geosciences

MDPI

Article

Time-Lapse Seismic Inversion Using a 3D Neural Operator

Xiao Ma * and Tariq Alkhalifah

Physical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia; tariq.alkhalifah@kaust.edu.sa

* Correspondence: xiao.ma@kaust.edu.sa

Abstract

Time-lapse (4D) seismic inversion is a key tool for monitoring subsurface changes caused by fluid injection and production, but model-based workflows require repeated forward and adjoint evaluations, a wavelet estimate for every vintage, and survey-specific regularization tuning. We propose a 3D neural operator framework for post-stack time-lapse impedance inversion. A U-shaped neural operator is trained on synthetic impedance models and their convolutional post-stack responses and serves as a mesh-independent, differentiable surrogate of the post-stack forward operator. During inversion, the operator is frozen, and the impedance model is updated by gradient descent with automatic differentiation using one set of settings for all vintages of a survey. On the Kimberlina synthetic CO_2 model, the inversion reduces the error of the initial model for every vintage (relative error from 3.1–3.2% to 2.8–2.9%). On the Sleipner field data, the recovered impedance changes are concentrated inside an independently mapped CO_2 plume (median change 4.7–7.3 times larger than outside), the predicted 4D data difference correlates with the observed one at $r = 0.92$–0.93 inside the plume, and the operator is applied without retraining to the full $1001 \times 468 \times 249$ survey grid. We also quantify the training and fine-tuning costs and discuss the limits of a learned convolutional surrogate.

Keywords: time-lapse seismology; 4D seismic inversion; neural operator; post-stack inversion; acoustic impedance; inverse theory; deep learning; carbon capture and storage; Sleipner

1. Introduction

Seismic inversion aims to quantitatively estimate subsurface elastic or acoustic properties from recorded seismic data and has long been a central topic in exploration geophysics [1], geothermal energy production [2], and carbon capture and storage [3]. Within this broader context, time-lapse (4D) seismic inversion is particularly important for monitoring temporal variations of reservoir properties induced by hydrocarbon production, CO_2 injection, or geothermal exploitation, as the inverted differences between baseline and monitor surveys serve as direct quantitative indicators of fluid migration, pressure variations, and saturation changes [4]. Over the past several decades, a variety of traditional time-lapse inversion strategies have been developed. The simplest is the parallel inversion scheme, in which baseline and monitor surveys are inverted separately and the 4D change is obtained by subtracting the two inverted models; this approach is straightforward to implement but tends to amplify non-repeatable noise and inversion artifacts [5]. Sequential difference inversion partially mitigates this issue by using the baseline inversion result as the starting model for the monitor inversion [6], thereby suppressing artifacts in

Check for updates

Academic Editor: John McBride

Received: 25 August 2026
Revised: 30 September 2026
Accepted: 15 October 2026
Published: 16 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

non-changing regions. A more widely adopted family of methods is double-difference inversion [7], in which the difference between the observed baseline and monitor data is directly inverted for the time-lapse model perturbation. By formulating the problem in the data and model difference domains, this strategy focuses the inversion on regions of true subsurface change. Simultaneous joint inversion further extends this idea by inverting baseline and monitor surveys jointly under a coupling regularization that explicitly enforces similarity in non-reservoir regions [8], while time-lapse full-waveform inversion (4D FWI) exploits the full wavefield to recover finer-scale velocity changes through repeated forward modeling and adjoint computation [9,10]. At the core of all these methods lies an explicit physical forward operator combined with an iterative optimization procedure that minimizes a misfit between observed and synthetic data, typically regularized by smoothness, sparsity, or geological priors and constrained by a provided low-frequency background model and an estimated source wavelet. While these methods generally provide improved resolution and physical consistency, they also exhibit several limitations. First, the reliance on repeatedly evaluating the forward operator and its adjoint makes large-scale 3D and 4D inversion extremely time-consuming and computationally demanding. Second, deterministic gradient-based schemes are highly sensitive to the choice of initial model, and inaccurate starting models can drive the optimization into local minima or cause cycle skipping [11]. Third, traditional time-lapse workflows usually require careful, problem-specific tuning of regularization parameters and wavelet estimation, which limits their portability and generalization across different time-lapse seismic data. Moreover, seismic inversion is inherently an ill-posed problem, as multiple models can produce similar seismic responses, and the available seismic data are band-limited and contaminated by noise. These limitations have motivated ongoing research into new strategies that aim to retain the physical meaning of model-based inversion while alleviating its computational burden.

In recent years, artificial intelligence (AI)-based methods have demonstrated promising performance in seismic inversion by leveraging data-driven learning to approximate the complex mapping between seismic data and subsurface properties [12]. Compared with traditional model-based approaches, these methods are capable of significantly improving computational efficiency once trained and have shown strong potential in handling large-scale seismic datasets. Broadly speaking, existing AI-based seismic inversion methods can be categorized into supervised and unsupervised frameworks, depending on whether labeled data are available during training. Early studies on supervised seismic inversion primarily focused on 1D problems, where individual seismic traces were treated independently [13–16]. In these approaches, neural networks were trained to learn a direct mapping from post-stack seismic traces to corresponding 1D acoustic impedance or velocity profiles. For example, Das et al. [17] proposed a convolutional neural network (CNN) to learn the mapping between seismograms and impedance. Through the careful construction of synthetic datasets that are consistent with realistic geological settings and sufficient training, the network demonstrates strong performance on both synthetic and real seismic data. In general, owing to the simplicity of the 1D formulation, these methods achieved fast inversion results. However, the neglect of lateral spatial correlations limited their ability to recover complex geological structures and led to discontinuous inversion results in 3D or 4D seismic inversion tasks. To overcome these limitations, Wu et al. [18] proposed a 2D CNN-based framework to improve the accuracy of seismic impedance prediction. During the training stage, the network takes seismic data together with a corresponding low-frequency impedance model as inputs, while the network output is compared against well-log-derived impedance profiles, which serve as incomplete labels. This training strategy effectively constitutes a weakly supervised learning scheme, as the well logs provide only sparse constraints in space. During the inference stage, the trained

network is directly applied to seismic data to generate 2D impedance prediction results. However, this approach relies on a reasonably good impedance model and the availability of well-log data, which restricts its applicability in many practical scenarios. Furthermore, motivated by the limited availability of labeled data required for supervised learning, a number of unsupervised seismic inversion approaches have been developed [19,20]. Deep learning has also been applied directly to 3D time-lapse CO_2 monitoring; for example, Celaya and Araya-Polo [21] jointly inverted time-lapse surface gravity and seismic data to recover 3D CO_2 plumes. Although the aforementioned methods have achieved promising results, the direct mapping from seismic data to impedance models often lacks explicit physical interpretability, making it difficult to ensure physical consistency and generalization beyond the training distribution. Consequently, several studies have sought to embed neural networks into traditional seismic inversion frameworks in order to enhance inversion performance while preserving physical interpretability. Representative examples include the use of Deep Image Prior (DIP) as a nonlinear preconditioner for solving post-stack seismic inverse problems, where the network architecture itself acts as an implicit regularization [22], as well as plug-and-play (PnP) prior-based methods, in which a pretrained denoiser is incorporated into iterative inversion schemes to impose prior information [23,24]. Notably, Romero et al. [25] introduced an inversion strategy based on implicit neural representations (INRs) for acoustic impedance estimation. In their framework, the acoustic impedance model is implicitly represented as a continuous function of spatial coordinates, which are fed into a multilayer perceptron (MLP). The inversion is carried out by iteratively updating the parameters of the MLP such that the predicted seismic response matches the observed seismic data. Applications to real field data have demonstrated the effectiveness and robustness of this approach. Alongside this line of research, another perspective naturally arises: can neural networks be employed to replace the conventional seismic modeling operator and subsequently leverage their differentiability to update an initial acoustic impedance model?

Fortunately, neural operators (NOs) have demonstrated remarkable success in replacing conventional numerical operators in recent years. By learning mappings from function spaces to function spaces, neural operators provide a powerful alternative to traditional physics-based operators and have been increasingly applied to geophysical problems. For example, Yang et al. [26] employed neural operators to approximate the forward operator of the 2D acoustic wave equation and achieved highly accurate forward modeling results. They further leveraged automatic differentiation to replace the adjoint-state method for full-waveform inversion, and experiments on synthetic datasets validated the feasibility of this approach as an alternative to conventional FWI. Subsequently, Zou et al. [27] extended this framework to 3D synthetic velocity model inversion and reported encouraging results. In addition, neural operators have also been employed to replace frequency-domain wavefield operators [28,29] and elastic wave propagation operators [30], enabling efficient surrogate modeling of wave-equation-based forward operators. Nevertheless, despite these advances, neural-operator-based inversion methods, particularly in 3D applications, remain largely untested on field data. This lack of practical validation has limited the broader adoption of neural operators in real-world geophysical applications.

Therefore, in this paper, we propose a neural-operator-based time-lapse inversion framework to address the challenges associated with time-lapse seismic inversion. First, we introduce the theoretical foundations of neural operators and describe the network architecture adopted in this study. Then, we provide an inversion workflow in which the neural operator is trained to learn the seismic modeling operator that maps impedance models to zero-offset (post-stack) seismic data. We stress the scope of this operator from the outset: the training data are generated with the normal-incidence convolutional model (Section 2.3), so the network

approximates a post-stack convolutional forward operator, not a wave-equation propagator. Its role in the inversion is that of a differentiable, mesh-independent surrogate for this operator that carries an implicit prior from its training distribution and does not require a wavelet to be estimated for each field vintage. The trained neural operator is then applied to time-lapse inversion of synthetic and field post-stack data.

2. Theory and Methods

In this section, we first review the fundamental concepts of post-stack seismic inversion and then introduce the neural operator architecture adopted in this study, followed by a description of the synthetic datasets used for training. Finally, we present the proposed time-lapse inversion framework.

2.1. Post-Stack Seismic Inversion

Post-stack seismic inversion aims to estimate subsurface acoustic impedance from stacked seismic data. In this context, the recorded seismic trace is commonly described by a convolutional model, in which the seismic response is expressed as the convolution of a source wavelet with the Earth's reflectivity series. This approximation provides a computationally efficient and physically interpretable framework for seismic inversion.

Let $I(x, y, z) = \rho(x, y, z)v(x, y, z)$ denote the acoustic impedance as a function in 3D of depth z and lateral position x and y, where ρ and v represent density and P-wave velocity, respectively. The corresponding reflectivity series $r(x, y, z)$ can be expressed as

$$r(x, y, z) = \frac{1}{2} \frac{\partial \ln I(x, y, z)}{\partial z}, \tag{1}$$

which is typically discretized using finite-difference operators in practice. Assuming a known seismic wavelet $w(t)$, the post-stack seismic data $d(x, y, t)$ can be approximated as

$$d(x, y, t) = w(t) * \hat{r}(x, y, t) + n(x, y, t), \tag{2}$$

where $\hat{r}(x, y, t) = r(x, y, 0.5v(x, y, t)t)$ and $v(x, y, t)$ is the time-depth conversion velocity (also known as vertical velocity). The operator $*$ denotes convolution and $n(x, y, t)$ represents additive noise.

Based on this forward model, post-stack seismic inversion is commonly formulated as an optimization problem that seeks an impedance model I minimizing the misfit between the observed post-stack seismic data d_{obs} and synthetic data generated by a forward modeling operator $\mathcal{Q}(\cdot)$ (Equations (1) and (2)):

$$\min_{I} \|\mathcal{Q}(I) - d_{\text{obs}}\|_2^2 + \lambda \mathcal{R}(I), \tag{3}$$

where $\mathcal{R}(I)$ denotes a regularization term, and λ controls the balance between data fidelity and regularization. Equation (1) is the small-contrast linearization of the exact normal-incidence reflection coefficient (Equation (6) below). With it, Equations (1) and (2) are linear in $\ln I$; this is the normal-incidence counterpart of the Born approximation used in linearized waveform inversion [9], and it is the operator used by the classical Tikhonov and double-difference baselines in Section 3.

2.2. Neural Operator Architecture

Neural operator learning provides a data-driven framework for approximating mappings between infinite-dimensional function spaces, rather than mappings between finite-dimensional vectors. Unlike conventional neural networks that learn discrete input–output relations tied to a fixed resolution, neural operators aim to learn resolution-invariant

operators that generalize across different discretizations of the same physical domain. Formally, let $\mathcal{A}$ and $\mathcal{U}$ denote function spaces. A neural operator seeks to approximate a nonlinear operator

$$\mathcal{G} : \mathcal{A} \to \mathcal{U}, \tag{4}$$

such that for an input function $a \in \mathcal{A}$, the output function $u = \mathcal{G}(a)$ represents the solution of an underlying physical process. In the context of post-stack inversion, this operator naturally corresponds to the seismic forward modeling operator that maps acoustic impedance models to seismic responses.

