysical Data. IEEE Transactions on Geoscience and Remote Sensing, 59(5), 3846-3856.

Grana, D., Mukerji, T. and Doyen, P. (2021) Seismic Reservoir Modeling: Theory, Examples, and Algorithms. John Wiley & Sons.

Hunziker, J., Laloy, E. and Linde, N. (2019) Bayesian full-waveform tomography with application to crosshole ground penetrating radar data. *Geophysical Journal International*, *218*(2), 913-931.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2021) An introduction to statistical learning. Springer.

Kaur, H., Zhong, Z., Sun, A. and Fomel, S. (2021) Time-lapse seismic data inversion for estimating reservoir parameters using deep learning. Interpretation, 10(1), 1-91.

Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S. and Yang, L. (2021) Physics-informed machine learning. Nature Reviews Physics, 3(6), 422-440.

Kazei, V., Ovcharenko, O., Plotnitskii, P., Peter, D., Zhang, X. and Alkhalifah, T. (2021) Mapping full seismic waveforms to vertical velocity profiles by deep learning. *Geophysics*, *86*(5), 1-50.

Kingma, D.P. and Ba, J. (2015) A method for stochastic optimization. In: 3rd International Conference on Learning Representations. Conference Track Proceedings.

Lazaratos, S.K. and Marion, B.P. (1997) Crosswell seismic imaging of reservoir changes caused by CO_2 injection. *The Leading Edge*, *16*(9), 1300-1308.

Li, D., Peng, S., Huang, X., Guo, Y., Lu, Y. and Cui, X. (2021) Time-lapse full waveform inversion based on curvelet transform: Case study of CO_2 storage monitoring. International Journal of Greenhouse Gas Control, 110, 103417.

Li, D., Peng, S., Guo, Y., Lu, Y. and Cui, X. (2021) CO_2 storage monitoring based on time-lapse seismic data via deep learning. *International Journal of Greenhouse Gas Control*, *108*, 103336.

Li, Y. and Yang, D. (2021) Electrical imaging of hydraulic fracturing fluid using steel-cased wells and a deep-learning method. *Geophysics*, *86*(4), E315-E332.

Lumley, D. (2010) 4D seismic monitoring of CO_2 sequestration. *The Leading Edge*, *29*(2), 150-155.

Mavko, G., Mukerji, T. and Dvorkin, J. (2020) The rock physics handbook. Cambridge university press.

Metz, B., Davidson, O., De Coninck, H.C., Loos, M. and Meyer, L. (2005) IPCC special report on carbon dioxide capture and storage. Cambridge University Press.

Mosser, L., Dubrule, O. and Blunt, M.J. (2020) Stochastic seismic waveform inversion using generative adversarial networks as a geological prior. Mathematical Geosciences, 52(1), 53-79.

Oh, S., Noh, K., Seol, S.J. and Byun, J. (2020) Cooperative deep learning inversion of controlled-source electromagnetic data for salt delineation. *Geophysics*, *85*(4), E121-E137.

Petrov, P.V. and Newman, G.A. (2012) 3D finite-difference modeling of elastic wave propagation in the Laplace-Fourier domain. *Geophysics*, *77*(4), T137-T155.

Pruess, K., Oldenburg, C.M. and Moridis, G.J. (1999) *TOUGH2 user's guide version 2* (No. LBNL-43134). Lawrence Berkeley National Lab.

Puzyrev, V. (2019) Deep learning electromagnetic inversion with convolutional neural networks. *Geophysical Journal International*, *218*(2), 817-832.

Queißer, M. and Singh, S.C. (2013) Full waveform inversion in the time lapse mode applied to CO_2 storage at Sleipner. *Geophysical prospecting*, *61*(3), 537-555.

Raissi, M., Perdikaris, P. and Karniadakis, G.E. (2019) Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational physics, 378, 686-707.

Raknes, E.B., Weibull, W. and Arntsen, B. (2015) Seismic imaging of the carbon dioxide gas cloud at Sleipner using 3D elastic time-lapse full waveform inversion. International Journal of Greenhouse Gas Control(42), 26-45.

Ringrose, P. (2020) How to Store CO_2 Underground: insights from early-mover CCS Projects. Springer.

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

Rutqvist, J. (2012) The geomechanics of CO_2 storage in deep sedimentary formations. *Geotechnical and Geological Engineering*, *30*(3), 525-551.

Sen, M.K. and Stoffa, P.L. (2013) Global optimization methods in geophysical inversion. Cambridge University Press.

Shin, C. and Cha, Y.H. (2008) Waveform inversion in the Laplace domain. *Geophysical Journal International*, *173*(3), 922-931.

Shin, C. and Ho Cha, Y. (2009) Waveform inversion in the Laplace—Fourier domain. *Geophysical Journal International*, *177*(3), 1067-1079.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, *15*(1), pp.1929-1958.

Tang, H., Fu, P., Sherman, C.S., Zhang, J., Ju, X., Hamon, F., Azzolina, N.A., Burton-Kelly, M. and Morris, J.P. (2021) A deep learning-accelerated data assimilation and forecasting workflow for commercial-scale geologic carbon storage. International Journal of Greenhouse Gas Control, 112, 103488.

Tarantola, A. (2005) Inverse problem theory and methods for model parameter estimation. Society for Industrial and Applied Mathematics.

Um, E.S., Commer, M. and Newman, G.A. (2012) Iterative finite-difference solution analysis of acoustic wave equation in the Laplace-Fourier domain. *Geophysics*, *77*(2), T29-T36.

