

1           **Neural network-based CO<sub>2</sub> interpretation from 4D**  
2           **Sleipner seismic images**

3           **Bei Li<sup>1</sup>and Yunyue Elita Li<sup>1</sup>**

4           <sup>1</sup>Department of Civil and Environmental Engineering, National University of Singapore, 117576 Singapore

5           **Key Points:**

- 6           • We train a 3D U-Net for an automatic end-to-end mapping from 4D seismic im-  
7           ages to 3D CO<sub>2</sub> distribution.
- 8           • Successful applications of the trained neural network demonstrate its robustness  
9           and consistency.
- 10          • We analyze the neural network interpretation standards and provide training strat-  
11          egy for 2D sparse labels.

---

Corresponding author: Yunyue Elita Li, [elita.li@nus.edu.sg](mailto:elita.li@nus.edu.sg)

12 **Abstract**

13 Time-lapse or 4D seismic survey is a crucial monitoring tool for CO<sub>2</sub> geological sequestration.  
 14 Conventional time-lapse interpretation provides detailed characterization of CO<sub>2</sub>  
 15 distribution in the storage unit. However, manual interpretation is labour-intensive and  
 16 often inconsistent throughout the long monitoring history, due to the inevitable changes  
 17 in seismic acquisition and processing technology and interpreter's subjectivity. We pro-  
 18 pose a neural network (NN)-based interpretation method that translates baseline and  
 19 monitoring seismic images to the probability of CO<sub>2</sub> presence. We use a simplified 3D  
 20 U-Net, whose training, validation and testing are all based on the Sleipner CO<sub>2</sub> storage  
 21 project. The limited labels for training are derived from the interpreted CO<sub>2</sub> plume out-  
 22 lines within the internal sandstone layers for 2010. Then we apply the trained NN on  
 23 different time-lapse seismic datasets from 1999 to 2010. The results suggest that our NN-  
 24 based CO<sub>2</sub> interpretation has the following advantages: (1) high interpretation efficiency  
 25 by automatic end-to-end mapping; (2) robustness against the processing-induced mis-  
 26 match between the baseline and time-lapse inputs, relaxing the baseline reprocessing de-  
 27 mands when compared to newly acquired or reprocessed time-lapse datasets; and (3) in-  
 28 herent interpretation consistency throughout multiple vintage datasets. Testing results  
 29 with crafted time-lapse images unveil that the NN takes both amplitude difference and  
 30 structural similarity into account for CO<sub>2</sub> interpretation. We also compare 2D and 3D  
 31 U-Nets under the scenario of sparse 2D labels for training. The results suggest that the  
 32 3D U-Net provides more continuous interpretation at the cost of larger computational  
 33 resources for training and application.

34 **Plain Language Summary**

35 To reduce the greenhouse effect, captured and compressed CO<sub>2</sub> can be injected into  
 36 subsurface area with special geological settings that accommodate the greenhouse gas.  
 37 Many pilot projects for such geological CO<sub>2</sub> sequestration have been ran for decades, dur-  
 38 ing which various monitoring techniques have been experimented to study the interac-  
 39 tions between injected CO<sub>2</sub> and the storage unit mainly for the purpose of evaluating  
 40 storage safety. Such studies suggest that time-lapse 3D seismic survey is the key tool for  
 41 detailed understanding of CO<sub>2</sub> behavior along time. However, the recorded seismic data  
 42 must go through rigorous processing route to match the baseline and time-lapse datasets,  
 43 so that the differences can be detected and interpreted by human labours. In this study,  
 44 we propose to use the neural network (NN) to facilitate the human interpretation. It of-  
 45 fers an efficient end-to-end mapping directly from the 4D seismic images to 3D CO<sub>2</sub> dis-  
 46 tribution. By incorporating differently processed data during training, the NN gains ro-  
 47 bustness against moderate mismatch between the baseline and time-lapse images. The  
 48 generalized applications of the trained NN on different time-lapse data show great con-  
 49 sistency throughout the monitoring history, which provides reliable analysis for CO<sub>2</sub> plume  
 50 development as a function of time.

51 **1 Introduction**

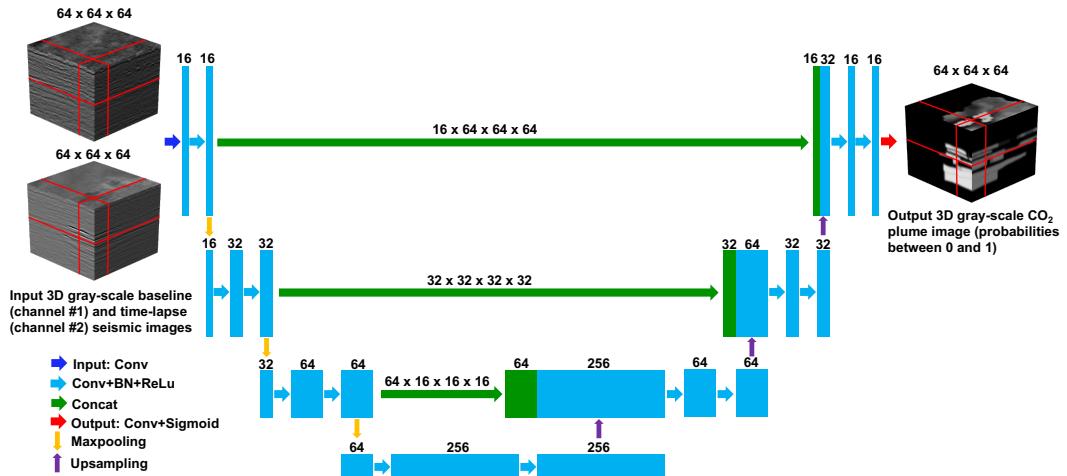
52 Excessive emission of greenhouse gases due to anthropogenic activities has signif-  
 53 icantly contributed to climate change since the industrial revolution (Bachu & Adams,  
 54 2003). With the increase of fossil fuel consumption, CO<sub>2</sub> is responsible for more than  
 55 64% of the enhanced greenhouse effect (Bryant et al., 1997). Hence, CO<sub>2</sub> capture and  
 56 storage (CCS) becomes an important measure for greenhouse gas mitigation, among which  
 57 Geological CO<sub>2</sub> Sequestration (GCS) aims at removing CO<sub>2</sub> from the atmosphere by keep-  
 58 ing them underground with suitable geomechanical conditions (Castelletto et al., 2013).  
 59 During GCS projects, long-term monitoring is necessary for the purposes of understand-  
 60 ing CO<sub>2</sub> behaviour in the reservoir, detecting CO<sub>2</sub> leakage from the storage unit, and  
 61 assessing effects of contingency measures in case of leakage (Furre et al., 2017). Serving

for various monitoring purposes, time-lapse or 4D seismic survey is the most informative tool for detailed and quantitative CO<sub>2</sub> characterization in the storage complex varying along time (Boait et al., 2012; Bourne et al., 2014).

For GCS monitoring, time-lapse seismic analysis estimates the subsurface parameter changes between two separate seismic experiments due to CO<sub>2</sub> injection. The density and bulk modulus differences between the injected supercritical CO<sub>2</sub> and the originally saturated brine leads to dramatic velocity decrease in the storage reservoir, and consequently creates strong reflections at interfaces between the CO<sub>2</sub> and brine saturated rocks (R. Chadwick et al., 2005). Hence, conventional quantitative interpretations for CO<sub>2</sub> plume thickness and saturation are mainly based on velocity pushdown and reflection amplitude (A. Chadwick et al., 2010). However, detailed analysis still requires tedious manual interpretation and the outcome heavily depends on reliable processing with relative amplitude preservation and satisfying match between the baseline and time-lapse seismic images. Furthermore, it is difficult to maintain the interpretation consistent throughout the long-term GCS project. To partially automate this process, stratigraphic inversion (Clochard et al., 2010) and full waveform inversion (FWI) (Romdhane & Querendez, 2014) have been utilized to build high-resolution models of elastic impedance or velocity, based on which the CO<sub>2</sub> plume can be interpreted more accurately and objectively. Nonetheless, the inversion-based interpretation requires satisfying initial models, which also rely on substantial manual interpretation, and the computational cost for 3D inversion is expensive.