Among the various neural operator architectures, the Fourier Neural Operator (FNO) has emerged as a particularly effective approach for learning PDE-governed operators. The key idea of FNO is to parameterize the operator in the Fourier domain, enabling efficient modeling of global interactions through spectral convolution. Given an input function $a(x)$, the FNO performs iterative updates of the form

$$v_{k+1}(x) = \sigma\Big(Wv_k(x) + \mathcal{F}^{-1}(R \cdot \mathcal{F}(v_k)(\xi))\Big), \tag{5}$$

where $v_k(x)$ denotes the feature representation at layer k, W is a pointwise linear operator, $\mathcal{F}$ and $\mathcal{F}^{-1}$ represent the Fourier and inverse Fourier transforms, respectively, ξ is the wavenumber, R is a learnable spectral weight tensor restricted to a finite number of low-frequency modes, and $\sigma(\cdot)$ is a nonlinear activation function.

While the original FNO architecture effectively captures global dependencies, its performance may degrade when recovering fine-scale local structures, which are critical for seismic inversion. To address this limitation, we adopt the U-shaped neural operator (U-NO) architecture [31], shown in Figure 1. Inspired by U-Net-style encoder–decoder networks [32], U-NO integrates multi-scale feature extraction with skip connections to preserve both global and local information. Specifically, the U-NO consists of a down-sampling path that progressively increases the channel dimension while reducing spatial resolution, followed by an upsampling path that reconstructs high-resolution features. Skip connections are introduced between corresponding encoder and decoder layers to enhance the recovery of detailed structures. Each block G_l represents a Fourier layer that performs the calculation in Equation (5).

2.3. Training Dataset

The reflection coefficient series is computed from the acoustic impedance model using the standard impedance-contrast formulation, after which the post-stack seismic response is obtained via wavelet convolution. Specifically, we construct synthetic acoustic impedance models following Wu et al. [33], Ovcharenko et al. [34], and Ma et al. [35]. Since the Earth is predominantly characterized by flat layers, we design the synthetic dataset to exhibit laterally continuous, horizontally layered structures. To enhance geological realism and improve model generalization, we further incorporate representative structural complexities, including faults and folds. In summary, the training impedance model samples should include features we expect in our solution corresponding to the area of investigation. After building the impedance model samples, we compute the reflectivity and, from it, the seismic data that serve as labels. Figure 2 illustrates an example workflow for generating one synthetic training sample. After obtaining the acoustic impedance model, we compute the corresponding normal-incidence reflectivity series using the standard impedance-contrast formulation

$$r_i = \frac{I_{i+1} - I_i}{I_{i+1} + I_i}, \tag{6}$$

where I denotes the acoustic impedance and r_i is the reflection coefficient at the interface between the i-th and $(i+1)$-th layers. Given the reflectivity series, the associated zero-offset seismic data can be synthesized using a convolutional model

$$d(t) = (w * r)(t) + n(t), \tag{7}$$

where $d(t)$ is the zero-offset (post-stack) seismic trace, $w(t)$ is the source wavelet, $r(t)$ is the reflectivity series expressed in the time domain, $*$ denotes convolution, and $n(t)$ represents additive noise. In this study, we use a 25 Hz Ricker wavelet to generate the synthetic seismic responses. Each training sample is a $128 \times 128 \times 300$ volume. The training volumes are cropped from larger $550 \times 550 \times 500$ models with base layer thicknesses of 15–30 samples, 3–6 faults per model and a maximum dip factor of 2, and noise is added to the seismic data at signal-to-noise ratios of 7.5, 12.5 or 17.5 dB. In total, 883 samples are used for training and 220 for validation.

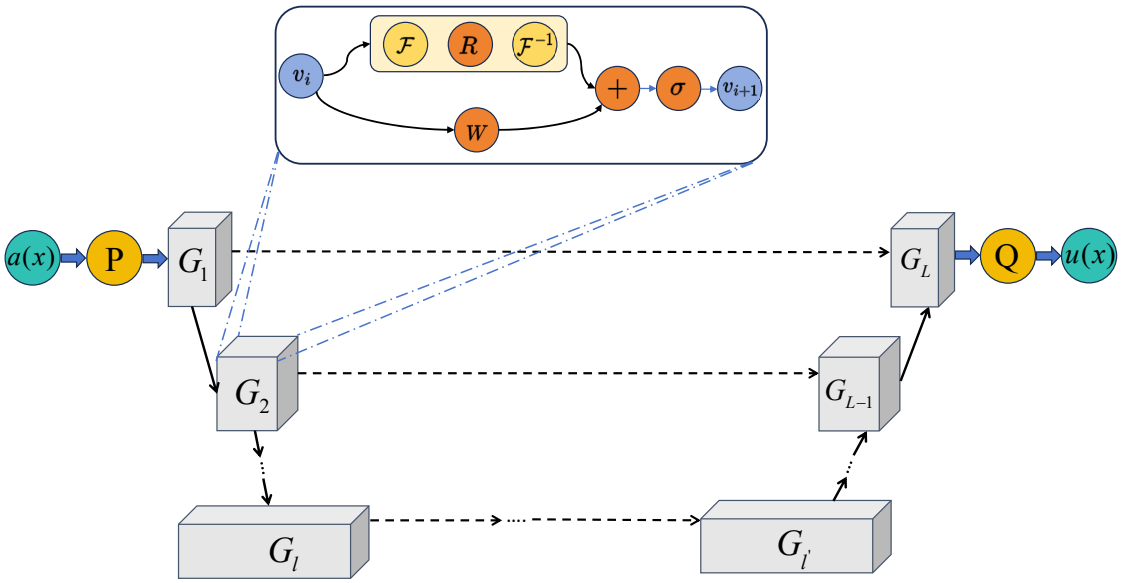


Figure 1. Architecture of the U-NO adopted in this study. The network follows a U-Net-style encoder–decoder backbone in which each block G_l is a Fourier layer implementing the spectral convolution defined in Equation (5). P and Q are the pointwise lifting and projection layers between the input function $a(x)$, the latent space and the output function $u(x)$. Solid arrows show the path through the encoder and the decoder, dashed horizontal arrows are skip connections, and dotted segments stand for omitted intermediate layers. The inset, linked by the blue dash-dotted lines, shows one Fourier layer (v_i and v_{i+1} correspond to v_k and v_{k+1} in Equation (5)). The colors only distinguish the types of operations.

2.4. Forward Modeling and Inversion

Instead of explicitly evaluating the conventional forward modeling operator $\mathcal{Q}$, we train a neural operator $\mathcal{G}_\theta$, parameterized by learnable parameters θ, to approximate it. Following Ma et al. [36], the network takes two input channels: a smooth reference impedance $\mathbf{I}_0$ and the impedance model $\mathbf{I}$ whose seismic response is sought,

$$\mathbf{d} \approx \mathcal{G}_\theta(\mathbf{I}_0, \mathbf{I}), \tag{8}$$

where $\mathcal{G}_\theta : \mathcal{M} \times \mathcal{M} \to \mathcal{D}$ maps a pair of impedance functions in the model space $\mathcal{M}$ to the seismic data function space $\mathcal{D}$ (the spaces $\mathcal{A}$ and $\mathcal{U}$ of Equation (4)). The three grid coordinates are appended to the input as additional channels, as is standard for FNO-type networks. During training, the first channel is a Gaussian-smoothed copy of the true model,

$\mathbf{I}_0 = S_\sigma \mathbf{I}_i$, where S_σ is a 3D Gaussian filter with a standard deviation of σ samples ($\sigma = 15$ for the operator used on Sleipner and $\sigma = 10$ for the Kimberlina fine-tuning), so that it mimics the smooth initial model available in practice. The operator is trained on synthetic pairs $\{(\mathbf{I}_i, \mathbf{d}_i)\}_{i=1}^{N}$ by minimizing (Figure 3a)

$$\min_{\theta} \sum_{i=1}^{N} \ell\big(\mathcal{G}_\theta(S_\sigma \mathbf{I}_i, \mathbf{I}_i),\, \mathbf{d}_i\big), \tag{9}$$

where the misfit ℓ is the sum of the mean squared error and the mean Euclidean distance between the predicted and reference data.

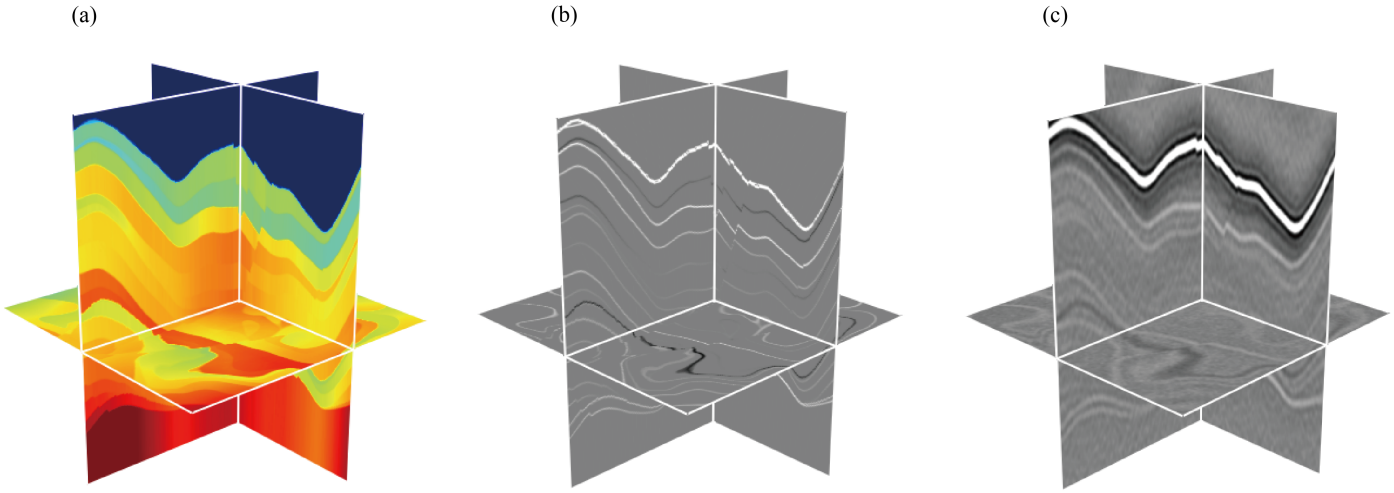


Figure 2. Workflow for generating one synthetic training sample. Starting from a horizontally layered acoustic impedance model with superimposed faults and folds (**a**), the normal-incidence reflectivity series is computed via the impedance-contrast formulation in Equation (6) (**b**), and the corresponding zero-offset post-stack seismic response is obtained by convolving the reflectivity with a 25 Hz Ricker wavelet following Equation (7) (**c**). In (**a**), the colors indicate acoustic impedance from low (blue) to high (red); (**b**,**c**) are displayed in gray scale.

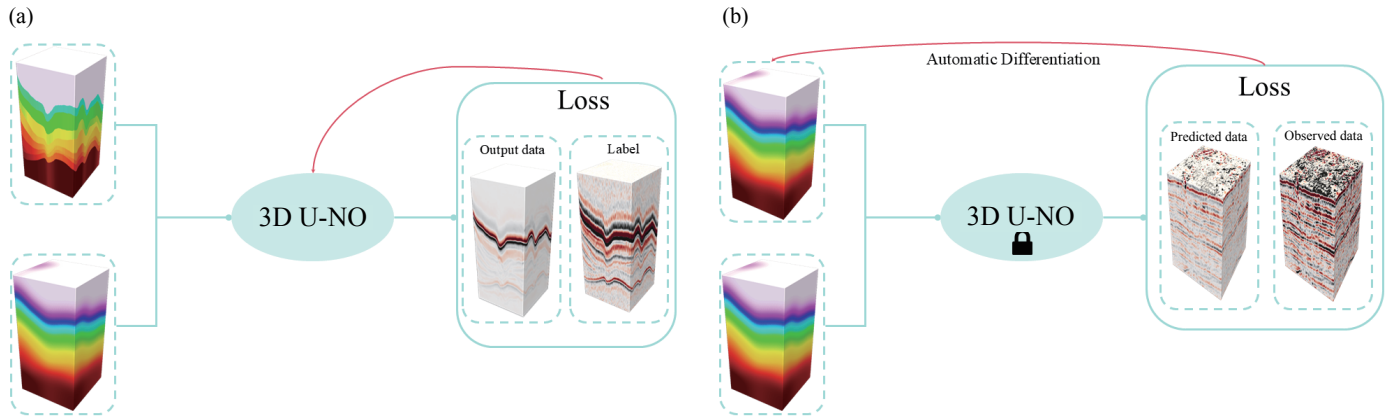


Figure 3. Schematic of the proposed neural-operator-based post-stack inversion framework. (**a**) Training stage: the neural operator $\mathcal{G}_\theta$ is trained on synthetic impedance–seismic pairs $\{(\mathbf{I}_i, \mathbf{d}_i)\}$ to learn the post-stack forward modeling operator. (**b**) Inversion stage: the trained neural operator is frozen and used as a differentiable surrogate for the forward operator; the acoustic impedance model is iteratively updated by minimizing the data misfit through gradient-based optimization with automatic differentiation. The colored volumes are illustrative examples of impedance models and seismic data; their colors carry no quantitative scale.