Um, E.S., Kim, J. and Wilt, M. (2020) 3D borehole-to-surface and surface electromagnetic modeling and inversion in the presence of steel infrastructure. *Geophysics*, *85*(5), E139-E152.

Vasco, D.W., Daley, T.M. and Bakulin, A. (2014) Utilizing the onset of time-lapse changes: A robust basis for reservoir monitoring and characterization. *Geophysical Journal International*, *197*(1), 542-556.

Virieux, J. and Operto, S. (2009) An overview of full-waveform inversion in exploration geophysics. *Geophysics*, *74*(6), WCC1-WCC26.

Virieux, J., Asnaashari, A., Brossier, R., Métivier, L., Ribodetti, A. and Zhou (2017) An introduction to full waveform inversion. *Encyclopedia of exploration geophysics,* R1-1.

Waheed, U., Haghighat, E., Alkhalifah, T., Song, C. and Hao, Q. (2021) PINNeik: Eikonal solution using physics-informed neural networks. Computers & Geosciences, 155, 104833.

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

Yang, F. and Ma, J. (2019) Deep-learning inversion: A next-generation seismic velocity model building method. *Geophysics*, *84*(4), R583-R599.

Yang, X., Buscheck, T.A., Mansoor, K., Wang, Z., Gao, K., Huang, L., Appriou, D. and Carroll, S.A. (2019) Assessment of geophysical monitoring methods for detection of brine and CO_2 leakage in drinking water aquifers. *International Journal of Greenhouse Gas Control*, *90*, 102803.

Zhang, F., Juhlin, C., Cosma, C., Tryggvason, A. and Pratt, R.G. (2012) Cross-well seismic waveform tomography for monitoring CO_2 injection: a case study from the Ketzin Site, Germany. *Geophysical Journal International*, *189*(1), 629-646.

Zhang, Z.D. and Alkhalifah, T. (2019) Regularized elastic full-waveform inversion using deep learning. Geophysics, 84(5), R741-R751.

Zhang, Z. and Lin, Y. (2020) Data-driven seismic waveform inversion: A study on the robustness and generalization. *IEEE Transactions on Geoscience and Remote sensing*, *58*(10), 6900-6913.

Zhao, Z. and Sen, M.K. (2021) A gradient-based Markov chain Monte Carlo method for full-waveform inversion and uncertainty analysis. *Geophysics*, *86*(1), R15-R30.

Zhu, W., Xu, K., Darve, E., Biondi, B. and Beroza, G.C. (2022) Integrating deep neural networks with full-waveform inversion: Reparameterization, regularization, and uncertainty quantification. Geophysics, 87(1), R93-R109.

Zhou, Q. and Birkholzer, J.T. (2011) On scale and magnitude of pressure build-up induced by large-scale geologic storage of CO_2. *Greenhouse Gases: Science and Technology*, *1*(1), 11-20.

Zhou, Z., Lin, Y., Zhang, Z., Wu, Y., Wang, Z., Dilmore, R. and Guthrie, G. (2019) A data-driven CO_2 leakage detection using seismic data and spatial–temporal densely connected convolutional neural networks. *International Journal of Greenhouse Gas Control*, *90*, 102790.