Different from conventional interpretation, machine learning-based seismic interpretation, trained by labels from experienced interpreter or realistic model building, can deliver satisfying results with much higher efficiency, e.g., in seismic facies analysis (Wrona et al., 2018), faults identification (Wu et al., 2019), salt bodies delineation (Guillen et al., 2015) and horizon detection (Geng et al., 2020). For CO<sub>2</sub> interpretation in CCS projects, machine learning should be more attractive due to its high efficiency and inherent consistency throughout the long-term monitoring history. Using synthetic data generated by flow simulation, rock physics modeling and acoustic wave equation modeling, Wang et al. (2020) applied different machine learning algorithms to predict CO<sub>2</sub> saturation from seismic attributes and other downhole measurements. Sinha et al. (2020) treated the CO<sub>2</sub> leakage detection as an anomaly detection problem from CO<sub>2</sub> injection rates and pressure data using various neural networks (NNs). These successful applications are limited to frequently repeated small-scale measurements, whereas large-scale investigations, e.g., time-lapsed 3D surface seismic survey, are rarely interpreted by machine learning.

In this study, we propose a NN-based CO<sub>2</sub> interpretation method that offers end-to-end mapping from 4D seismic images, consisting of the baseline and time-lapse pairs, to 3D probability of CO<sub>2</sub> distribution. We employ a simplified 3D U-net which has been successfully utilized for fault detection from 3D seismic images (Wu et al., 2019). We train it by the shared seismic dataset and the benchmark model from the Sleipner CCS project, which is the world first industrial offshore CCS project starting injection from 1996, and storing around 18.5 million tonnes CO<sub>2</sub> by 2020 (Williams & Chadwick, 2021). The trained NN is applied to different monitoring time-lapse datasets acquired and processed in different years compared with the originally processed baseline dataset. With a single TITAN RTX GPU (24 G), the runtime for training and validation is around 3 hours, while the application takes only few seconds to obtain a complete 3D volume of CO<sub>2</sub> distribution. The NN interpretation results show high resolution with valid consistency throughout all available datasets. In addition, the NN also exhibits reasonable robustness against processing-induced mismatch between the baseline and time-lapse images, which can potentially alleviate the reprocessing demands on baseline dataset to be compared with newly acquired or reprocessed time-lapse datasets. To understand how the trained NN interprets, we test its performance based on crafted data samples. The outcome suggests that our trained NN considers both amplitude difference and structural similarity for detect-



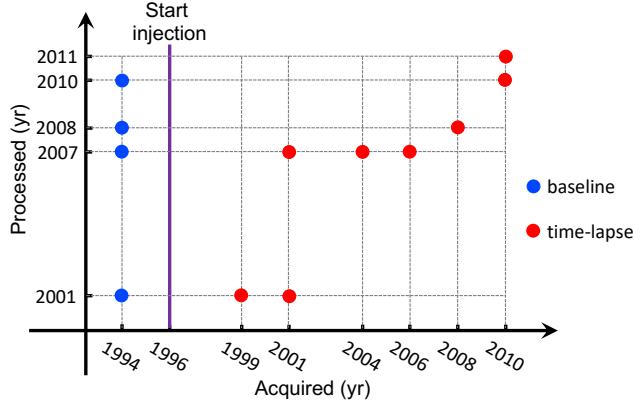
**Figure 1.** Simplified 3D U-Net for CO<sub>2</sub> interpretation from 4D seismic images.

ing CO<sub>2</sub> distribution. Finally, we discuss a realistic situation when only sparse 2D human interpretations (labels) are available for training. The comparison between 2D and 3D U-Nets reveals that the 3D U-Net is advantageous in terms of interpretation resolution and continuity, but requires much more computational resources.

## 2 Simplified 3D U-Net for CO<sub>2</sub> interpretation

The CO<sub>2</sub> interpretation in this study aims at depicting 3D CO<sub>2</sub> distribution in the reservoir from 4D seismic images. Therefore, we consider it as an image segmentation task, which assigns ones to CO<sub>2</sub> saturated parts, and zeros to the other parts. Similar image segmentation has been achieved in fault identification from seismic images (Wu et al., 2019), using a simplified version of the original U-Net (Ronneberger et al., 2015). We make some adjustments on the fault-detection network for CO<sub>2</sub> interpretation as shown in Figure 1. Firstly, our network input has two channels accommodating both baseline and time-lapse seismic images for CO<sub>2</sub> interpretation. Such input structure attempts to mimic conventional interpretation, where we identify CO<sub>2</sub> according to amplitude changes and velocity pushdown in the time-lapse image w.r.t the baseline image. The analysis (downward) and synthesis (upward) paths of the NN are almost identical to those used for fault segmentation. However, our network uses half the number of feature maps at the bottom of the U-Net, to reduce memory and computational cost while preserving the performance for CO<sub>2</sub> interpretation. In addition, we add an extra layer of batch normalization (BN) (Ioffe & Szegedy, 2015) before each rectified linear unit (ReLU), for faster convergence, as suggested by (Çiçek et al., 2016). The kernel size for max pooling in the downward layer is  $2 \times 2 \times 2$  with a stride of 2 along each dimension. Correspondingly, we use trilinear algorithm for upsampling with the scale factor of 2. The kernel size for the 3D convolutional layer is  $3 \times 3 \times 3$ , except for the output layer, where convolution kernel size becomes  $1 \times 1 \times 1$ . The final Sigmoid function outputs a 3D probability volume for CO<sub>2</sub> presence in the corresponding seismic image domain.

The dimension of 3D cubes for both input and output is  $64 \times 64 \times 64$ , with the voxel size of  $25 \text{ m} \times 25 \text{ m} \times 8 \text{ ms}$  along inline, crossline and traveltime directions, respectively. Hence, the receptive field of each voxel in the bottom feature map is  $200 \text{ m} \times 200 \text{ m} \times 64 \text{ ms}$ , which provides a satisfying balance between the computational cost and feature detection for CO<sub>2</sub> interpretation. The 3D cube covers the volume of  $1.6 \text{ km} \times 1.6 \text{ km} \times 512 \text{ ms}$ , sitting in the entire data volume of  $3.35 \text{ km} \times 7.0 \text{ km} \times 2000 \text{ ms}$ . To



**Figure 2.** Acquiring and processing years for available seismic images in the 4D Sleipner seismic dataset.

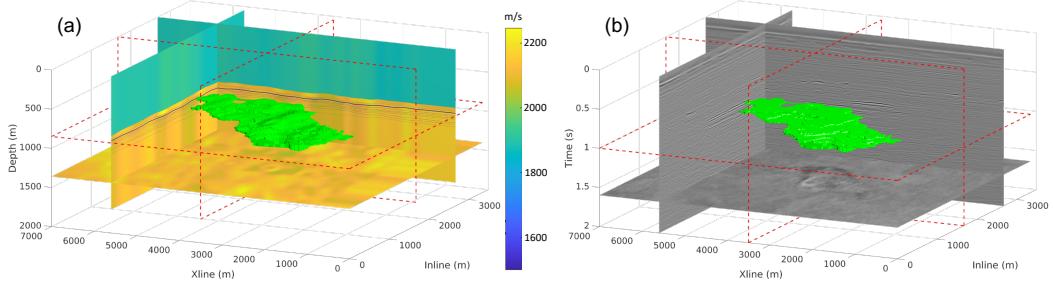
147 avoid the influence of disparate amplitude ranges between multi-vintages with different  
 148 processing routes, we normalize the input cubes of baseline and time-lapse images sep-  
 149 arately, i.e., each image cube is subtracted by its mean value and divided by its stan-  
 150 dard deviation before being concatenated and fed to the NN.

### 151 3 Training and validation

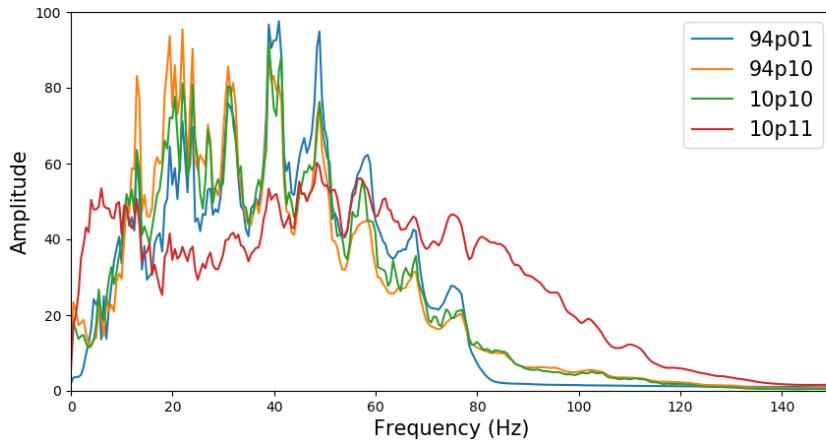
#### 152 3.1 Datasets generation

153 For the training input, publicly shared Sleipner 4D seismic dataset (Equinor, 2020a)  
 154 includes abundant pairs of baseline and time-lapse datasets that have gone through time-  
 155 lapse processing in different years. Figure 2 shows the acquired and processed years of  
 156 different datasets. The baseline data is acquired in 1994, before the CO<sub>2</sub> injection started  
 157 in 1996 (Baklid et al., 1996). Then multiple time-lapse surveys have been repeated. In  
 158 the shared datasets, the time-lapse data acquired in 1999, 2001, 2004, 2006, 2008 and  
 159 2010, are available. We refer to a certain dataset as *xxpyy*, if it is acquired in the year  
 160 of *xx* and processed in the year of *yy*. Each time-lapse dataset *xxpyy* has a correspond-  
 161 ingly reprocessed baseline dataset 94pyy, except for 10p11, which only has gone through  
 162 image processing route without a matched baseline. Although all shared datasets con-  
 163 tain near-, middle-, far-, and full-offset stacked images, we only utilize the near-offset im-  
 164 ages, which are adequate for the CO<sub>2</sub> interpretation in this study.