After training, θ is frozen and $\mathcal{G}_\theta$ serves as a differentiable surrogate for the forward operator (Figure 3b). Given an initial model $\mathbf{I}_{\text{init}}$, the first channel is fixed to $\mathbf{I}_{\text{init}}$ and only the second channel is updated:

$$\hat{\mathbf{I}} = \arg\min_{\mathbf{I}} \left\| \mathcal{G}_\theta(\mathbf{I}_{\text{init}}, \mathbf{I}) - g\,\mathbf{d}_{\text{obs}} \right\|_2^2 + \lambda_{\text{TV}}\,\text{TV}(\mathbf{I}), \qquad \mathbf{I}^{(0)} = \mathbf{I}_{\text{init}}, \tag{10}$$

where g is a scalar that brings the observed data to the amplitude level of the training data, TV is the isotropic 3D total variation, and λ_{TV} is its weight. Both channels therefore start from the initial model, and only the second one changes during the inversion. The fixed first channel plays the role it had during training: it tells the operator which smooth background the model in the second channel is a perturbation of. Because the network has only seen inputs whose second channel is a detailed version of the first, the fixed channel also anchors the update to the background trend; Section 3.3.1 compares this choice with two alternatives. The problem is solved with the Adam optimizer, and the gradient with respect to $\mathbf{I}$ is obtained by automatic differentiation, so no adjoint operator has to be derived. The same operator and the same settings (g, λ_{TV}, step size and number of iterations) are applied to every vintage of a survey, so the observed data can be the baseline or any monitor survey. The trained operator also carries an implicit prior: its predictions are reliable only for models that resemble its training distribution, which guides the solution toward plausible layered models and reduces the impact of noise.

3. Results

In this section, we first describe the details of the training stage. We then demonstrate the effectiveness of the proposed method for 3D and 4D seismic inversion using both synthetic and field data.

3.1. Training Details

During the training stage, a total of 883 3D synthetic samples are used to train the neural operator, while 220 samples are reserved for validation. The corresponding training and validation loss curves are shown in Figure 4. Although the training loss is noticeably lower than the validation loss, subsequent forward modeling tests indicate that no apparent overfitting is observed. This behavior suggests that the trained neural operator exhibits good generalization capability. The network is trained with the misfit of Equation (9), a learning rate of 1×10^{-3} and a batch size of 1 on a single NVIDIA A100 (80 GB) GPU, which takes about 0.63 h per epoch. The checkpoint with the lowest validation loss (epoch 23) is used in all experiments; it was reached after about 14.5 h of training. The operator has about 2.1×10^9 trainable parameters.

3.2. Forward Modeling Test

We first evaluate the forward modeling capability of the trained neural operator on a test sample. By comparing Figure 5b,c, we can observe that the trained neural operator produces accurate seismic data that closely match the reference data. This agreement is further confirmed by the corresponding difference shown in Figure 5d, where only minor discrepancies are observed, indicating that the neural operator provides a reliable approximation of the seismic modeling operator.

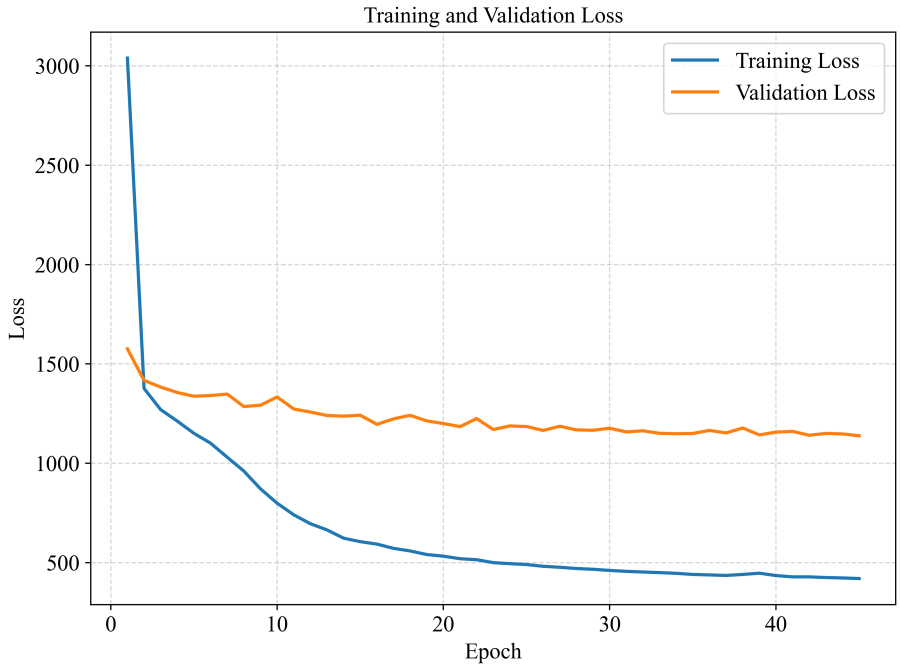


Figure 4. Training and validation loss curves of the U-NO over 45 training epochs. Although the validation loss remains slightly higher than the training loss, no apparent overfitting is observed in the subsequent forward modeling tests, indicating good generalization capability of the trained neural operator.

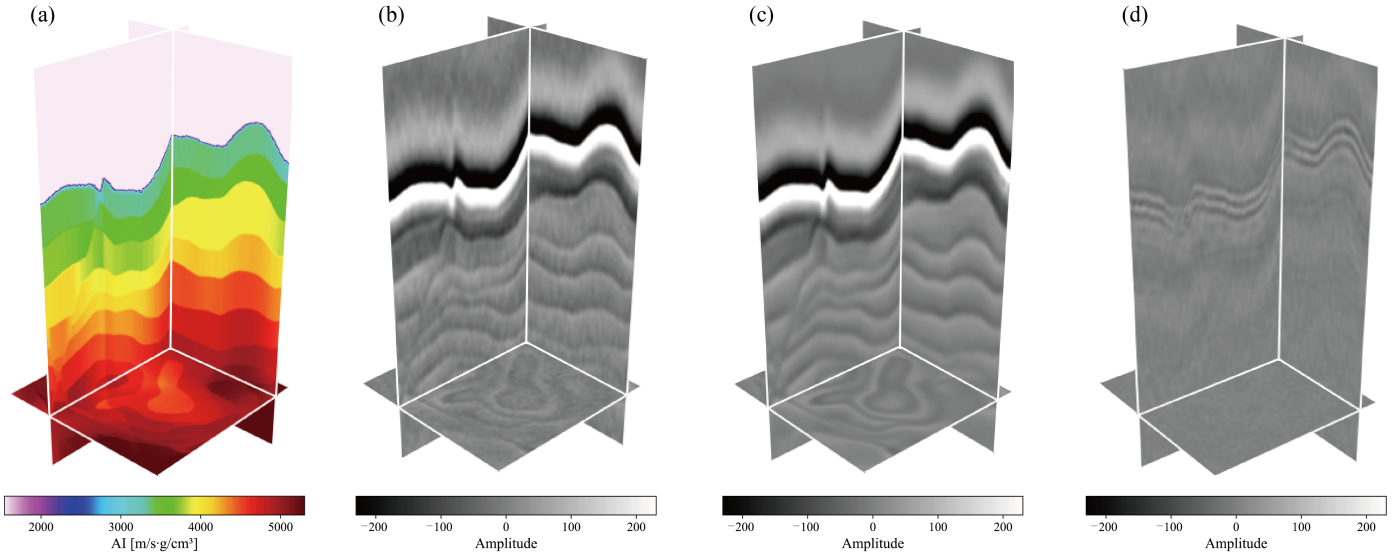


Figure 5. Forward modeling test of the trained neural operator on a synthetic test sample. (**a**) Acoustic impedance model. (**b**) Reference post-stack seismic data. (**c**) Post-stack seismic data predicted by the trained neural operator. (**d**) Difference between (**b**,**c**) plotted at the same scale. Only minor discrepancies are visible, confirming that the neural operator provides a reliable approximation of the post-stack forward modeling operator.

3.3. Time-Lapse Inversion Test (Synthetic Data)

Subsequently, we assess the capability of the neural operator to address time-lapse inversion problems. In this study, we adopt the Kimberlina synthetic dataset [37]. The Kimberlina reservoir model was developed based on comprehensive geological investigations of the southern San Joaquin Basin in California, integrating geological and hydrogeological data collected from numerous oil fields in the region. This dataset provides 200 years of

simulated CO_2 migration. The first 20 years correspond to the CO_2 injection phase, while the subsequent years represent the post-injection stage characterized by pressure-controlled CO_2 migration. In addition, this open access dataset includes P-wave velocity, S-wave velocity, and density models, which enable the rapid construction of acoustic impedance. The post-stack seismic data used for inversion are generated with PyLops [38] by convolving the reflectivity series with a 40 Hz Ricker wavelet. Because the neural operator was pre-trained on synthetic data generated with a 25 Hz Ricker wavelet, it has to be adapted to the 40 Hz wavelet of this data set. We fine-tune it on patches of $100 \times 100 \times 350$ samples extracted from the Year-0 models and their seismic responses (learning rate 10^{-4}, batch size 2, $\sigma = 10$ for the smooth channel, 50 epochs, about 2 h on three A100 GPUs). This cost is not negligible, and it shows that a trained operator does not transfer to data with a different source wavelet without adaptation (Section 4.2). After fine-tuning, the same operator is applied to all vintages without further retraining. The Year-20 and Year-200 surveys used for evaluation were not part of the fine-tuning set.

For each vintage, the initial model is obtained by smoothing the true impedance of that vintage with a 3D Gaussian filter ($\sigma = 15$ samples, i.e., 150 m; Figure 6a shows the one for Year 0), which removes the layering and keeps only the large-scale trend. The volume is inverted in 100×100 lateral patches (full depth, 20-sample overlap, Hann-window blending) with 100 Adam iterations and a step size of 5, $g = 1$ and $\lambda_{TV} = 0.7$; these settings are the same for all vintages.

Figure 6 shows the results. Each volume has $400 \times 400 \times 350$ samples (4 km $\times$ 4 km $\times$ 3.5 km), and one inversion takes about 15 min on four A100 GPUs. The inverted models (Figure 6e–g) add layering and lateral detail of the true models (Figure 6b–d) to the smooth initial models, and the inverted time-lapse differences (Figure 6h,i) reproduce the location, shape and lateral extent of the true CO_2-induced change (Figure 6j,k). Table 1 quantifies the accuracy of the inverted impedance. For each vintage, it lists the relative error, the root-mean-square error (RMSE), and the correlation coefficient with the true model over the whole volume and over the reservoir interval (the depth range in which the impedance changes over time), together with the same numbers for the initial model. For all three vintages, the inversion reduces the error of the initial model: the relative error decreases from 3.1–3.2% to 2.8–2.9% over the whole volume and from 4.2–4.4% to 3.8–4.0% in the reservoir, where the RMSE decreases by 8–10%, and the correlation with the true model increases from 0.978–0.979 to 0.982–0.983 (from 0.959–0.960 to 0.968 in the reservoir). The correlation between the model update and the true residual, $\mathbf{I}_{\text{true}} - \mathbf{I}_{\text{init}}$, is 0.53–0.58 over the whole volume and 0.58–0.63 in the reservoir, so the update points in the right direction. The improvement for Years 20 and 200, which were not part of the fine-tuning set, is as large as for Year 0.

Figure 7 compares the observed data with the data predicted by passing the final inverted models through the operator. The predicted data correlate with the observed data at $r = 0.97$ (normalized root-mean-square difference (NRMS) of 25–26%), and the predicted time-lapse data differences correlate with the observed ones at $r = 0.97$–0.98. The remaining differences are small; they show little lateral coherence, and their amplitude is largest at the strongest reflections. Figure 8 shows the data misfit during the inversion for every patch of the Year-20 and Year-200 inversions. The misfit decreases by about two orders of magnitude within 40–60 iterations and then levels off at about 2% of its initial value; the spread between patches is small, and the curves of the two vintages are nearly identical.