Figures

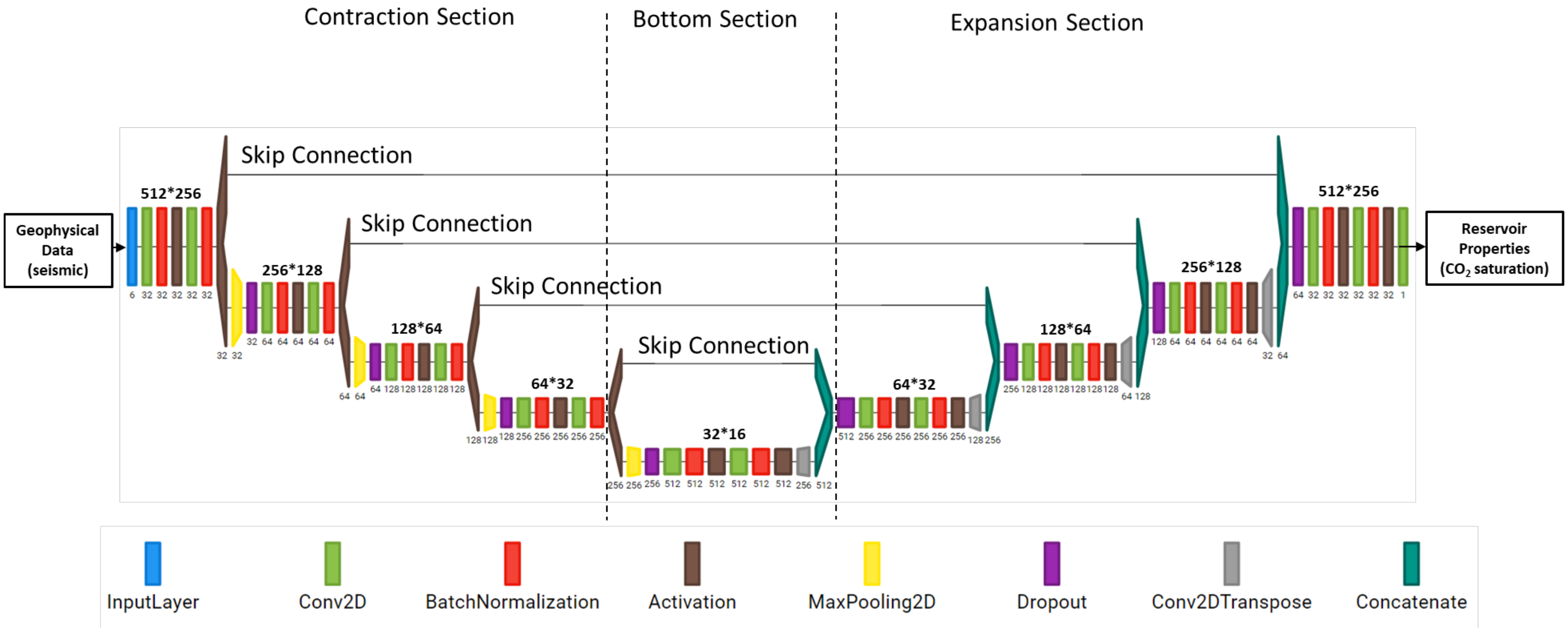


Figure 1. The U-Net architecture used for the DL inversion. The x-y size of the data/feature maps is shown at the top of the layers. The number of channels of data/feature maps is shown at the bottom of the layers.

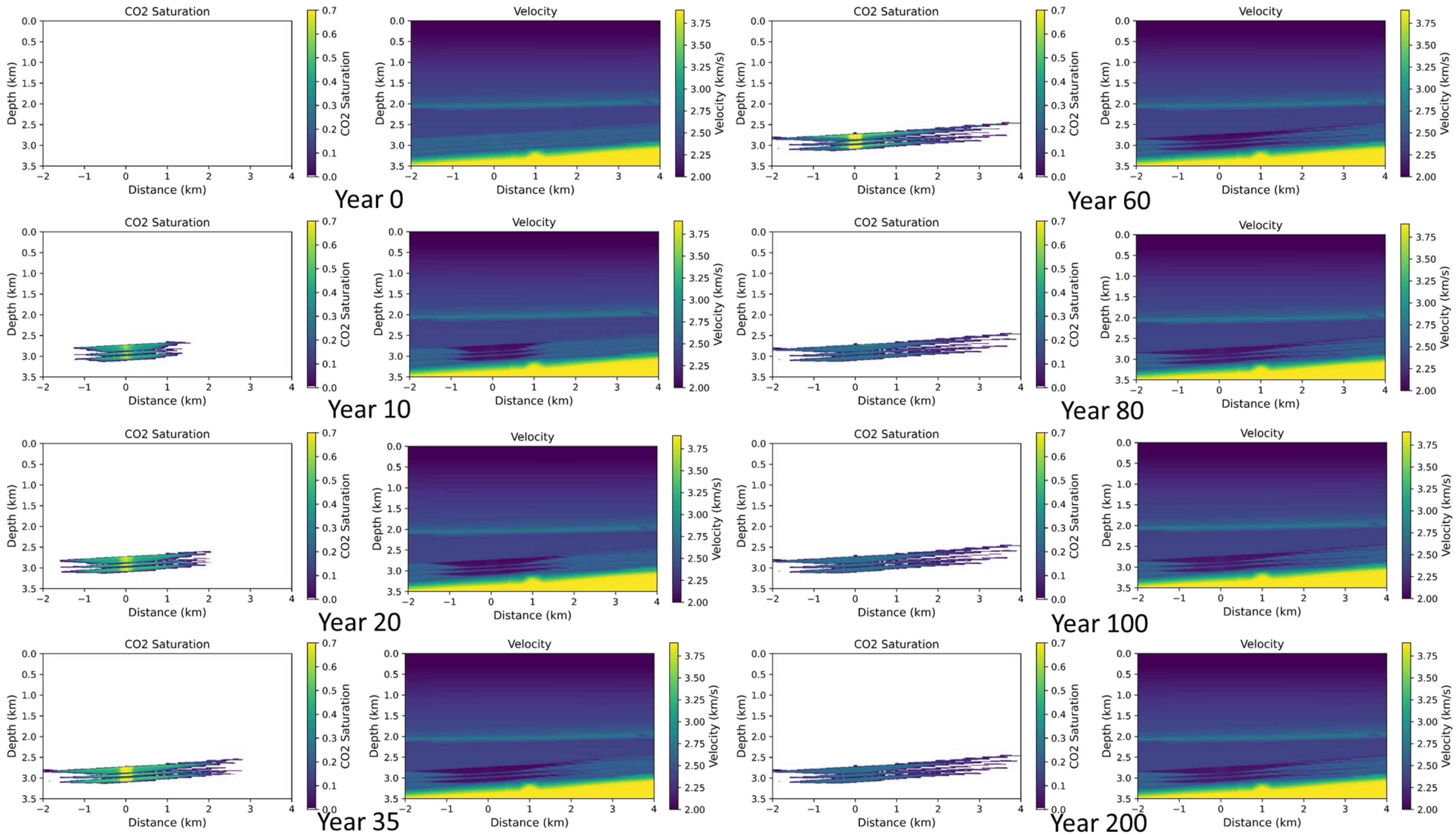


Figure 2. Cross-sectional (*y*-*z* plane) views of the Kimberlina CO_2 saturation models (left column) and their corresponding P-wave velocity models (right column).