165 For the training labels, it requires attested CO<sub>2</sub> interpretation corresponding to  
 166 the input data. Here, we utilize the interpreted CO<sub>2</sub> plume boundaries of 2010 in nine  
 167 internal sandstone layers provided in Sleipner 2019 Benchmark Model (Equinor, 2020b).  
 168 Due to the lack of information for the CO<sub>2</sub> layers' thickness, we assume that the whole  
 169 corresponding sandstone layer is fully saturated with CO<sub>2</sub> within the plume's lateral bound-  
 170aries. Figure 3a illustrates the labeled CO<sub>2</sub> distribution in the model domain, along with  
 171 the available interval velocity and reservoir interfaces from the benchmark model. Us-  
 172 ing these information, we convert the CO<sub>2</sub> labels along depth into the image domain along  
 173 travelttime (Figure 3b). The generated label can be seen as the 3D probability volume  
 174 of CO<sub>2</sub> distribution with lateral resolution of 100 m (as indicated in the benchmark model),  
 175 and vertical resolution of corresponding sandstone layers' thickness whose average is about  
 176 30 m (in terms of two-way travelttime about 28 ms). Considering the interpretation un-  
 177 certainties in the interval velocity and interface positions for the depth to travelttime con-  
 178 version, we slightly smooth the generated image-domain label by a 3D Gaussian filter,  
 179 so that the label margins are adjusted to lower confidence.



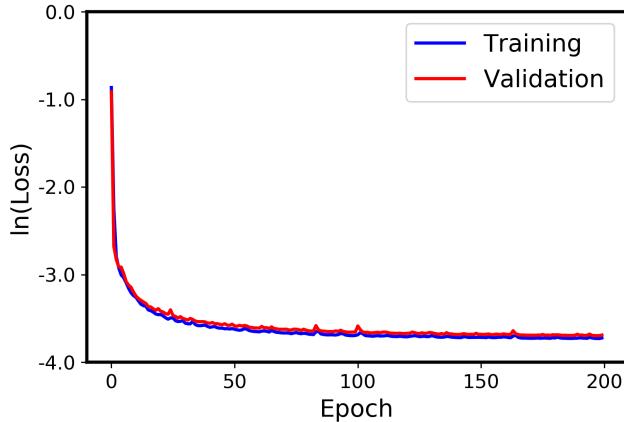
**Figure 3.** CO<sub>2</sub> plume labels for 2010 in (a) the model domain and (b) the image domain. The green blobs represent the CO<sub>2</sub> plume mask in 3D model or image domain. The red dashed squares indicate the sampled slice positions in 3D. The background color of the slices in (a) represent the interval velocity, while the wiggles represent the depth of interfaces in Utsira formation. The slices shown in (b) are stacked near-offset images from 10p10.



**Figure 4.** Comparison among amplitude spectra of 94p01, 94p10, 10p10, and 10p11.

Since the labels are limited to 2010, it is natural to choose the baseline and time-lapse dataset pair for 2010, i.e., 94p10 and 10p10, to generate training inputs. However, we also include the originally processed baseline data 94p01 and the newly processed time-lapse data 10p11 in training dataset generation. Consequently, there are four types of combinations between baseline and time-lapse images as 94p01 vs. 10p10, 94p01 vs. 10p11, 94p10 vs. 10p10, and 94p10 vs. 10p11. Figure 4 shows the comparison among amplitude spectra of the four datasets. Average spectra of 94p10 and 10p10 are mostly identical to each other, only slightly different from that of 94p01. However, 10p11 shows marked difference in the bandwidth compared to the others, since it has not gone through the time-lapse processing route. Hence, the various dataset combinations include processing discrepancies between the baseline and time-lapse inputs which are necessary to improve the NN's tolerance against processing-induced image-space mismatch.

We randomly sample corresponding 3D cubes in one baseline, one time-lapse and the label volumes. Originally, we have 720 cubes whose centers distributing randomly in the entire 3D volume. For each cube, there are the four combinations of baseline and time-lapse inputs. Therefore, total 2880 samples are composed. Among these samples, around 60% do not have any overlap with the labeled CO<sub>2</sub> plume, having all zero probability values in their labels. This bias in the training dataset significantly slows down



**Figure 5.** Training and validation losses in natural log scale.

198 the convergence. Therefore, we reduce the number of all-zero CO<sub>2</sub> samples by randomly  
 199 discarding some of them. Eventually, we have 1576 samples, of which only 480 (around  
 200 1/3) are all-zero CO<sub>2</sub> samples. We divide these samples randomly into training and val-  
 201 idation datasets with 1500 and 76 samples, respectively.

### 202 3.2 Training and validation

203 We use the binary cross-entropy (BCE) loss function since our labels are proba-  
 204 bility between zero and one. The total training epoch is 200. We use the Adam optimizer  
 205 (Kingma & Ba, 2014) with a learning rate of 0.0002. The batch size is 30 to prevent over-  
 206 fitting and improve the training efficiency with a single TITAN RTX GPU (24G). We  
 207 use pytorch to implement the training and validation process and it takes approximately  
 208 3 hours.

209 Figure 5 shows the training and validation losses varying with epoch. Both train-  
 210 ing and validation losses decrease by nearly 3 orders of magnitudes. In Figure 6, we show  
 211 the NN predictions for two different samples from the training datasets. The NN pre-  
 212 dictions for both samples are quite consistent with the labels, despite of slightly lower  
 213 resolution. More epochs or smaller batch size could further improve the resolution, but  
 214 the robustness and generalization of the NN may be compromised.

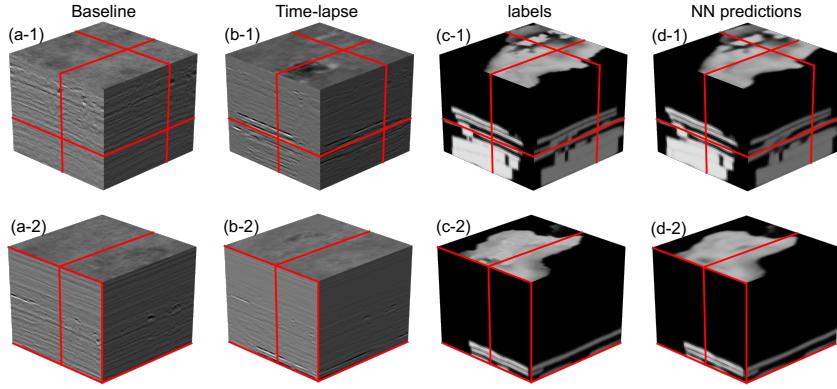
## 215 4 Applications

### 216 4.1 Robustness test

217 After our NN is trained based on random samples from the seismic images and cor-  
 218 responding CO<sub>2</sub> labels for 2010, we firstly apply it to the same 4D seismic datasets, but  
 219 with regularly sampled cubes on the entire 3D volume. The sampling number is 3×5×  
 220 6 along inline, crossline and traveltime directions, respectively, resulting in 90 samples  
 221 with approximately 40% overlapping in 3D. The runtime for such 90-sample test takes  
 222 only a few seconds. The obtained NN predictions are reconstructed into the original seis-  
 223 mic image dimension through weighted summation. The 3D weighting function is:

$$224 \quad w(x, y, t) = f(x)f(y)f(t), \quad (1)$$

226 where  $x$ ,  $y$  and  $t$  represent inline, crossline and traveltime directions, respectively;  $f$  de-  
 227 notes a 1D weighting function with ones in the middle and gradually decaying to zero