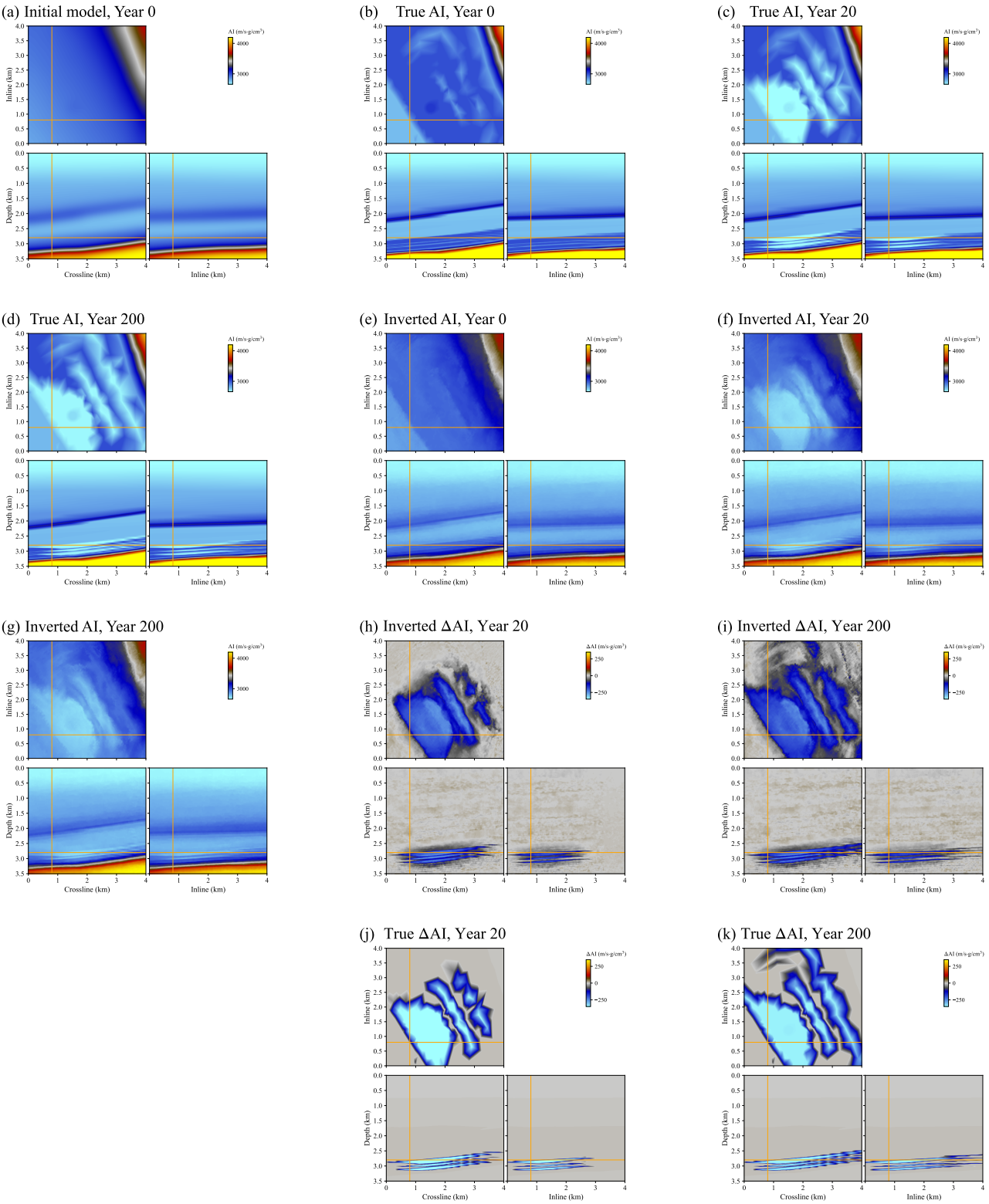


Figure 6. Time-lapse acoustic impedance inversion on the Kimberlina synthetic CO_2 data set. Each panel shows a depth slice (**top**) and vertical sections (**bottom**) along the orange lines; in the vertical sections, the orange lines mark the depth of the depth slice and the intersection of the two sections. (**a**) Initial model for Year 0, obtained by smoothing the true impedance with a 150 m Gaussian filter. (**b**–**d**) True impedance for Years 0, 20 and 200. (**e**–**g**) Inverted impedance for the same years. (**h**,**i**) Inverted and (**j**,**k**) true time-lapse changes relative to Year 0. Panels (**h**–**k**) share a symmetric color scale.

Table 1. Error metrics for the Kimberlina test: relative error $\|\hat{\mathbf{I}} - \mathbf{I}\|_2 / \|\mathbf{I}\|_2$, RMSE (m/s·g/cm^3) and Pearson correlation r between the inverted and true impedance for the inverted models and for the initial models. "Reservoir" is the depth interval in which the impedance changes over time. Year 0 was part of the fine-tuning set; Years 20 and 200 were not.

Vintage	Model	Whole Volume		Reservoir	
		Rel. Error/RMSE	r	Rel. Error/RMSE	r
Year 0	initial	3.1%/92	0.979	4.2%/130	0.959
	inverted	2.8%/85	0.983	3.8%/119	0.968
Year 20	initial	3.1%/95	0.978	4.3%/134	0.960
	inverted	2.9%/87	0.982	3.9%/122	0.968
Year 200	initial	3.2%/96	0.978	4.4%/136	0.960
	inverted	2.9%/88	0.982	4.0%/123	0.968

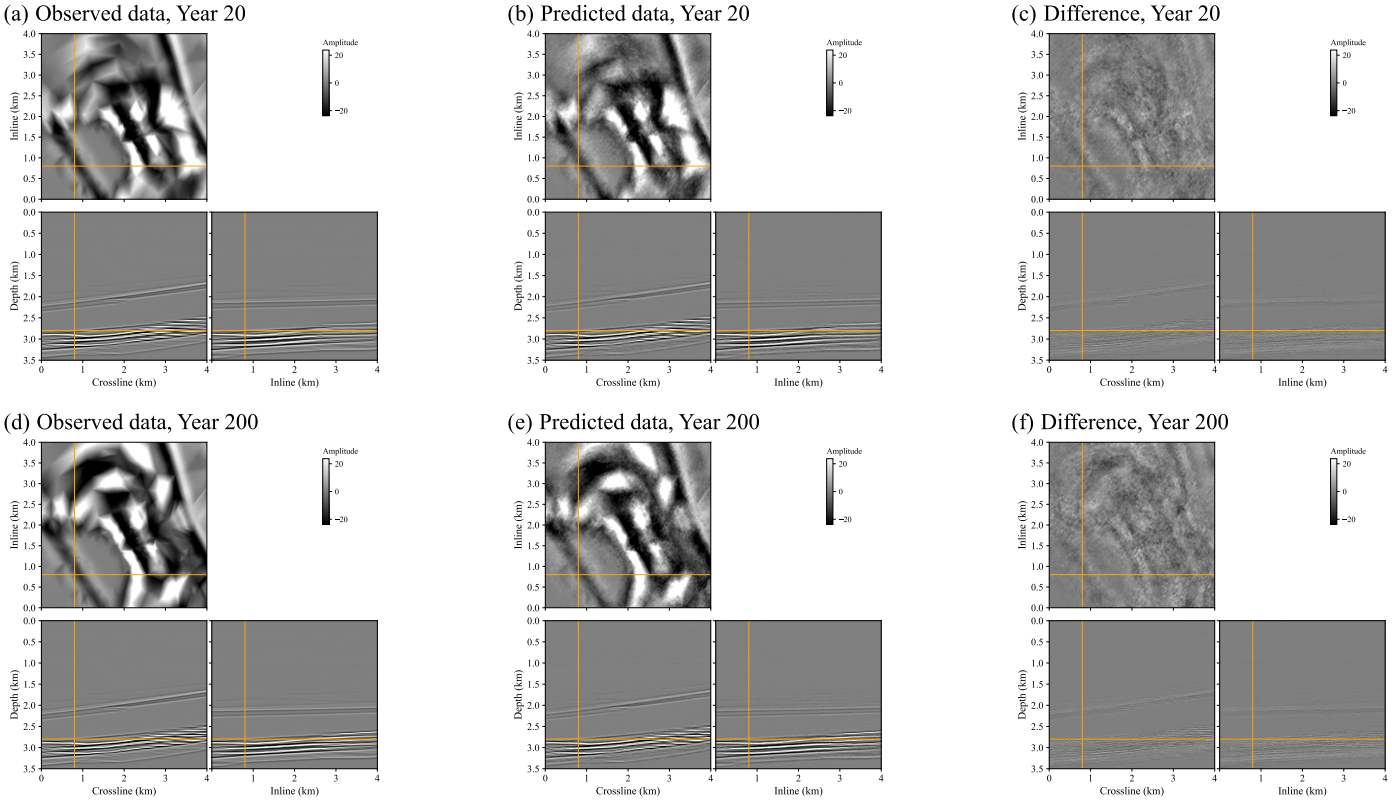


Figure 7. Data fit of the Kimberlina inversion for Year 20 (**top**) and Year 200 (**bottom**): (**a**,**d**) observed data; (**b**,**e**) data predicted by passing the final inverted models through the operator; (**c**,**f**) difference. All panels share the same color scale. The orange lines mark the slice positions, as in Figure 6.

3.3.1. Role of the Fixed Input Channel

We repeated the Year-20 inversion on the 2×2 patch block over the plume with two alternatives for the first channel: *tied*, where it is set to the model being updated, and *smoothed update*, where it is recomputed at every iteration as a Gaussian-smoothed copy ($\sigma = 10$) of the updated model. With the tied channel, the inversion fails: the misfit stalls at 22% of its initial value, and the model ends up less accurate than the initial model (Table 2). The smoothed update converges but gives a slightly less accurate model than the fixed channel. The first channel must therefore carry a smooth background that does not follow the update, and fixing it to the initial model works best.

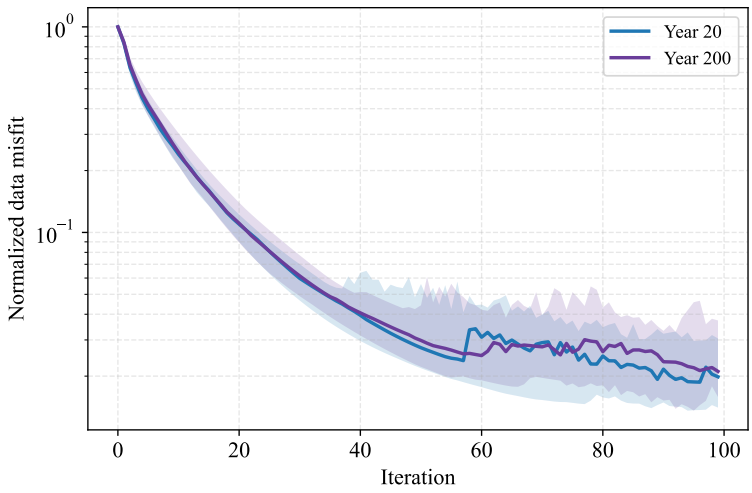


Figure 8. Data misfit versus iteration for the Kimberlina Year-20 and Year-200 inversions. The misfit of each patch is normalized by its initial value; the lines show the median over all patches, and the shaded areas show the 10th–90th percentile range.

Table 2. Ablation of the first input channel (Kimberlina, Year 20, 2×2 patch block over the plume). Relative error and correlation with the true model, correlation of the update with the true residual, and final data misfit relative to its initial value.

First Channel	Rel. Error	r	r, Update	Final Misfit
(initial model)	3.27%	0.976	–	–
Fixed (proposed)	3.01%	0.981	0.54	1.8%
Tied to the update	3.48%	0.974	0.13	22%
Smoothed update	3.10%	0.980	0.40	2.2%

3.4. Time-Lapse Inversion Test (Field Data)

The publicly available Sleipner 4D seismic data set comprises one baseline seismic survey acquired in 1994 and six subsequent monitor surveys acquired in 1999, 2001, 2004, 2006, 2008, and 2010. This extensive time-lapse seismic acquisition program was designed to continuously monitor the injection of CO_2 into the saline aquifer of the Middle Miocene/Early Pliocene Utsira Formation. The seismic surveys cover an area of approximately 4×7 km^2, with a record length of 2 s and a temporal sampling interval of 2 ms. Geologically, the data are characterized by predominantly subhorizontal stratification, intersected by strong reflection responses associated with the presence and migration of injected CO_2 within the primary reservoir. First, to focus on the CO_2-affected region, a localized subregion around the injection area was selected for inversion. The 1994 seismic survey is used as the baseline data for the proposed inversion framework, whereas the 2001, 2004, and 2006 surveys serve as the monitor data, and they are inverted using the same neural operator. The same 2D background acoustic impedance model, obtained from [25], is adopted in this case. The operator of Section 3.1 is used without fine-tuning. The volume is split into 100×100 patches in time and crossline (full inline extent, 20-sample overlap, Hann-window blending), and each patch is inverted with 100 Adam iterations, a step size of 50, $g = 3$ and $\lambda_{\mathrm{TV}} = 5$. These settings are identical for all four vintages. Each inversion takes about 18 min on one A100 GPU. The 3D visualization of the inversion results is presented in Figure 9. It is worth emphasizing that the initial model is constructed by smoothing well-log data and extending it along the two horizontal dimensions; it is therefore laterally homogeneous and contains no time-slice information, which poses a significant challenge for the inversion. Nevertheless, as shown in Figure 9, clear variations in acoustic impedance can be observed

over time, reflecting the effects of CO_2 migration. This behavior is further illustrated in Figure 10, where the imprints of CO_2 migration are particularly evident in the time-slice views, allowing the migration pathways to be clearly identified.