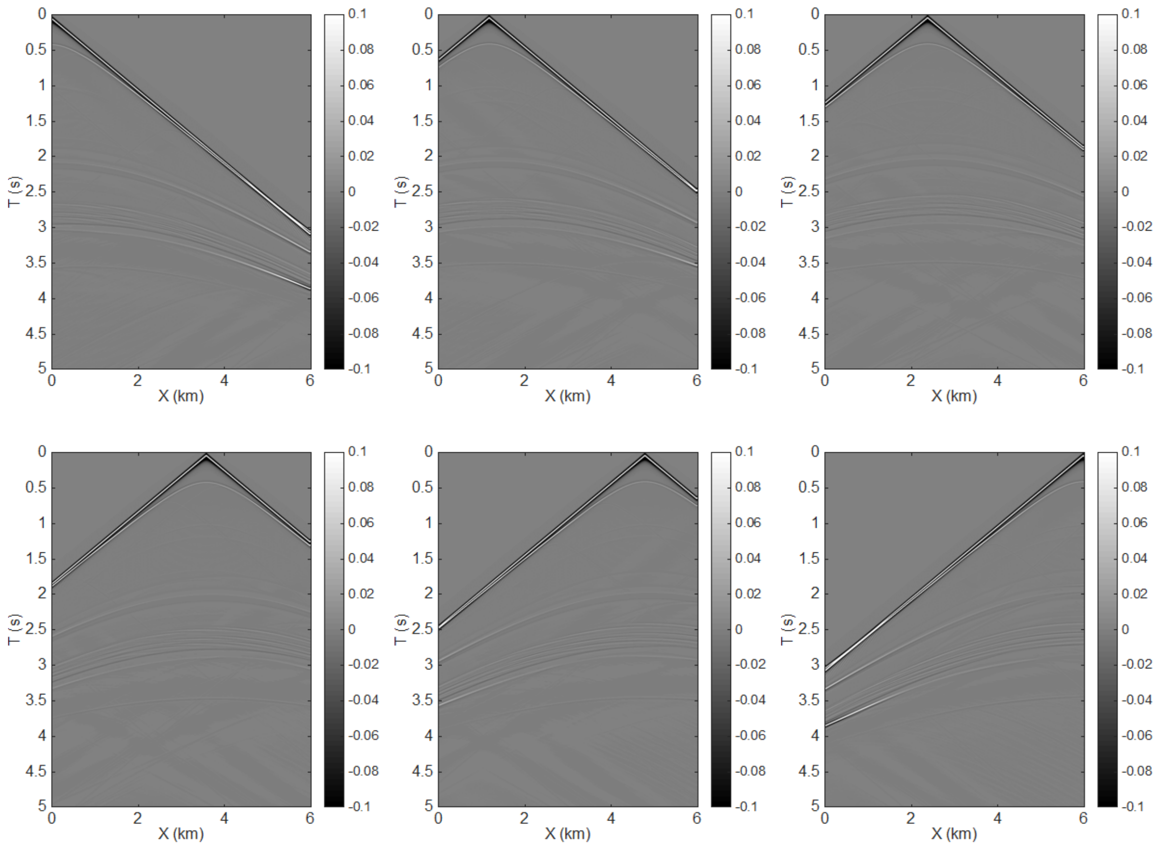
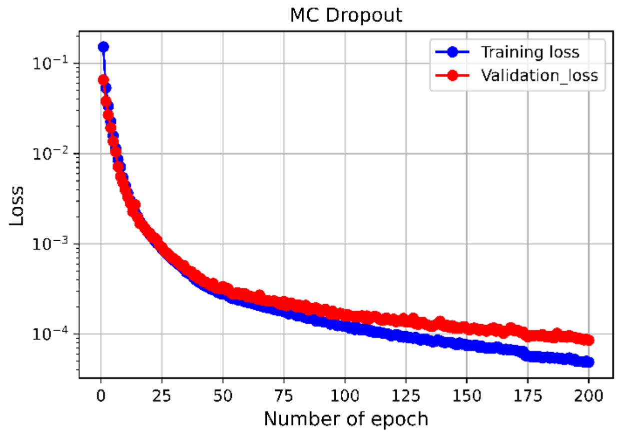
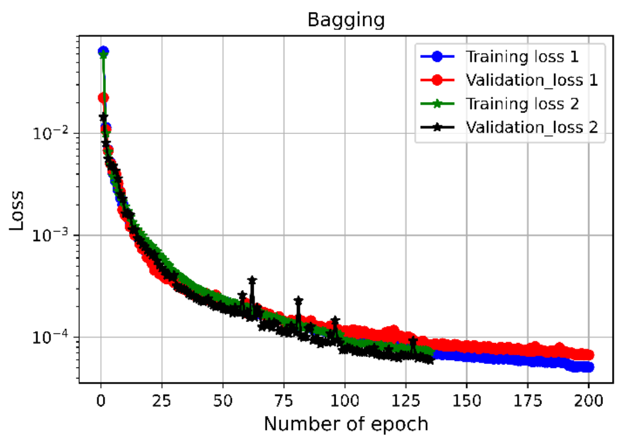


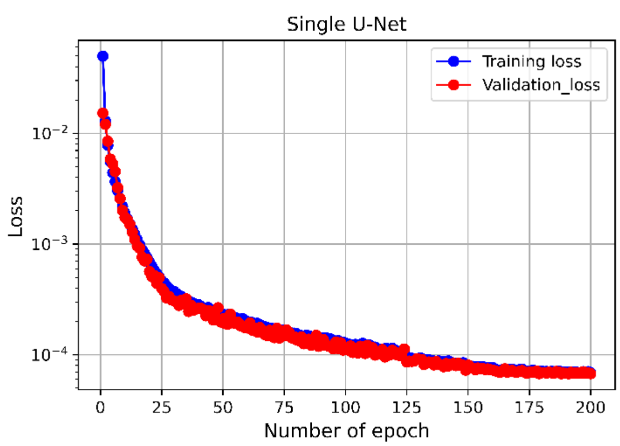
Figure 3. Examples of common shot gathers with a 2D velocity model.