**Figure 6.** Trained NN predictions for two samples from the training dataset. The first two columns are the baseline and time-lapse seismic images as the NN input, and the last two columns are the human interpreted labels and NN predictions. The red lines indicate the slice positions shown by the cube surfaces.

towards the edges using a Hanning window as follows:

$$f(a) = \begin{cases} 1, & |a| \leq \frac{1}{2}\alpha L \\ \cos^2\left(\frac{\pi}{L}(|a| - \frac{1}{2}\alpha L)\right), & \frac{1}{2}\alpha L < |a| \leq \frac{L}{2} \\ 0, & |a| > \frac{L}{2}, \end{cases} \quad (2)$$

(3)

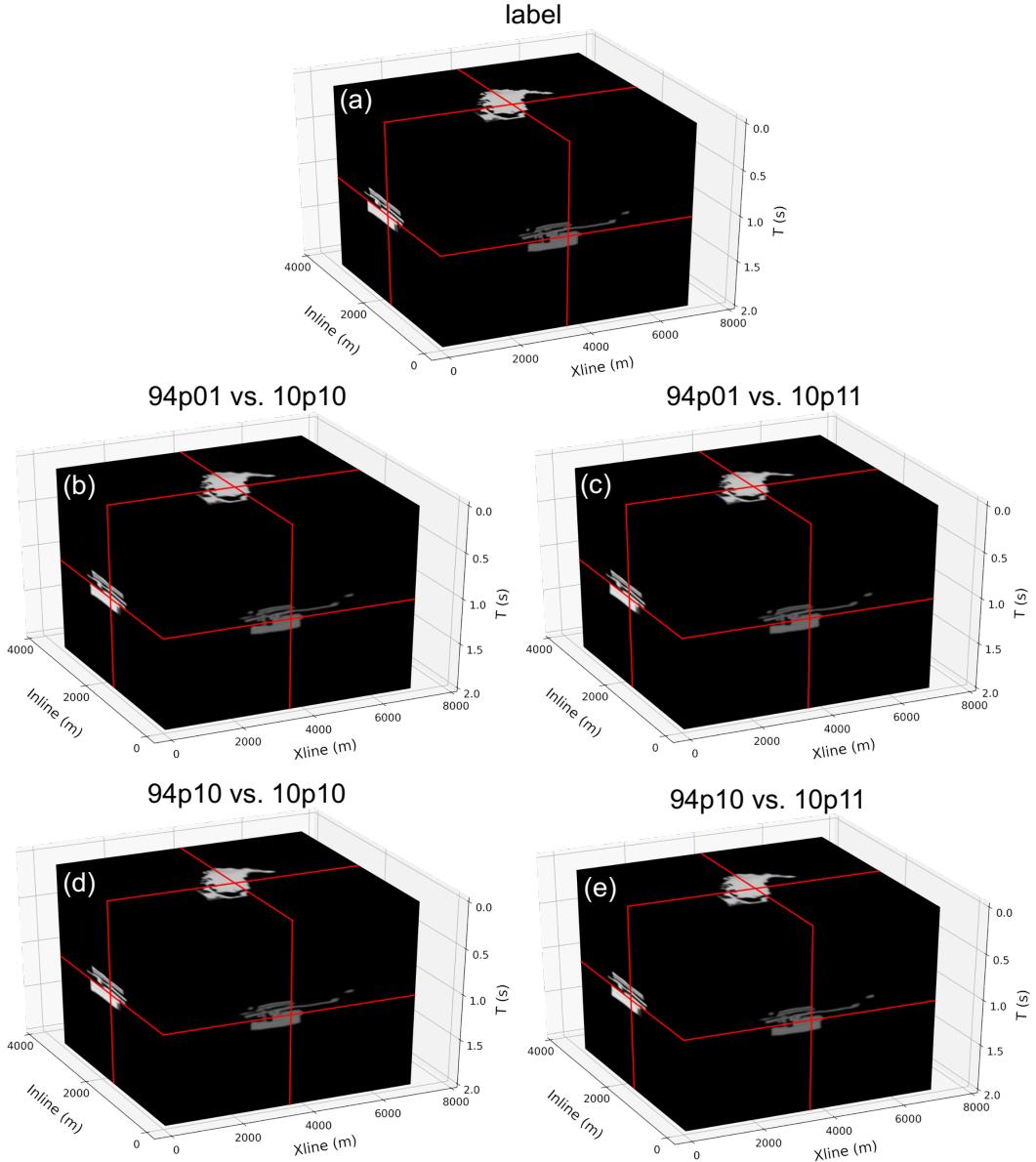
where  $L = 64$  is the NN output size along each dimension,  $\alpha = 0.6$  indicates the portion of the full confidence in the middle of the weighting function. Thus, the weighted summation of the 90 NN predictions is

$$M(x, y, t) = \frac{\sum_{i=1}^{90} p_i w(x - x_i, y - y_i, t - t_i)}{\sum_{i=1}^{90} w(x - x_i, y - y_i, t - t_i)}, \quad (4)$$

where  $M$  is the reconstructed 3D CO<sub>2</sub> prediction,  $p_i$  is the NN prediction for the  $i$ th sample, whose cube center is  $(x_i, y_i, t_i)$ .

To test the NN robustness against processing-induced mismatch between baseline and time-lapse inputs, we apply our NN on the four different sets of 90 samples corresponding to the four different combinations as 94p01 vs. 10p10, 94p01 vs. 10p11, 94p10 vs. 10p10, and 94p10 vs. 10p11, respectively. Figure 7 shows the reconstructed CO<sub>2</sub> distributions based on different NN predictions. We can see that the predicted CO<sub>2</sub> distributions for different combinations of baseline and time-lapse images are almost identical to the label (Figure 7e). Further calculating the BCE loss between each reconstructed prediction and the label, Table 1 shows that 94p10 vs. 10p10 provides the best result, and using 10p11 as time-lapse input always leads to larger losses in comparison with 10p10. Such observations are consistent with the dataset similarity represented by the spectra comparison shown in Figure 4. Regardless of the insignificant differences in the BCE losses, the NN predictions are visually undifferentiated, reflecting strong robustness of the trained NN against moderate processing mismatch between the baseline and time-lapse images.

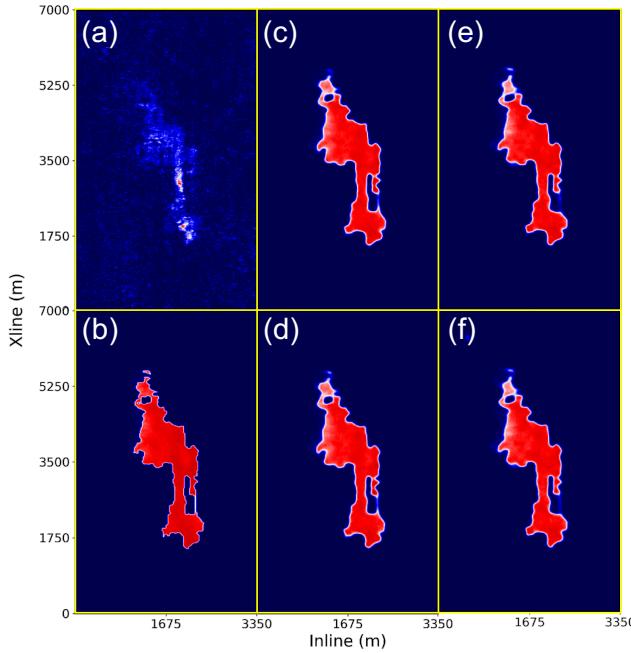
In Figure 8, we compare the seismic amplitude anomaly with the human interpreted and NN predicted CO<sub>2</sub> distributions in the top sand wedge layer above the Utsira Formation. We take a vertical average of the absolute amplitude differences between 94p10 and 10p10 within this layer, to indicate the footprint of CO<sub>2</sub>. Similarly, we take a vertical average of the the interpreted CO<sub>2</sub> distributions for an areal comparison. Although



**Figure 7.** Comparison between (a) the label and (b,c,d,e) the reconstructed CO<sub>2</sub> distributions based on NN predictions using four different combinations of baseline and time-lapse images.