To demonstrate the reliability of the inversion, Figure 11 compares the observed and predicted seismic data for Year 1994 and Year 2001. As indicated by the residuals, although some signal loss is present, the overall agreement between the predicted and observed data remains strong. Here the predicted data are obtained by passing the final inverted impedance through the frozen operator. Over the whole volume, they correlate with the observed data at $r = 0.81$–0.83 for the four vintages (about zero for the initial model), and at the trace level they also reproduce the strong CO_2-related reflections of 2001 (Figure 12). More relevant for monitoring, the predicted 4D data differences correlate with the observed ones at $r = 0.74$–0.77 over the volume and at $r = 0.92$–0.93 inside the CO_2 plume mask defined below. These predictions are those of the learned operator itself (see Section 4).

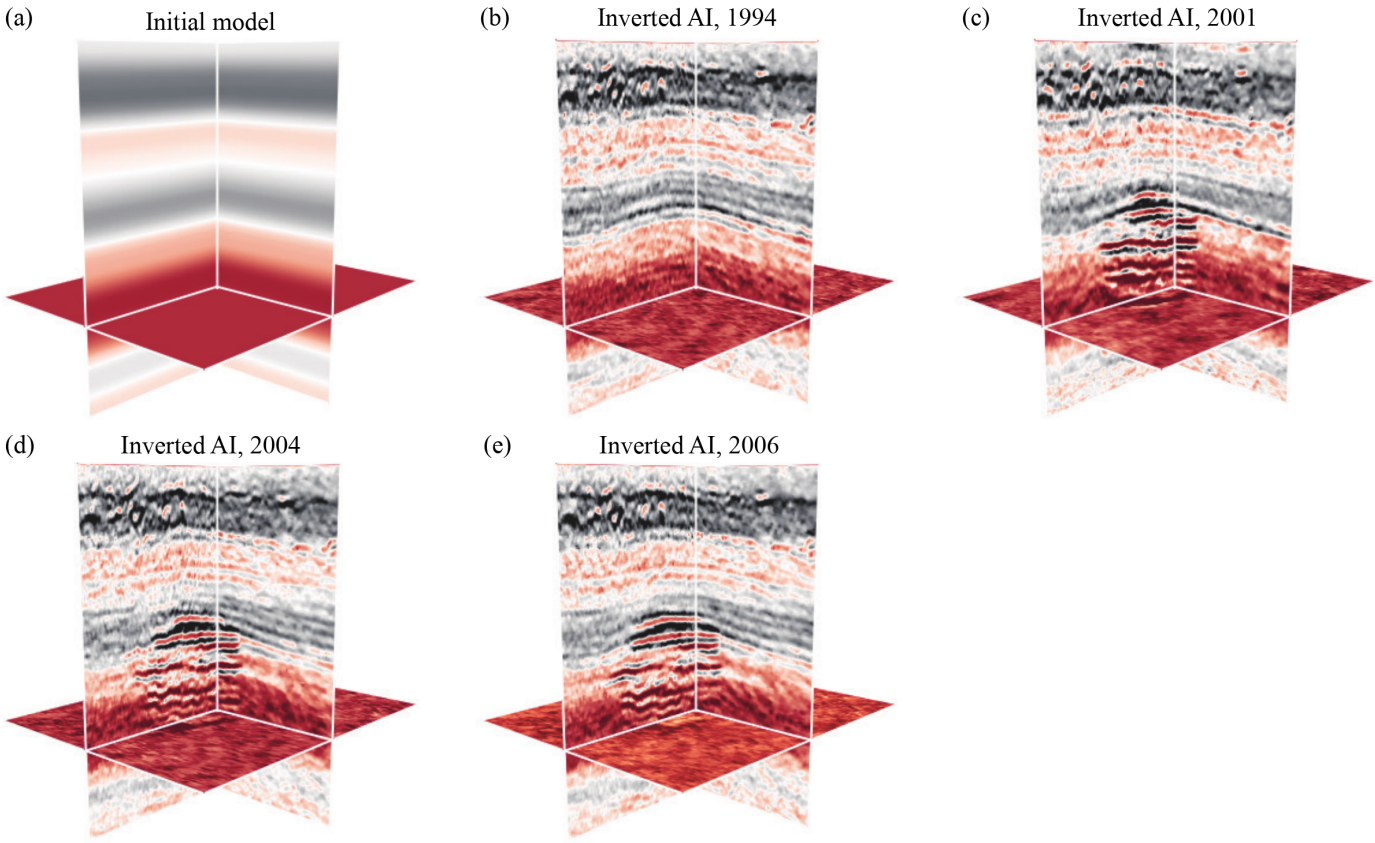


Figure 9. Three-dimensional visualization of the initial model and of the acoustic impedance volumes inverted from the Sleipner field data using the proposed neural operator: baseline survey (1994) and monitor surveys (2001, 2004 and 2006). The same trained neural operator, the same initial model, and the same inversion settings are used for all four vintages. Clear temporal variations of acoustic impedance are visible across surveys, reflecting the progressive migration of injected CO_2 within the Utsira reservoir. The colors show acoustic impedance with the same color map as in Figure 10 (red: low; black: high).

To quantify the time-lapse signal in the model domain, we use a plume mask derived from the observed data alone: the smoothed envelope of the observed 2001–1994 seismic difference, thresholded at four times the noise level of a shallow, CO_2-free window (0.81% of the voxels). Inside this mask, the median $|\Delta I|$ of the 2001, 2004 and 2006 inversions is 4.7–7.3 times that outside, and the mean change becomes increasingly negative (-70, -119

and -144 m/s·g/cm^3), whereas outside it stays within ± 5 m/s·g/cm^3. The impedance therefore changes where the data indicate CO_2 and remains stable elsewhere.

Figure 13 shows the data misfit during the 1994, 2001 and 2006 Sleipner inversions. It decreases more slowly than on Kimberlina and reaches 22–27% of its initial value after 100 iterations. The misfit is still decreasing at that point, but we kept the 100 iterations used for all Sleipner results because the field data contain noise and signal that a post-stack operator cannot explain. The spread between patches is small, and the curves of the three vintages are nearly identical.

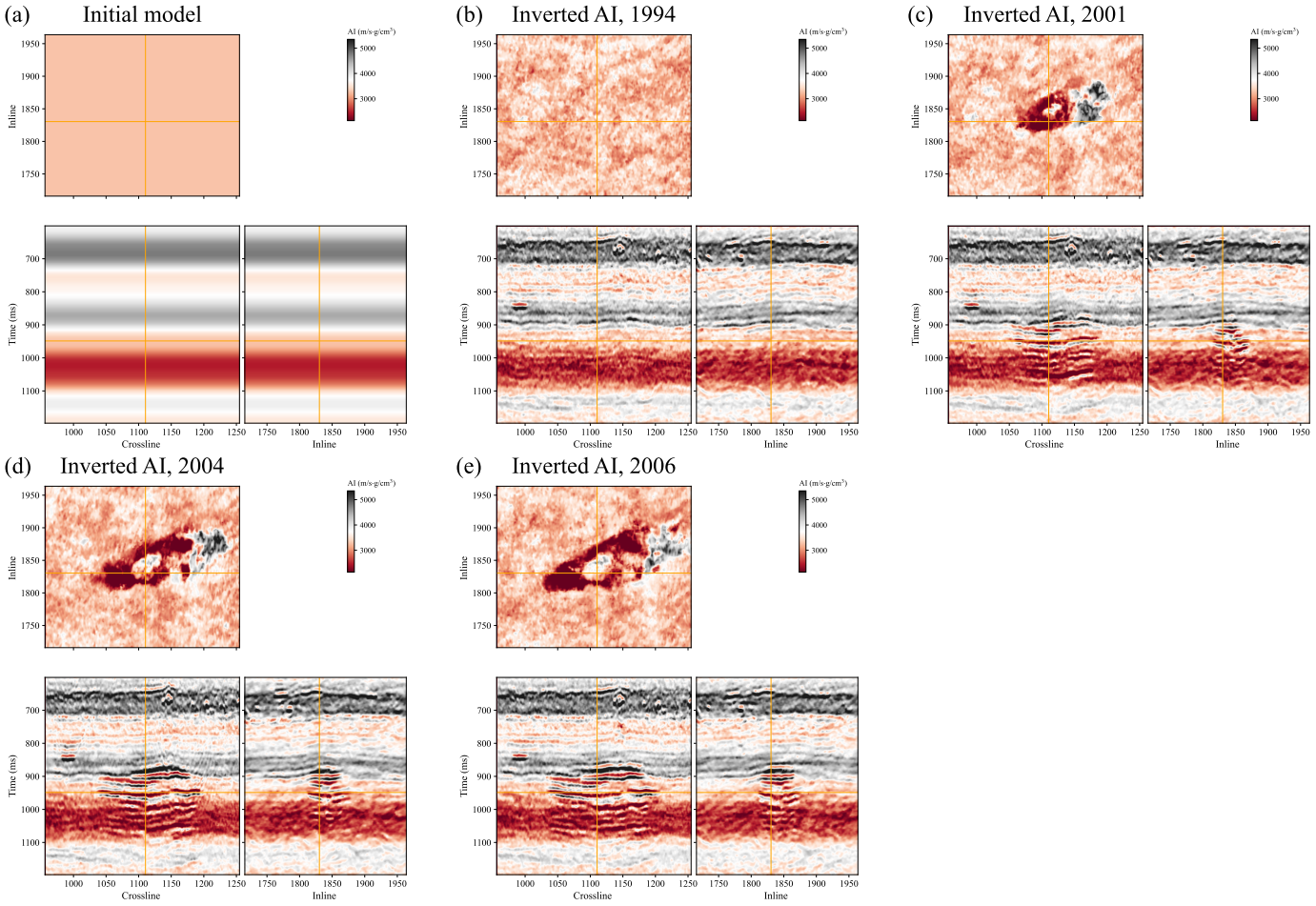


Figure 10. Slice views of the inverted acoustic impedance for the Sleipner data set at the baseline (1994) and monitor (2001, 2004, and 2006) surveys. Each panel shows a time slice at 950 ms (**top**) and vertical sections along the crossline and inline directions (**bottom**); the orange lines mark the slice positions, and the colors show acoustic impedance (red: low; black: high). The imprints of CO_2 migration are clearly identifiable on the time slices, and the lateral expansion of the low-impedance anomalies between successive vintages delineates the CO_2 migration pathways within the Utsira reservoir.

To contextualize the performance of the proposed neural operator, we compare it against a conventional 3D Tikhonov-regularized post-stack inversion applied to the same Sleipner datasets. The Tikhonov objective minimizes

$$\min_{\mathbf{m}} \; \|\mathbf{G}\mathbf{m} - \mathbf{d}\|_2^2 \; + \; \epsilon_{\text{id}}^2 \, \|\mathbf{m} - \mathbf{m}_{\text{back}}\|_2^2 \; + \; \epsilon_{\text{lap}}^2 \, \|\nabla^2 \mathbf{m}\|_2^2, \tag{11}$$

where $\mathbf{G}$ is the post-stack convolutional modeling operator, $\mathbf{m} = \ln I$ is the log-impedance model, $\mathbf{m}_{\text{back}}$ is the background model, ϵ_{id} controls proximity to the background (identity regularization), and ϵ_{lap} enforces spatial smoothness via a 3D Laplacian operator. The inversion is solved iteratively using conjugate gradients for 100 iterations.