(a)

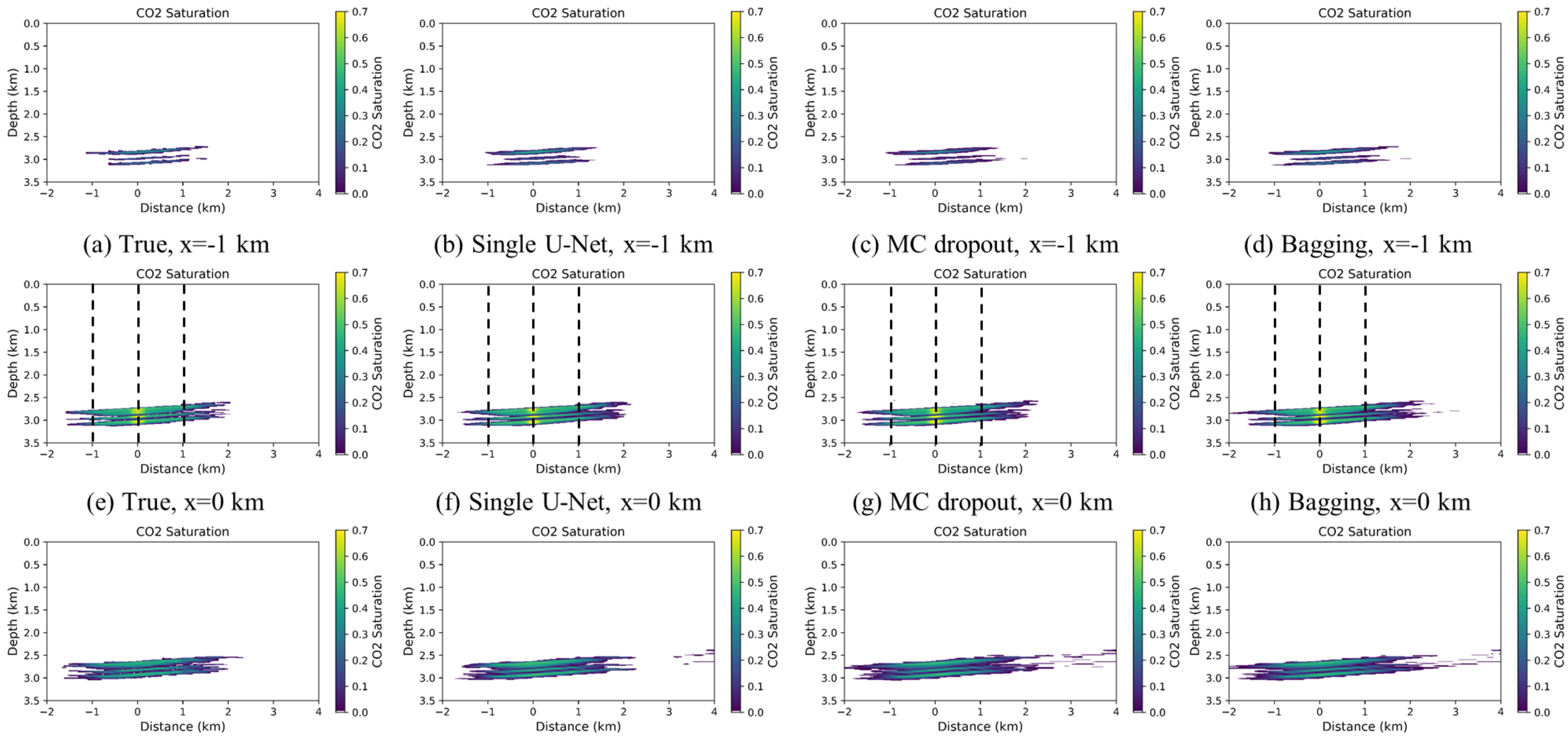


(b)



(c)

Figure 4. Training and validation loss comparison for Training the U-Net. (a) The MC dropout method. (b) The bagging method. (c) The plain single U-Net method.

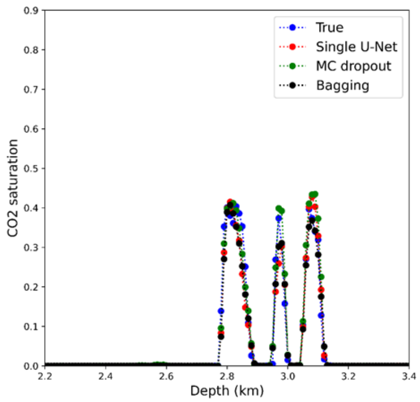


(a) True, x=-1 km (b) Single U-Net, x=-1 km (c) MC dropout, x=-1 km (d) Bagging, x=-1 km

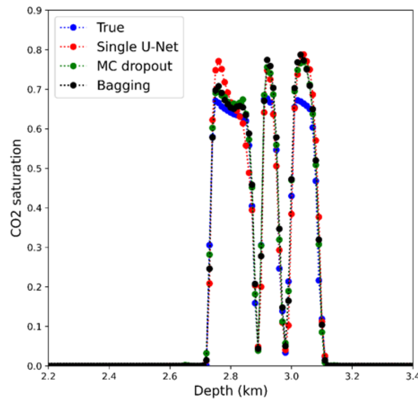
(e) True, x=0 km (f) Single U-Net, x=0 km (g) MC dropout, x=0 km (h) Bagging, x=0 km