**Table 1.** BCE loss of the reconstructed 2010 CO<sub>2</sub> distributions w.r.t the label.

	BCE loss	baseline	94p01	94p10
time-lapse				
10p10			4.86e <sup>-3</sup>	4.85e <sup>-3</sup>
10p11			4.91e <sup>-3</sup>	4.93e <sup>-3</sup>

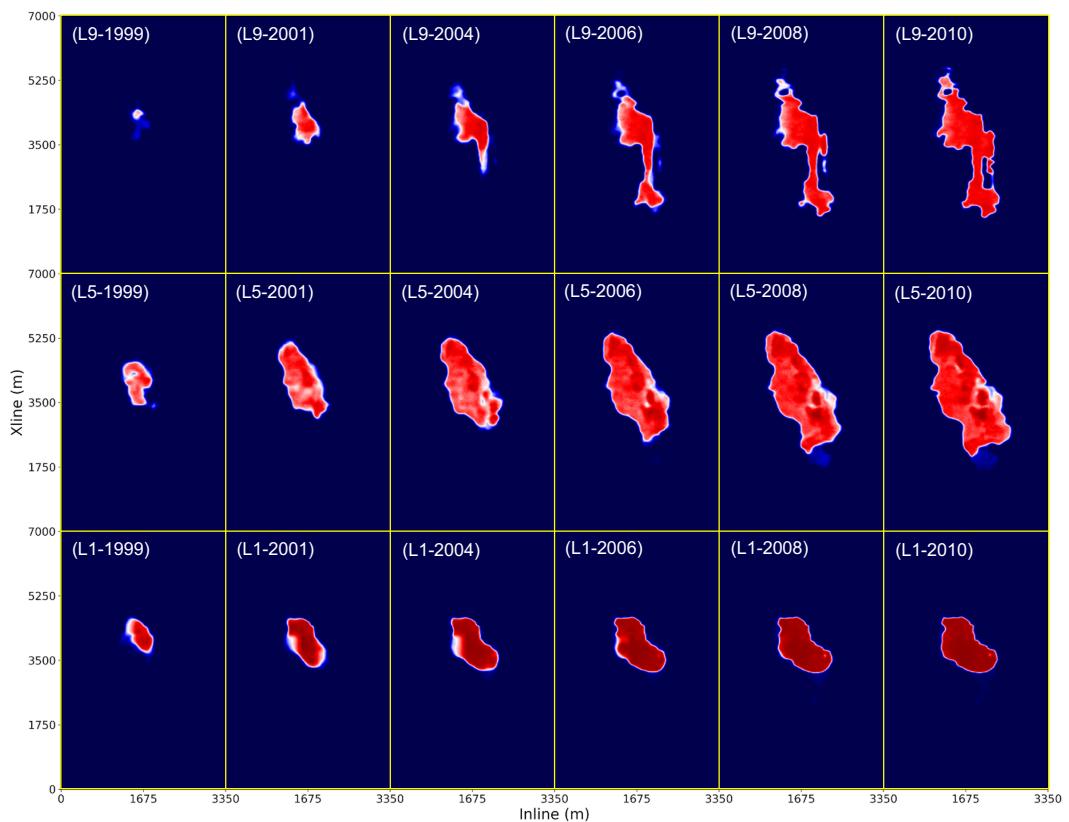


**Figure 8.** Comparison between (a) the absolute amplitude anomaly with (b) the human interpreted and (c,d,e,f) NN predicted CO<sub>2</sub> distributions in the top sand wedge layer above the Utsira Formation. For baseline input, (c) and (e) use 94p01, while (d) and (f) use 94p10; for time-lapse input, (c) and (d) use 10p10, while (e) and (f) use 10p11.

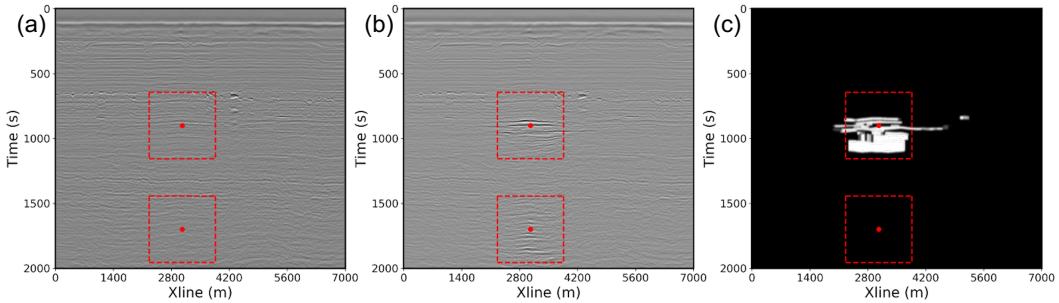
the amplitude indicator does provide a rough clue for the CO<sub>2</sub> distribution, it requires further processing and more detailed analysis to obtain the high-resolution CO<sub>2</sub> plume as shown in the label (Figure 8b). The trained NN, on the other hand, achieves accurate CO<sub>2</sub> depiction with high resolution directly from the baseline and time-lapse images, even when noticeable processing mismatch exists.

#### 4.2 Consistency test

To test the generalizability of our trained NN, we apply it to other available 4D seismic vintages shared by the Sleipner CO<sub>2</sub> storage project. We use the same originally processed 94p01 as the NN baseline input for all time-lapse inputs: 99p01, 01p01, 04p07, 06p07, 08p08, and 10p10. Figure 9 displays the NN interpreted CO<sub>2</sub> distribution in the top (L9), middle (L5) and base (L1) of the internal sandstone layers in Utsira Formation, developing from 1999 to 2010. In all displayed layers, the NN predictions are compacted with clear and continuous boundaries. Moreover, they grow steadily throughout the decade, although the CO<sub>2</sub> plume in L1 expands noticeably slower due to the buoyancy of the supercritical CO<sub>2</sub> in the saline aquifer (Arts et al., 2008). In 1999, the top layer result (Figure 9(L9-1999)) shows a singularity on the probability map, indicating the injected CO<sub>2</sub> has just reached the top of the formation (A. Chadwick et al., 2010). Similar singularities are also visible in Figures 9 (L9-2004), (L9-2008) and (L5-1999), suggesting that our NN interpretation has the potential for high-resolution leakage detection or feeder recovery. Finally, we can also identify the migration directions of CO<sub>2</sub> plume in different layers. Generally, the NN interpretations along time are reasonably consistent in terms of CO<sub>2</sub> migration and plume expansion in the storage unit. A movie of the predicted 3D CO<sub>2</sub> growth from 1999 to 2010 are provided in the supplementary materials of this paper.



**Figure 9.** NN interpreted CO<sub>2</sub> plume expanding along L9 (top), L5 (middle) and L1 (base) of the Utsira Formation, from 1999 to 2010.



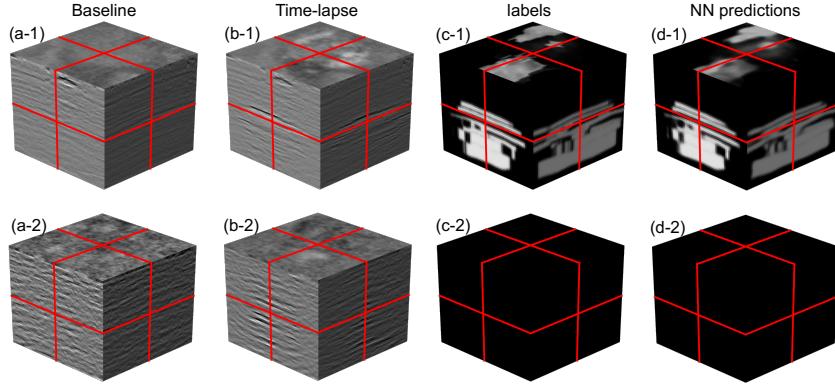
**Figure 10.** The inline assembles at 1625 m for (a) baseline and (b) time-lapse images from 94p10 and 10p10, along with the (c) CO<sub>2</sub> distribution label. The red dots and dashed lines indicate the centers and boundaries of the sampled cubes within the inline assemble.