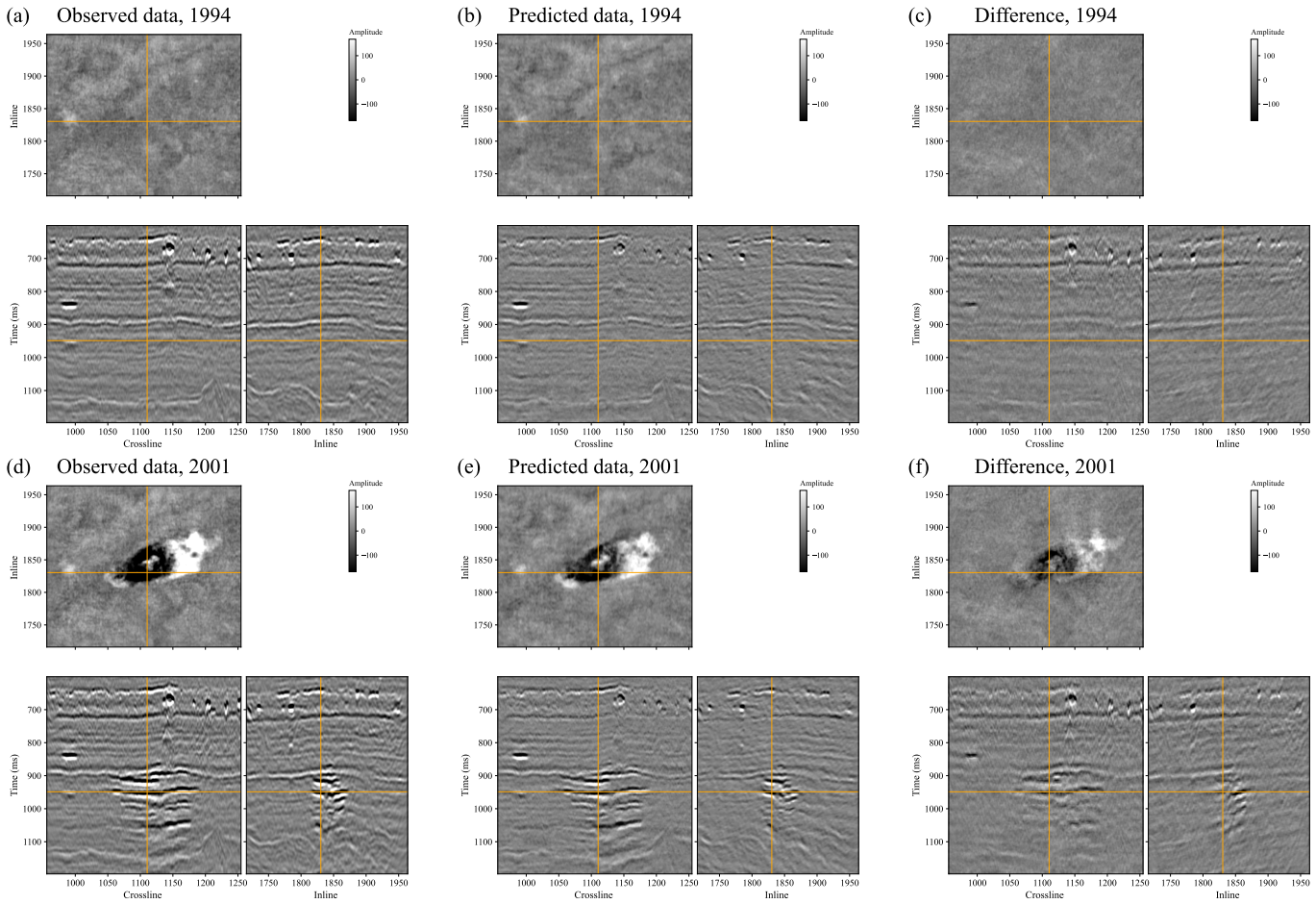


Figure 11. Comparison between observed and predicted post-stack seismic data on the Sleipner field for the 1994 baseline survey (**top row**) and the 2001 monitor survey (**bottom row**). For each year, the panels show the observed data (**a**,**d**), the data predicted by passing the final inverted impedance through the trained neural operator (**b**,**e**), and their difference (**c**,**f**). All panels share the same color scale. The orange lines mark the slice positions, as in Figure 10.

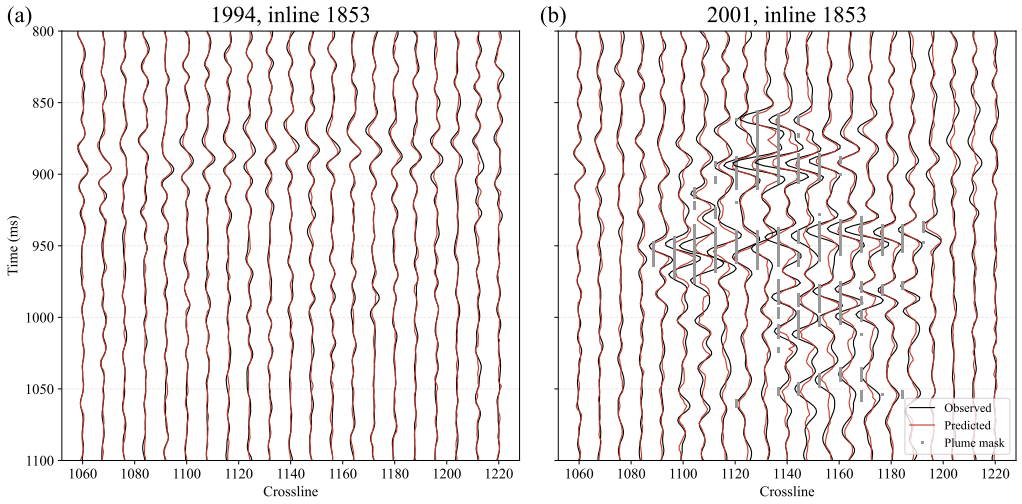


Figure 12. Observed (black) and predicted (red) post-stack traces along inline 1853, which crosses the CO_2 plume, for (**a**) 1994 and (**b**) 2001 (every eighth crossline, 800–1100 ms). The predicted traces are obtained by passing the final inverted impedance through the operator; both panels use the same display scale. Gray markers in (**b**) show where the plume mask intersects each trace.

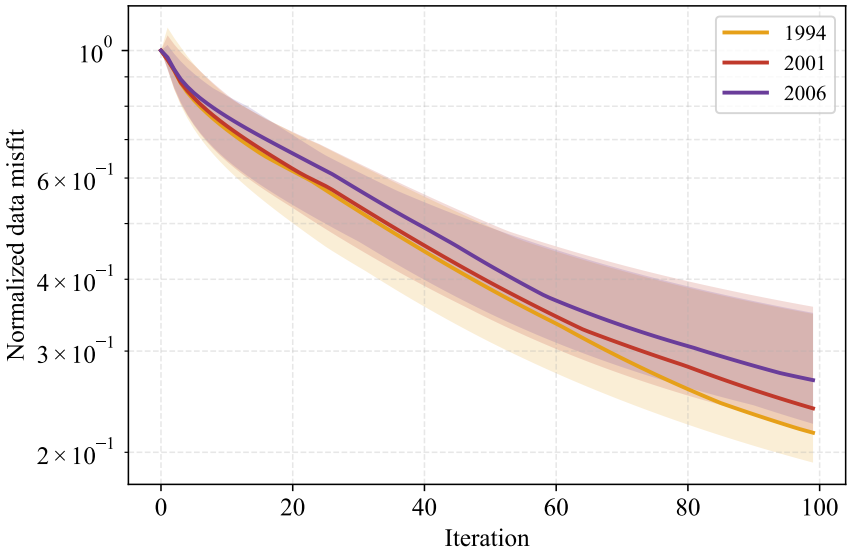


Figure 13. Data misfit versus iteration for the Sleipner inversions of the cropped volume (1994, 2001 and 2006). The misfit of each patch is normalized by its initial value; the lines show the median over all patches, and the shaded areas show the 10th–90th percentile range.

Traditional Tikhonov inversion faces an inherent trade-off between data fit and solution stability, directly controlled by ϵ_{id}. To demonstrate this trade-off, Figure 14 presents two Tikhonov inversions of the 2001 Sleipner monitor survey obtained with the same data and the same background model, but at opposite ends of the regularization spectrum: (a) a strong-regularization solution ($\epsilon_{\text{id}} = 1.5$) and (b) a weak-regularization solution ($\epsilon_{\text{id}} \to 0$). The strong-regularization result in Figure 14a yields an impedance model that is visually very similar to the one recovered by the proposed neural operator: the large-scale layering is preserved, and the impedance range is comparable, because the dominant background term anchors the solution close to $\mathbf{m}_{\text{back}}$. However, in Figure 15a, this seemingly plausible model fails to reproduce the observed seismic data: the predicted amplitudes are only a small fraction of the observations, and the residual retains most of the signal energy, revealing that the inversion is dominated by the regularization term rather than by the data. The weak-regularization result in Figure 14b sits at the opposite end of the trade-off: the data fit improves substantially (Figure 15b), but the impedance model now departs strongly from any geophysically reasonable shape, exhibiting amplified noise and unphysical extrema. No single value of ϵ_{id} can simultaneously deliver both a clean impedance model and an accurate data fit; reaching an acceptable balance would require careful, case-by-case tuning of ϵ_{id} (and ϵ_{lap}) for every survey year.

The proposed neural operator is less exposed to this trade-off. Because the learned operator produces reliable responses only for models that resemble its training distribution, it acts as an implicit regularizer, and a single set of inversion settings (g, λ_{TV}, step size and number of iterations) was sufficient for all four Sleipner vintages without per-survey tuning. The recovered impedance models remain geologically plausible while the predicted seismic responses match the observations (Figure 11). The inversion is not free of parameters: these settings and the smoothing of the initial model were chosen once for each data set, and different values were used for Kimberlina and Sleipner. The practical advantage is that they do not have to be re-tuned for every vintage.

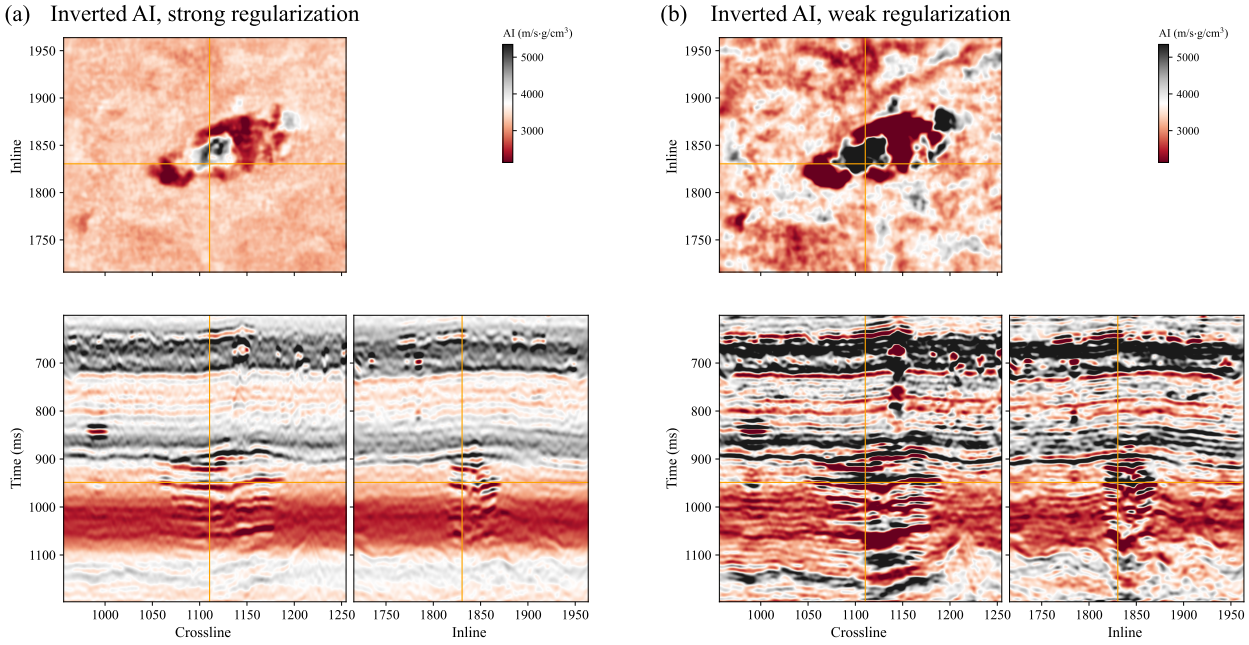


Figure 14. 3D Tikhonov acoustic impedance inversion of the 2001 Sleipner monitor survey under two different regularization settings (same data, same background model $\mathbf{m}_{\text{back}}$, $\epsilon_{\text{lap}} = 1.5$, 100 conjugate-gradient iterations in both cases): (**a**) strong regularization ($\epsilon_{\text{id}} = 1.5$); (**b**) weak regularization ($\epsilon_{\text{id}} \to 0$).