(i) True, x=1 km (j) Single U-Net, x=1 km (k) MC dropout, x=1 km (l) Bagging, x=1 km

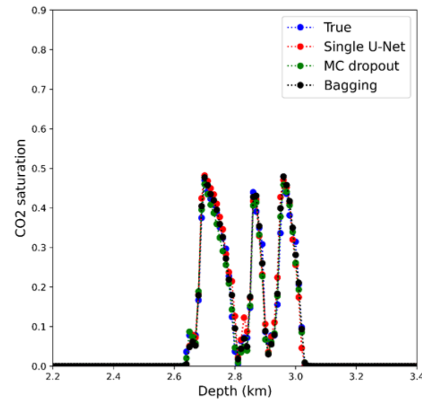
Figure 5. Comparison of the Kimberlina CO_2 saturation model (the 1st column) and the mean values of CO_2 as recovered from the three different DL methodologies (the 2nd to 4th columns). The networks are trained and tested on noise-free seismic data. The recovered CO_2 saturation values are compared to the true values along the black broken lines as shown in Figure 6.



(a) y=-1 km



(b) y=0 km



(c) y=1 km

Figure 6. 1D CO_2 saturation profile comparison of the four cases along the three black broken lines as shown in **Figure** 5e-h.

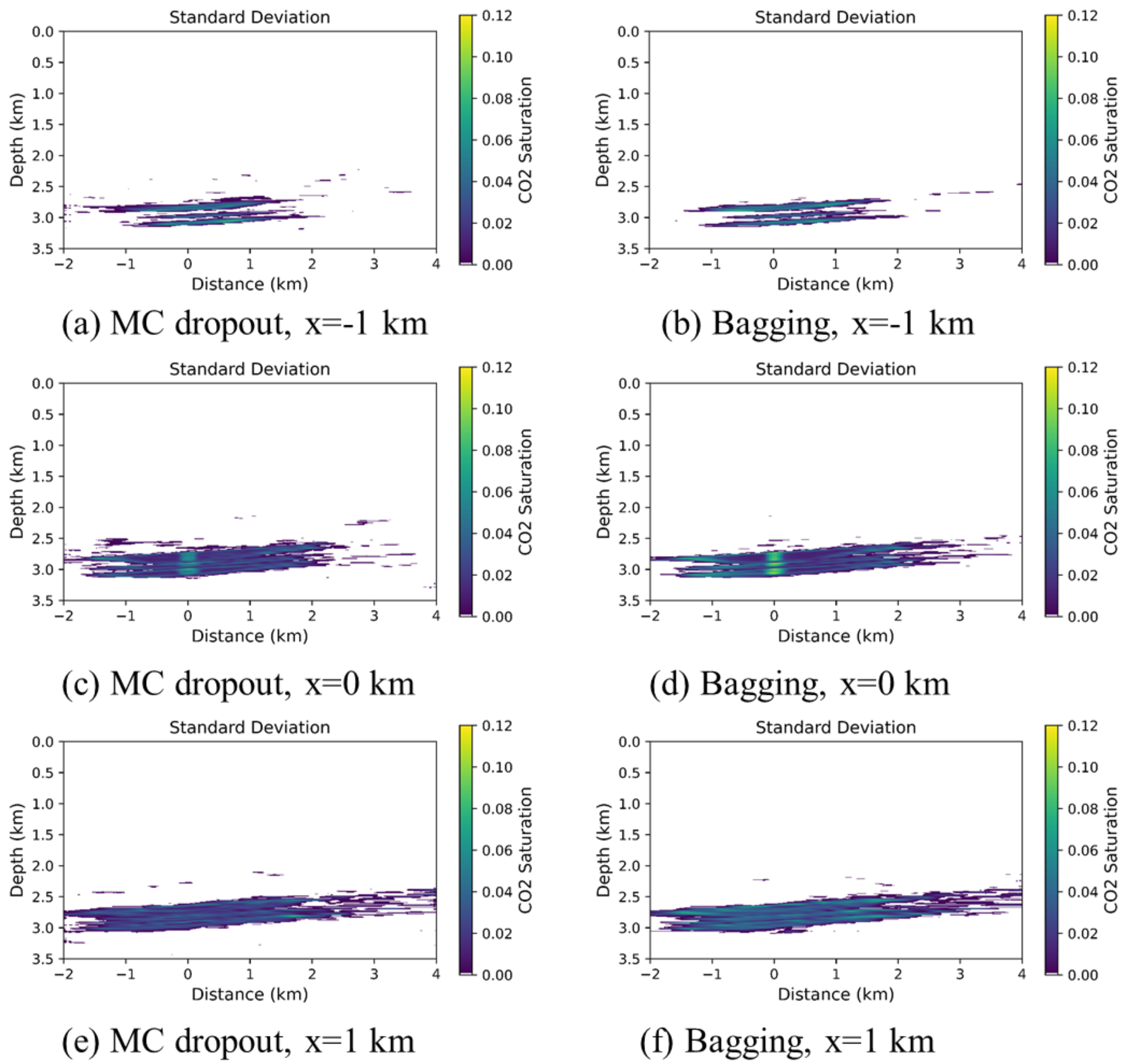
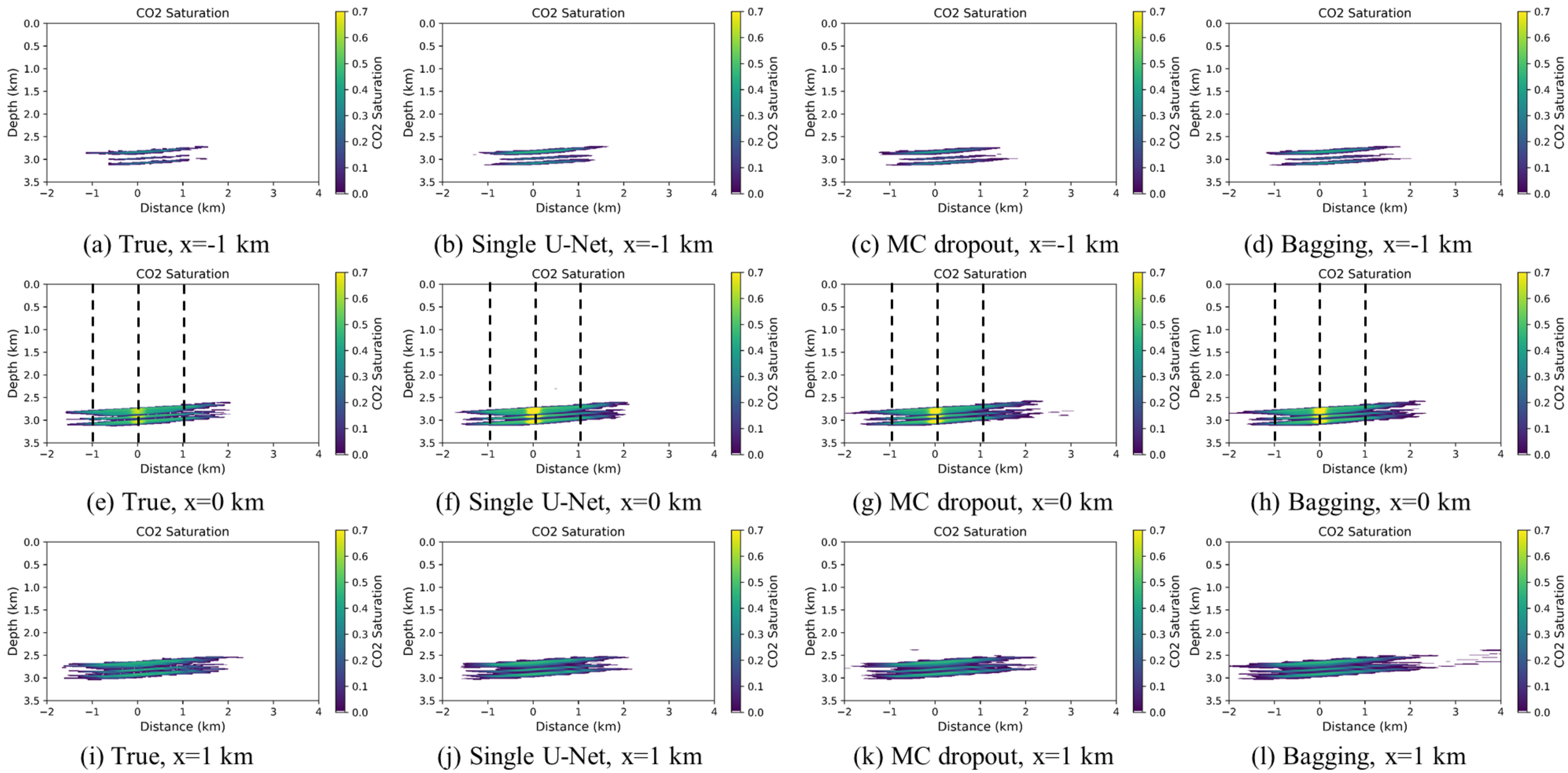