## 5 Discussion

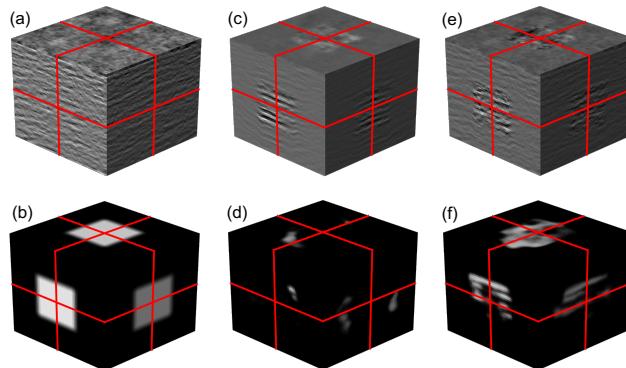
### 5.1 Analysis for NN interpretation standards

We design experiments with controlled inputs to test the performance of the trained NN, in order to shine some light on the hidden mechanism of the trained NN. Figure 8 suggests that large amplitude differences between the time-lapse image and the baseline image are strong indicators of CO<sub>2</sub> presence. In view of this intuitive hypothesis, we sample two cubes whose centers are both in the inline assemble at 1625 m from 94p10 and 10p10 as shown in Figure 10. The first sample includes the primary reflections caused by CO<sub>2</sub> accumulation in the storage unit, and the second sample contains the corresponding surface-related multiples right below the first sampling position. Human interpreters who understand the seismic imaging principles have no trouble differentiating the true amplitude anomaly and the artifacts. Hence, despite amplitude anomalies in both samples, their labels are drastically different as shown in Figure 10c. We feed these two samples of baseline and time-lapse cubes (Figures 11a and b) into the trained NN, and the predictions are shown in Figure 11d, which are consistent with the labels (Figure 11c). This implies that the NN does not solely rely on the amplitude difference between the baseline and time-lapse images to determine the CO<sub>2</sub> distribution.

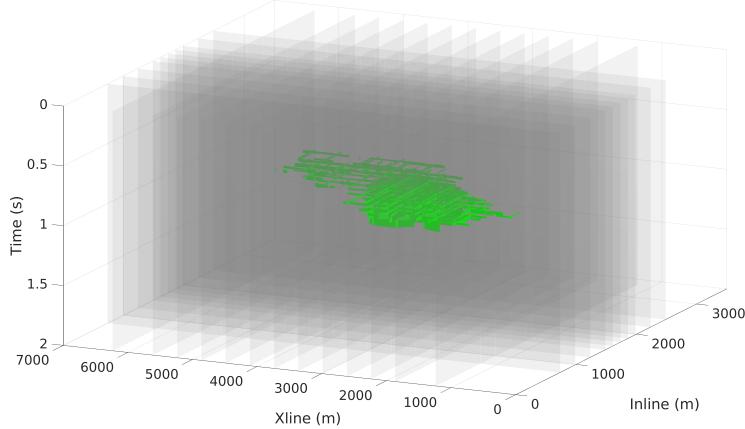
Another aspect human interpreters may utilize is the similarity of the images. Hence, we modify the time-lapse image in the second sample shown in Figure 11. The modifications and corresponding NN predictions are shown in Figure 12. In the first modified sample, we scale up the center area of the *time-lapse image* (Figure 11b-2) as the new time-lapse image (Figure 12c). For comparison, in the second modified sample, we scale up the center area of the *baseline image* (Figure 11b-1) as the new time-lapse image (Figure 12e). Both modified samples now share the same baseline image as shown in Figure 12a and the locally scaling multiplier is displayed in Figure 12b, whose center area has the value of 4, whereas the outer area is 1. By feeding the modified samples to the trained NN, we obtain the predictions shown in Figures 12d and f. It appears that increasing the amplitude of the multiples only slightly increase the chance of mis-interpretation in the NN prediction. However, when we directly scale up the baseline amplitude in specific area as the time-lapse image, the trained NN predicts a much higher probability for CO<sub>2</sub> presence in the scaled-up area. These testing results suggest that the trained NN also considers the structural similarity between the baseline and time-lapse images, in addition to their amplitude difference. This indicates that our trained NN performs similarly to a human interpreter.



**Figure 11.** Trained NN predictions for the two samples shown in Figure 10. The first two columns are the baseline and time-lapse seismic images as the NN input, and the last two columns are the human interpreted labels and NN predictions.



**Figure 12.** Modifications of the time-lapse images in the second sample shown in Figure 11. (a) is the common baseline; (b) is the locally scaling multiplier; (c) and (e) are new time-lapse images; (d) and (f) are NN predictions corresponding to (c) and (e), respectively.



**Figure 13.** 2D slice labels for CO<sub>2</sub> distribution, indicated by the green patches on each slice represented by translucent surfaces.

## 315 5.2 2D vs. 3D NN using sparse 2D labels

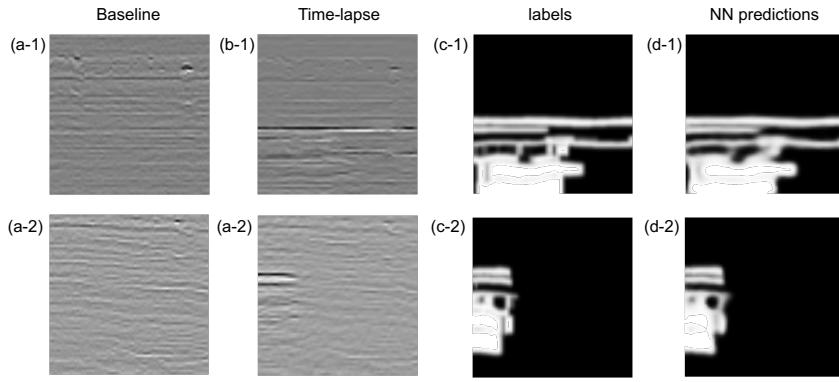
316 The presented NN based on the Sleipner 4D seismic dataset is trained by a com-  
 317 plete 3D label generated from CO<sub>2</sub> plume boundary interpretation for 2010. However,  
 318 in more general cases, the interpreted labels for CO<sub>2</sub> distribution are often available on  
 319 sparse 2D slices of inline and/or crossline assemblies, since human perception excels in  
 320 2D. We create an example of such scenario as illustrated in Figure 13. We sample 10 slices  
 321 along the inline direction, and 13 slices along the crossline direction. The slice interval  
 322 is smaller (125 m along inline direction and 375 m along crossline direction) near the stor-  
 323 age unit, and larger (250 m along inline direction and 625 m along crossline direction)  
 324 away from the storage unit, for a better representation of the label target.

325 One way of utilizing such sparse 2D labels is to directly train a 2D U-Net. Another  
 326 way is to train the 3D U-Net with corresponding weight on the sparse labels during loss  
 327 evaluation (Çiçek et al., 2016). Here, we compare these two strategies to offer a guide-  
 328 line under such realistic scenario.

### 329 5.2.1 2D U-Net with sparse 2D labels

330 We use the same architecture shown in Figure 1 for the 2D U-Net, except we re-  
 331 duce the dimensionality from 3D to 2D for convolution, max pooling and upsampling.  
 332 To generate the training dataset, we also utilize both 94p01 and 94p10 as the baseline  
 333 input, along with 10p10 and 10p11 as the time-lapse input. Since trace intervals along  
 334 both inline and crossline directions are the same, we can train a 2D U-Net applicable  
 335 on both inline and crossline assemblies. The 2D patch size is 64×64, with the grid size  
 336 of 25 m × 8 ms along inline/crossline and traveltime directions, respectively. We sam-  
 337 ple 2D patches on all available slices. Similar to the 3D case, we keep 1500 samples for  
 338 training, of which around 25% are all-zero CO<sub>2</sub> samples. The 2-D U-Net training uses  
 339 the same training parameters as for the 3D U-Net. Due to the dimensionality reduction,  
 340 the training time has reduced significantly from 3 hours to 5 minutes for 200 epochs us-  
 341 ing the same TITAN RTX GPU. Figure 14 shows the results of two training samples.  
 342 It appears that the NN predictions are consistent with the corresponding 2D labels.