In addition, we exploit the mesh-independent property of the neural operator to perform inversion on the full-scale Sleipner dataset. Unlike conventional neural networks that are tied to a fixed grid resolution, neural operators learn mappings between function spaces and thus can generalize across discretizations with different grid sizes and resolutions. In principle, the entire dataset could be directly fed into the neural operator for inversion. However, due to GPU memory limitations, we adopt a patch-based inversion strategy using two A100 GPUs, and subsequently stitch all inverted acoustic impedance patches to obtain the final full volume result. Here, we present the inversion results for full-scale data from Year 1994, Year 2001, and Year 2006. As shown in Figure 16, although the neural operator is trained only on volumes of $128 \times 128 \times 300$ samples, it can be applied to field data on a considerably larger grid. Figure 17 shows two traces of the full-scale results, with the reservoir interval enlarged. Well 15/9-13 (inline 1773, crossline 1099) lies outside the CO_2 plume, so its pre-injection log is a valid reference for every vintage; the log is aligned to the seismic by a 10 ms bulk shift from the well tie. The inverted trace follows the trend of the log and adds band-limited detail, although individual thin beds are not resolved. At the well, the 1994, 2001 and 2006 inversions are nearly identical, whereas at a trace inside the plume (inline 1853, crossline 1140) they differ strongly over the plume interval. The full-scale inversion therefore separates a stable location from the CO_2-affected one, without retraining the operator for the larger grid.

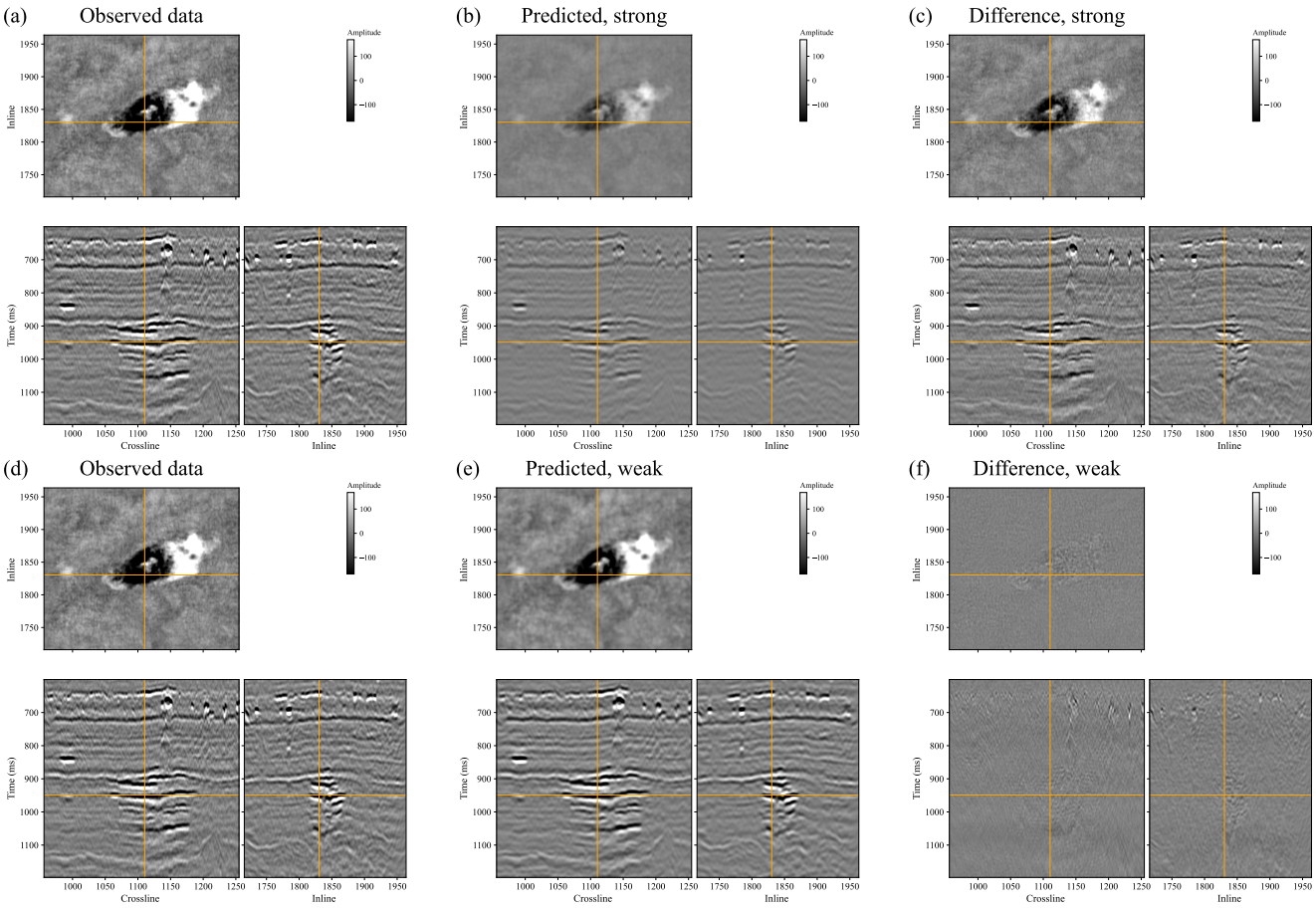


Figure 15. Seismic data fit of the two Tikhonov inversions in Figure 14: observed data (**a**,**d**), predicted data (**b**,**e**) and their difference (**c**,**f**), displayed as in Figure 11; the orange lines mark the slice positions. First row (**a**–**c**) corresponds to the strong regularization case ($\epsilon_{\mathrm{id}} = 1.5$); the predicted seismic data severely underestimate the observed amplitudes. Second row (**d**–**f**) corresponds to the weak-regularization case ($\epsilon_{\mathrm{id}} \to 0$); the data fit improves markedly, and the residual is much smaller, but the underlying impedance model in Figure 14b is geophysically unreasonable.

3.5. Comparison with Double-Difference Inversion

A standard baseline for time-lapse seismic inversion is the classical double-difference (DD) inversion. In its sequential, data-domain form, the baseline impedance is first recovered from the baseline survey

$$\mathbf{m}_b = \arg\min_{\mathbf{m}} \; \|\mathbf{G}_b\mathbf{m} - \mathbf{d}_b\|_2^2 + \epsilon_b^2 \, \|\nabla^2\mathbf{m}\|_2^2, \tag{12}$$

and the time-lapse change is then obtained by inverting the difference of the two surveys for a model perturbation

$$\Delta\mathbf{m} = \arg\min_{\delta\mathbf{m}} \; \left\|\mathbf{G}_m\delta\mathbf{m} - \left[\Delta\mathbf{d} - (\mathbf{G}_m - \mathbf{G}_b)\,\mathbf{m}_b\right]\right\|_2^2 + \epsilon_\Delta^2 \, \|\nabla^2\delta\mathbf{m}\|_2^2, \tag{13}$$

where $\Delta\mathbf{d} = \mathbf{d}_m - \mathbf{d}_b$ is the observed data difference and the monitor model is $\mathbf{m}_m = \mathbf{m}_b + \Delta\mathbf{m}$. As in Equation (11), $\mathbf{m} = \ln I$ is the log-impedance, so $\Delta\mathbf{m} = \ln(I_m/I_b)$ is a relative change, and the impedance volumes in Figure 18 are obtained as $\exp(\mathbf{m})$. Here $\mathbf{G}_b$ and $\mathbf{G}_m$ are the post-stack convolutional operators built from the 1994 and 2001 wavelets, $\mathbf{d}_b$ and $\mathbf{d}_m$ are the baseline and monitor surveys, and ∇^2 is the 3D Laplacian. We run 80 conjugate-gradient iterations for each stage. Equation (12) starts from the background model $\mathbf{m}_{\mathrm{back}}$ with $\epsilon_b = 1.5$, and Equation (13) starts from $\delta\mathbf{m} = \mathbf{0}$ with a larger weight,

$\epsilon_\Delta = 3.0$, because the 4D change is spatially confined. The cross term $(\mathbf{G}_m - \mathbf{G}_b)\,\mathbf{m}_b$ is zero when the two vintages share a wavelet, and Equation (13) then reduces to the simple-difference form. We keep it because the 1994 and 2001 wavelets differ. The double-difference formulation works on the data difference and solves directly for $\Delta\mathbf{m}$. The static part of the wavefield is common to both surveys, so it cancels in $\Delta\mathbf{d}$ and is never re-estimated, and the regularization on $\Delta\mathbf{m}$ keeps the update inside the regions that actually change. We applied this scheme to the 1994/2001 Sleipner surveys with the same crop and the same laterally homogeneous background model as the neural operator inversion. The operators $\mathbf{G}_b$ and $\mathbf{G}_m$ also need a wavelet for each vintage; the neural operator does not.

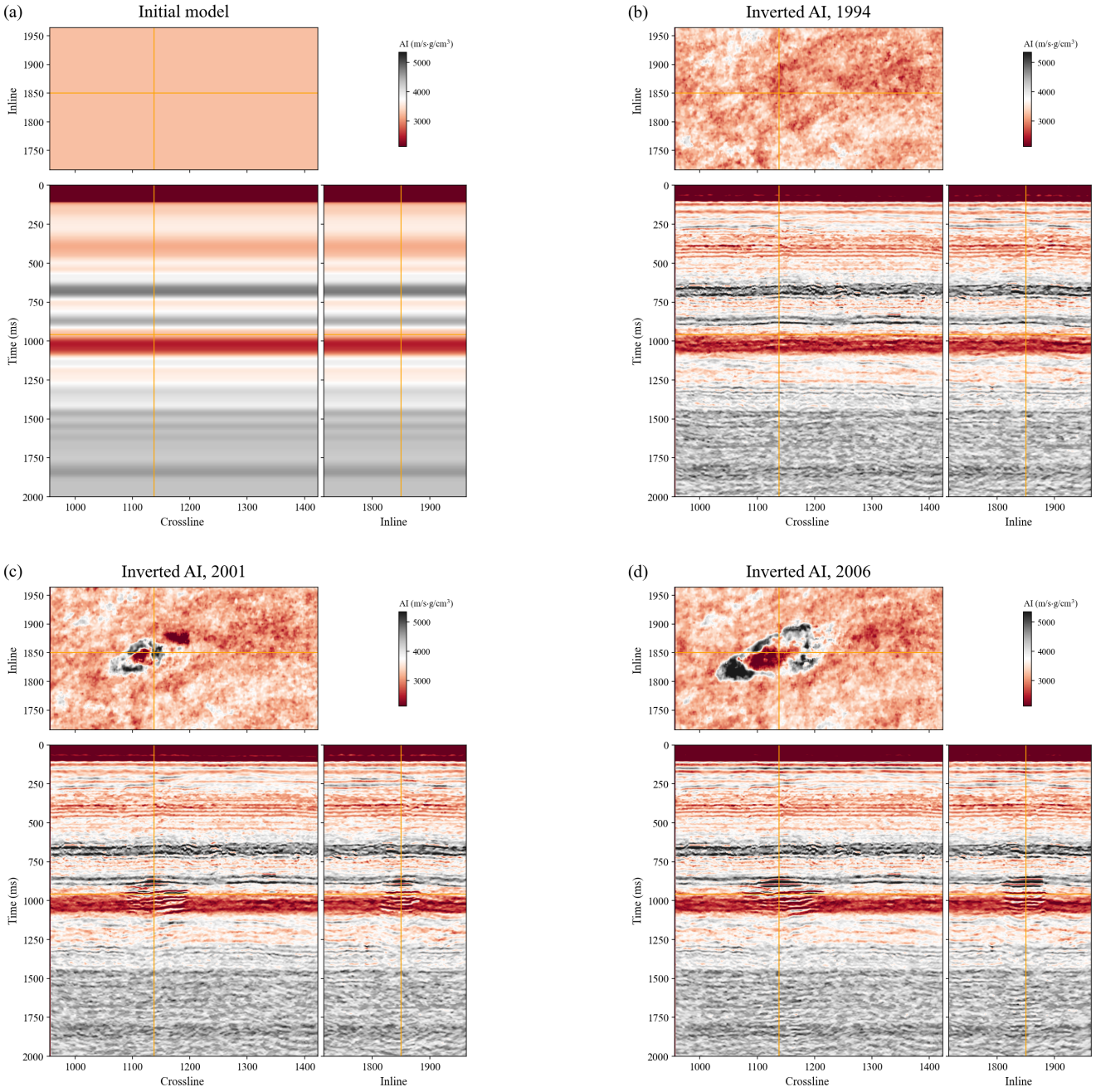


Figure 16. (**a**) Full-scale initial model used to verify the mesh-independent property of the neural operator. (**b**–**d**) Inversion results for the full-scale 1994, 2001, and 2006 surveys, respectively. The time slice is at 960 ms; at this time, the high-impedance rim around the plume is a side lobe of the thin CO_2-saturated layers. The colors show acoustic impedance (red: low; black: high), and the orange lines mark the slice positions.