Figure 7. Cross-sectional views of standard deviation of CO_2 saturation from the MC dropout method (the left column) and the bagging method (the right column).

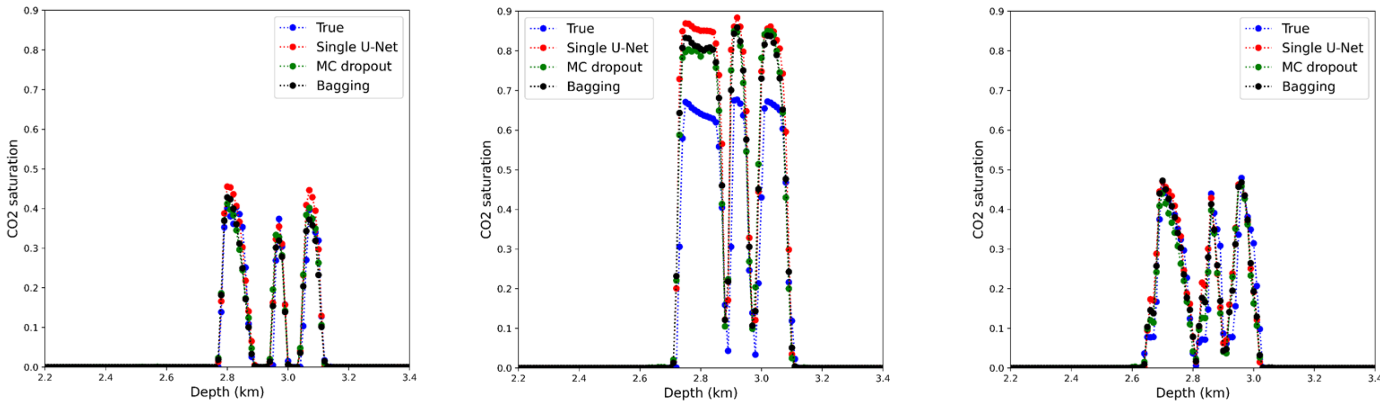


(a) True, x=-1 km (b) Single U-Net, x=-1 km (c) MC dropout, x=-1 km (d) Bagging, x=-1 km