343 To test the trained 2D U-Net, we apply it on 94p01 vs. 04p07 and 94p01 vs. 10p10,  
 344 respectively, by regularly sampling 2D patches along all inline and crossline assemblies  
 345 for the 3D dataset volume. The sampling numbers are also (3, 5, 6) along inline, crossline



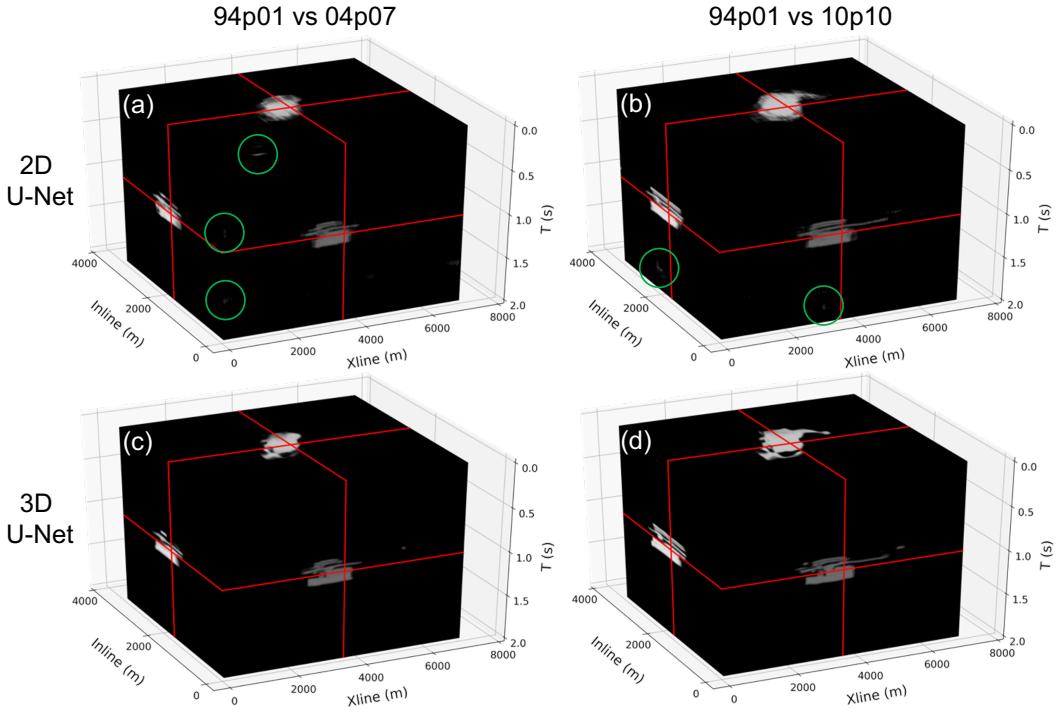
**Figure 14.** Trained 2D NN predictions for two samples from the training dataset with sparse 2D slice labels. The first two columns are the baseline and time-lapse seismic images as the NN input, and the last two columns are the human interpreted labels and NN predictions.

and traveltime directions. We combine these 2D NN predictions together by the weighted summation as shown in Equation 4 with 2D Hanning-windowed weighting functions. Figures 15a and b display the reconstructed 3D CO<sub>2</sub> distributions for the two tests. Compared to the label and the predictions from the original 3-D U-Net shown in Figure 7, the 2010 result predicted by 2D U-Net (Figures 15a) shows lower resolution and more artifacts as indicated by the green circles. Similar defects are also visible in the 2004 result obtained from 2D U-Net (Figures 15b). We display the reconstructed 2D CO<sub>2</sub> distributions along the top sand wedge layer for both tests in Figures 16a and b. Distinct stripping artifacts are visible in comparison with the corresponding 3D U-Net predictions shown in Figure 9. This is because the 2D U-Net cannot preserve the continuity along the 3rd dimension perpendicular to the plane where the 2D U-Net is applied. Although we apply the 2D U-Net along both inline and crossline assemblies before combining them by weighted summation, the outcomes simply present discontinuities along both directions as shown in Figures 16a and b.

### 5.2.2 3D U-Net with sparse 2D labels

Although the labels are sparsely generated in 2D, we can still train a 3D U-Net with the same structure. For each sampled cube, the baseline and time-lapses images are unchanged, but the label now has a corresponding weight, in which the sampled 2D slice position is assigned as one, whereas other part is zero. The weight is implemented during the BCE loss calculation and then backpropagated to influence the NN update. The runtime for 200 epochs are basically the same as for the original 3D U-Net training. Figure 17 shows the same samples displayed in Figure 6. The NN inputs are exactly the same, whereas the labels only contain limited vertical slices (Figures 17a-3 and b-3) instead of the whole cube (Figures 6a-3 and b-3). Nonetheless, the NN predictions are reasonably consistent. The horizontal slices shown in Figures 17a-4 and b-4 have been retrieved with satisfying resolution and continuity, even though they are not exactly the same as predictions from NN trained by full 3D labels shown in Figures 6a-4 and b-4.

We also test the sparsely trained 3D U-Net on 94p01 vs. 04p07 and 94p01 vs. 10p10, respectively. The 3D reconstructed CO<sub>2</sub> distributions are shown in Figures 15c and d. It appears that the sparsely trained 3D U-Net results provide much higher resolution than those obtained from 2D U-Net (Figures 15a and b). Figures 16c and d display the top sand wedge layer CO<sub>2</sub> distributions from the sparsely trained 3D U-Net. Compared to



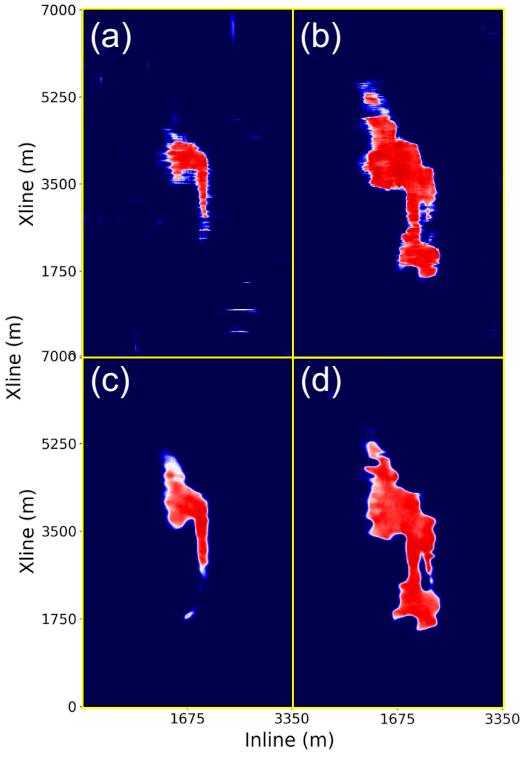
**Figure 15.** Reconstructed 3D CO<sub>2</sub> distributions from NN predictions for the tests of 94p01 vs. 04p07 and 94p01 vs. 10p10, using trained 2-D U-Net and 3-D U-Net, respectively. The green circles indicate the positions of artifacts.

the 2D U-Net results (Figures 16a and b), there are no stripping artifacts and the plume boundaries exhibit more continuity.

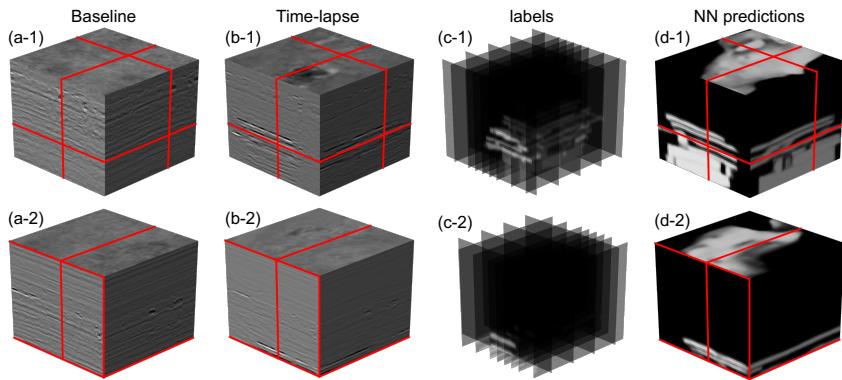
In summary, the 3D U-Net trained by weighted sparse labels generally present higher-quality interpretation than the 2D U-Net trained by the same labels, in terms of resolution, boundary continuity and artifacts. However, the 3D U-Net training and applications require much more computational resources than the 2D U-Net. Hence, we suggest to use 3D U-Net even for sparse 2D labels as long as necessary computational power (GPU with large enough memory) is accessible.

## 6 Conclusions

We utilized a simplified 3D U-Net to interpret the 3D CO<sub>2</sub> distribution from large 4D seismic images. We trained the NN using seismic images acquired at different times and processed with different routes to improve its robustness against processing-induced mismatch between the baseline and time-lapse images. When applying the NN to seismic data acquired in other years, we obtained high-resolution interpretation results with inherent consistency. Probing the trained NN with crafted inputs suggests that the trained NN predicts the presence of CO<sub>2</sub> not only based on the amplitude of the image differences, but also considering their structural similarity to the baseline image. We also provide two NN training strategies under more realistic scenarios where only sparse 2D labels are available. Both strategies achieve reasonable predictions and can be utilized according to the availability of computational resources. Our study demonstrates that the proposed 3D U-Net speeds up CO<sub>2</sub> interpretation with guaranteed consistency during the life time of the CCS projects.



**Figure 16.** NN interpreted CO<sub>2</sub> distributions in the top sand wedge layer above the Utsira Formation. (a) and (b) are the 2D U-Net results from 94p01 vs. 04p07 and 94p01 vs. 10p10, respectively, while (c) and (d) are the 3D U-Net results from 94p01 vs. 04p07 and 94p01 vs. 10p10, respectively.



**Figure 17.** Trained 3D NN predictions for two samples from the training dataset with sparse 2D slice labels. The first two columns are the baseline and time-lapse seismic images as the NN input, and the last two columns are the human interpreted labels and NN predictions.

400 **Acknowledgments**

401 The authors acknowledge the Singapore Economic Development Board for its financial  
 402 support through the Petroleum Engineering Professorship. This study is made possible  
 403 by the publicly shared Sleipner 4D seismic dataset at [https://co2datashare.org/dataset/  
sleipner-4d-seismic-dataset](https://co2datashare.org/dataset/sleipner-4d-seismic-dataset) and the Sleipner 2019 Benchmark Model at [https://  
co2datashare.org/dataset/sleipner-2019-benchmark-model1](https://co2datashare.org/dataset/sleipner-2019-benchmark-model1). The reproducible codes  
 404 and supplementary materials related to this article can be found online [https://github  
.com/nusbei/C02\\_Sleipner](https://github.com/nusbei/C02_Sleipner).

408 **References**

- 409 Arts, R., Chadwick, A., Eiken, O., Thibeau, S., & Noonan, S. (2008). Ten years'  
 410 experience of monitoring CO<sub>2</sub> injection in the Utsira Sand at Sleipner, offshore  
 411 Norway. *First Break*, 26(1), 65–72. doi: 10.3997/1365-2397.26.1115.27807
- 412 Bachu, S., & Adams, J. (2003). Sequestration of CO<sub>2</sub> in geological media in re-  
 413 sponse to climate change: capacity of deep saline aquifers to sequester CO<sub>2</sub>  
 414 in solution. *Energy Conversion and Management*, 44(20), 3151–3175. doi:  
 415 10.1016/S0196-8904(03)00101-8
- 416 Baklid, A., Korbol, R., Owren, G., et al. (1996). Sleipner Vest CO<sub>2</sub> disposal, CO<sub>2</sub>  
 417 injection into a shallow underground aquifer. In *SPE Annual Technical Confer-  
418 ence and Exhibition* (p. SPE-36600-MS). doi: 10.2118/36600-MS
- 419 Boait, F., White, N., Bickle, M., Chadwick, R., Neufeld, J., & Huppert, H. (2012).  
 420 Spatial and temporal evolution of injected CO<sub>2</sub> at the Sleipner Field, North  
 421 Sea. *Journal of Geophysical Research: Solid Earth*, 117(B3), B03309. doi:  
 422 10.1029/2011JB008603
- 423 Bourne, S., Crouch, S., & Smith, M. (2014). A risk-based framework for mea-  
 424 surement, monitoring and verification of the Quest CCS project, Alberta,  
 425 Canada. *International Journal of Greenhouse Gas Control*, 26, 109–126. doi:  
 426 10.1016/j.ijggc.2014.04.026
- 427 Bryant, E., Bryant, E. A., & Edward, B. (1997). *Climate Process and Change*. Cam-  
 428 bridge University Press.
- 429 Castelletto, N., Gambolati, G., & Teatini, P. (2013). Geological CO<sub>2</sub> sequestration  
 430 in multi-compartment reservoirs: Geomechanical challenges. *Journal of Geo-  
431 physical Research: Solid Earth*, 118(5), 2417–2428. doi: 10.1002/jgrb.50180
- 432 Chadwick, A., Williams, G., Delépine, N., Clochard, V., Labat, K., Sturton, S., ...  
 433 others (2010). Quantitative analysis of time-lapse seismic monitoring data at  
 434 the Sleipner CO<sub>2</sub> storage operation. *The Leading Edge*, 29(2), 170–177. doi:  
 435 10.1190/1.3304820
- 436 Chadwick, R., Arts, R., & Eiken, O. (2005). 4D seismic quantification of a growing  
 437 CO<sub>2</sub> plume at Sleipner, North Sea. In *Geological Society, London, Petroleum  
438 Geology Conference series* (Vol. 6, pp. 1385–1399). doi: 10.1144/0061385
- 439 Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D  
 440 U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In  
 441 *International Conference on Medical Image Computing and Computer-Assisted  
442 Intervention* (pp. 424–432).
- 443 Clochard, V., Delépine, N., Labat, K., & Ricarte, P. (2010). CO<sub>2</sub> plume imaging  
 444 using 3D pre-stack stratigraphic inversion: A case study on the Sleipner field.  
 445 *First Break*, 28(1), 91–96. doi: 10.3997/1365-2397.28.1.38083
- 446 Equinor. (2020a). *Sleipner 2019 Benchmark Model*. Retrieved from [https://  
co2datashare.org/dataset/e6f67cbd-abf3-4d85-a118-ed386a994c2c](https://co2datashare.org/dataset/e6f67cbd-abf3-4d85-a118-ed386a994c2c) doi:  
 447 10.11582/2020.00004
- 448 Equinor. (2020b). *Sleipner 4D Seismic Dataset*. Retrieved from [https://  
co2datashare.org/dataset/cbdc354c-fa61-4ab4-a0b4-134e1350a82b](https://co2datashare.org/dataset/cbdc354c-fa61-4ab4-a0b4-134e1350a82b)  
 449 doi: 10.11582/2020.00005
- 450 Furre, A.-K., Eiken, O., Alnes, H., Vevatne, J. N., & Kiær, A. F. (2017). 20 years of  
 451

- 453 Monitoring CO<sub>2</sub>-injection at Sleipner. *Energy Procedia*, 114, 3916–3926. doi:  
454 10.1016/j.egypro.2017.03.1523
- 455 Geng, Z., Wu, X., Shi, Y., & Fomel, S. (2020). Deep learning for relative geologic  
456 time and seismic horizons. *Geophysics*, 85(4), WA87–WA100. doi: 10.1190/  
457 geo2019-0252.1
- 458 Guillen, P., Larrazabal, G., González, G., Boumber, D., & Vilalta, R. (2015). Su-  
459 pervised learning to detect salt body. In *SEG Technical Program Expanded Ab-*  
460 *stracts* (pp. 1826–1829). doi: 10.1190/segam2015-5931401.1
- 461 Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network  
462 training by reducing internal covariate shift. In *International Conference on*  
463 *Machine Learning* (pp. 448–456).
- 464 Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization.  
465 *arXiv preprint arXiv:1412.6980*.
- 466 Romdhane, A., & Querendez, E. (2014). CO<sub>2</sub> Characterization at the Sleipner Field  
467 with Full Waveform Inversion: Application to Synthetic and Real Data. *En-*  
468 *ergy Procedia*, 63, 4358–4365. doi: 10.1016/j.egypro.2014.11.470
- 469 Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks  
470 for Biomedical Image Segmentation. In *International Conference on Medical*  
471 *Image Computing and Computer-Assisted Intervention* (pp. 234–241). doi: 10  
472 .1007/978-3-319-24574-4\_28
- 473 Sinha, S., de Lima, R. P., Lin, Y., Sun, A. Y., Symons, N., Pawar, R., & Guthrie,  
474 G. (2020). Normal or abnormal? Machine learning for the leakage detection in  
475 carbon sequestration projects using pressure field data. *International Journal*  
476 *of Greenhouse Gas Control*, 103, 103189. doi: 10.1016/j.ijggc.2020.103189
- 477 Wang, Z., Dilmore, R. M., & Harbert, W. (2020). Inferring CO<sub>2</sub> saturation from  
478 synthetic surface seismic and downhole monitoring data using machine learn-  
479 ing for leakage detection at CO<sub>2</sub> sequestration sites. *International Journal of*  
480 *Greenhouse Gas Control*, 100, 103115. doi: 10.1016/j.ijggc.2020.103115
- 481 Williams, G., & Chadwick, R. (2021). Influence of reservoir-scale heterogeneities  
482 on the growth, evolution and migration of a CO<sub>2</sub> plume at the Sleipner Field,  
483 Norwegian North Sea. *International Journal of Greenhouse Gas Control*, 106,  
484 103260. doi: 10.1016/j.ijggc.2021.103260
- 485 Wrona, T., Pan, I., Gawthorpe, R. L., & Fossen, H. (2018). Seismic facies analysis  
486 using machine learning. *Geophysics*, 83(5), O83–O95. doi: 10.1190/geo2017  
487 -0595.1
- 488 Wu, X., Liang, L., Shi, Y., & Fomel, S. (2019). FaultSeg3D: Using synthetic data  
489 sets to train an end-to-end convolutional neural network for 3D seismic fault  
490 segmentation. *Geophysics*, 84(3), IM35–IM45. doi: 10.1190/geo2018-0646.1