(e) True, x=0 km (f) Single U-Net, x=0 km (g) MC dropout, x=0 km (h) Bagging, x=0 km

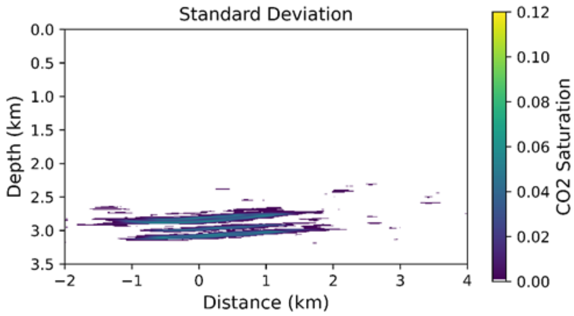
(i) True, x=1 km (j) Single U-Net, x=1 km (k) MC dropout, x=1 km (l) Bagging, x=1 km

Figure 8. Comparison of the Kimberlina CO_2 saturation model (the 1st column) and the mean values of CO2 as recovered from the three different DL methodologies (the 2nd to 4th columns). The networks are trained on the noise-free seismic data, and the test data are contaminated with 10% Gaussian noise. The recovered CO_2 saturation values are compared to the true values along the black broken lines as shown in **Figure** 9.

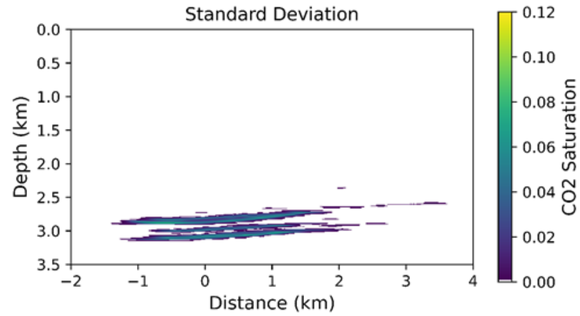


(a) y=-1 km (b) y=0 km (c) y=1 km

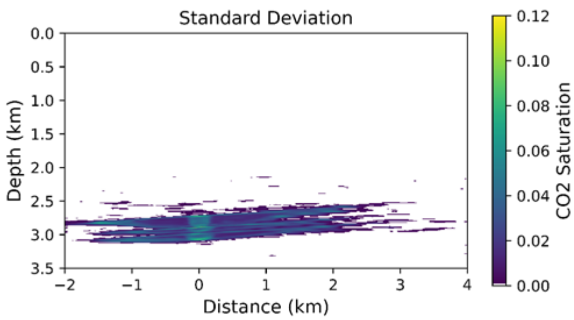
Figure 9. 1D CO_2 saturation profile comparison of the four cases along the three black broken lines as shown in Figure 8e-h.



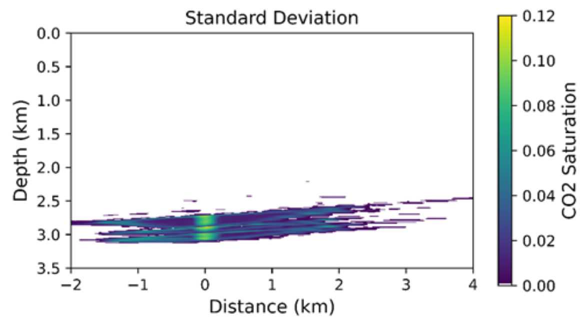
(a) MC dropout, x=-1 km



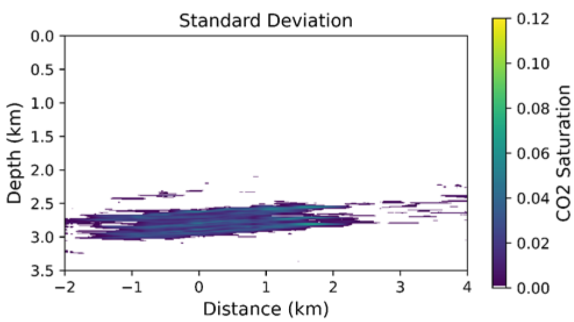
(b) Bagging, x=-1 km



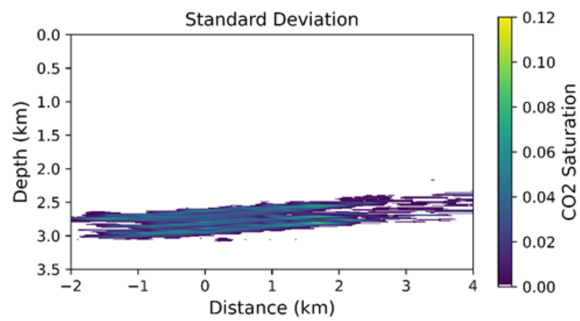
(c) MC dropout, x=0 km



(d) Bagging, x=0 km



(e) MC dropout, x=1 km



(f) Bagging, x=1 km

Figure 10. Cross-sectional views of standard deviation of CO_2 saturation from the MC dropout method (the left column) and the bagging method (the right column). The networks are trained on the noise-free seismic data, and the test data are contaminated with 10% Gaussian noise.