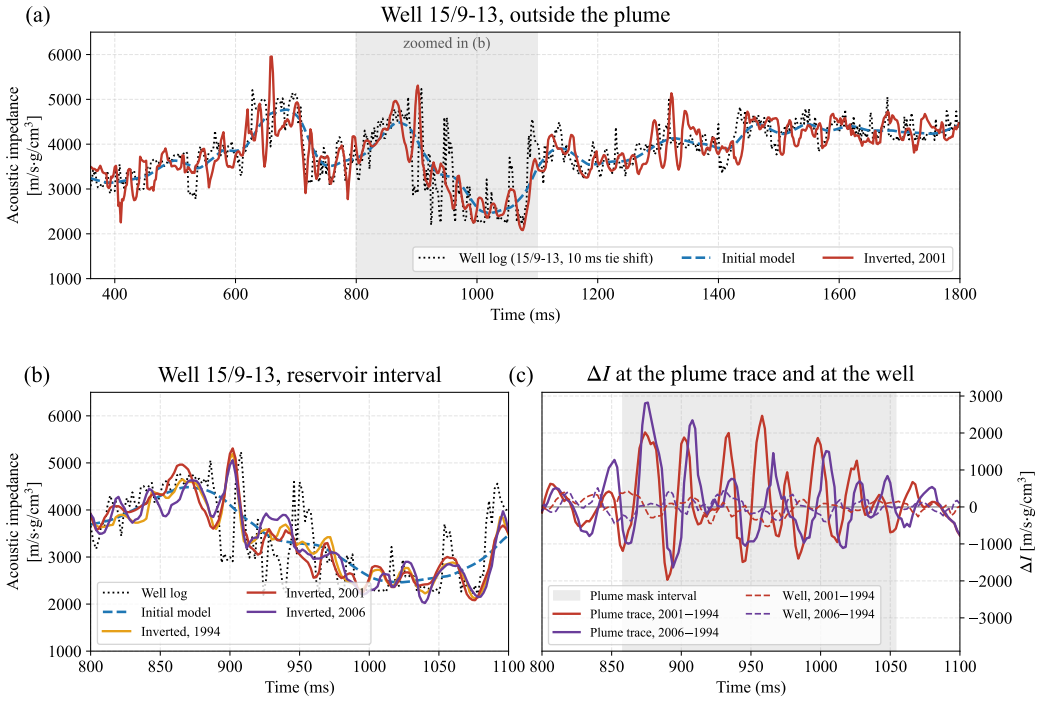


Figure 17. Full-scale neural operator inversion at a control location and inside the CO_2 plume. (**a**) Well 15/9-13, which lies outside the plume: well log (aligned to the seismic by a 10 ms bulk shift from the well tie), initial model and 2001 inversion; the shaded interval is enlarged in (**b**). (**b**) Reservoir interval at the well: the 1994, 2001 and 2006 inversions nearly coincide, as expected where no CO_2 is present. (**c**) Impedance change relative to 1994 at the plume trace (solid) and at the well (dashed); the shaded interval is where the plume mask derived from the observed 4D seismic difference intersects this trace.

Figure 18 compares, on two representative reservoir time slices (890 and 948 ms), the impedance recovered by the proposed neural operator, the impedance recovered by the classical DD inversion, and the corresponding observed 2001 seismic, with the objectively defined CO_2 plume outline overlaid. Because the initial model is laterally homogeneous and therefore carries no time-slice information, any coherent lateral structure appearing in a recovered time slice must originate from the data rather than from the starting model. The neural-operator time slices exhibit laterally coherent structure that closely follows the observed seismic: its median absolute spatial correlation with the observed seismic, computed over the reservoir time slices, is 0.54, reaching 0.78 and 0.74 on the two slices shown in Figure 18a,d. By contrast, the classical DD time slices remain close to the laterally homogeneous background except for a small localized anomaly, and their correlation with the observed seismic is only 0.08 (0.05 and 0.21 on the same two slices shown in Figure 18b,e). Under an identical homogeneous starting model and without any per-survey parameter tuning, the neural operator thus extracts substantially more spatially coherent, data-consistent information in the time-slice view.

We emphasize that, because no ground-truth impedance is available for the field data, this field comparison is qualitative and structural in nature; the quantitative accuracy of the proposed framework is established on the synthetic benchmarks with known ground truth (Figure 6). The two schemes also do not use the same forward physics. The neural operator is a learned mapping trained on synthetic impedance–seismic pairs, while the DD inversion uses an explicit convolutional operator. The network was trained to produce seismic-like imprints, so this may account for part of its higher correlation with the observed seismic. The correlation measures how much lateral information each scheme carries from the data into the model, not how accurate the impedance is. Within these bounds, the field

experiment shows that, from an identical laterally homogeneous starting model and with a single, fixed set of inference settings, the neural operator recovers laterally coherent, data-correlated impedance structure that the classical double-difference inversion does not, consistent with the fixed-settings behavior discussed above.

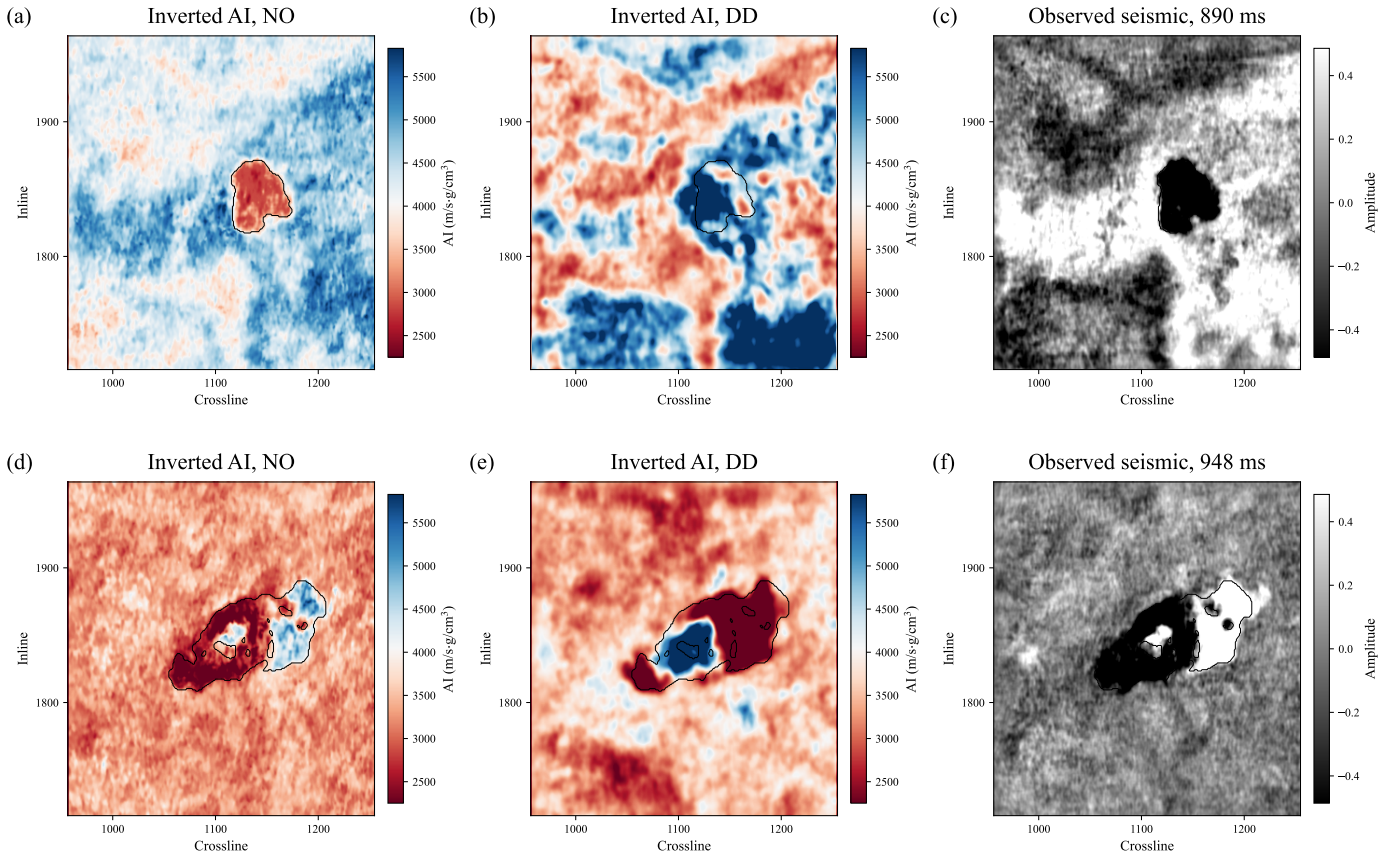


Figure 18. Time-slice comparison on the Sleipner 2001 monitor survey. Each of the two rows is a reservoir-level time slice (at 890 and 948 ms, map view); the columns show the impedance recovered by the proposed neural operator (**a**,**d**), the impedance recovered by the classical data-domain double-difference inversion (**b**,**e**), and the observed 2001 post-stack seismic (**c**,**f**). The black contour is the objectively defined CO_2 plume mask, and the same laterally homogeneous background model and identical crop are used for both inversions. The neural-operator time slices show laterally coherent structure that tracks the observed seismic.

4. Discussion

4.1. What the Neural Operator Approximates

The training data are generated with the normal-incidence convolutional model (Equations (6) and (7)). The neural operator therefore approximates a post-stack convolutional operator, not a wave-equation propagator: it does not model multiples, transmission losses, offset-dependent amplitudes or diffractions, and the method has the same assumptions as conventional post-stack inversion.

For small contrasts (Equation (1)), this operator is linear in $\ln I$ and is the normal-incidence counterpart of the Born approximation [9]. The Tikhonov and double-difference inversions in Section 3 use this linear operator. In wave-equation time-lapse inversion, Born and target-oriented formulations reduce the cost because the 4D change is small and local [6]. At the post-stack level, the operator is already linear and cheap, so the neural operator does not make a forward evaluation cheaper. What it provides is a differentiable operator with a prior learned from the training data, which needs no wavelet estimate

for each vintage and allows the same inversion settings for all vintages. Replacing an expensive pre-stack or wave-equation operator with a neural operator [26,27] is left for future work.

The predicted traces in Figure 11 are computed with the learned operator. If the inverted impedance is instead re-modeled with a convolutional operator and a wavelet estimated from the field data, the fit is poorer than that of a classical post-stack inversion, because the operator was trained with a 25 Hz Ricker wavelet and partly absorbs the difference to the field wavelet. The absolute impedance values and thin layers are therefore uncertain. The differences between vintages, obtained with the same operator and settings, are less affected, as indicated by the 4D data correlation of 0.92–0.93 inside the plume and the stable traces at the control well (Figure 17).

4.2. Computational Cost

Table 3 lists the cost of each stage. The main cost is the one-off training of the operator (about 14.5 GPU-hours). This cost is shared by all vintages: the same trained operator was used, without retraining, for the four Sleipner vintages and for both the cropped and the full-scale grid. For the Kimberlina data, whose 40 Hz wavelet differs from that of the training data, fine-tuning took about 6 GPU-hours; a similar step is needed for a new data set with a different wavelet or geology.

Once the operator is trained, a new vintage is inverted in 15–18 min for the Kimberlina and cropped Sleipner volumes, and because the patches are inverted independently, they can be distributed over several GPUs. The classical Tikhonov and double-difference inversions of the cropped volume run faster on a CPU because their convolutional operator is cheap. Their times in Table 3, however, do not include the steps these inversions need for each vintage: estimating a wavelet and tuning the regularization weights, to which the result is sensitive (Figures 14 and 15). The neural operator needs no wavelet estimate, and its settings are chosen once and kept fixed for all vintages, so a new survey can be inverted as soon as it is available and its result is directly comparable with the earlier ones. This advantage grows with the number of vintages, over which the training cost is also spread. A cost comparison with time-lapse FWI would not be like-for-like because FWI estimates velocity from pre-stack data with a wave-equation operator.

Table 3. Computational cost. GPU times refer to NVIDIA A100 (80 GB) cards; CPU times to a 40-core node.

Stage	Hardware	Wall Time	Needed
Operator training (to the selected checkpoint)	1 GPU	14.5 h	once
Fine-tuning to the 40 Hz Kimberlina wavelet	3 GPUs	≈2 h	per new data set
NO inversion, Kimberlina ($350 \times 400 \times 400$)	4 GPUs	≈15 min	per vintage
NO inversion, Sleipner crop ($300 \times 300 \times 249$)	1 GPU	≈18 min	per vintage
NO inversion, Sleipner full ($1001 \times 468 \times 249$)	2 GPUs	≈47 min (est.)	per vintage
Tikhonov, Sleipner crop, 100 iterations	CPU	16 min	per vintage
Double difference (L2), Sleipner crop, 1994/2001	CPU	13 min	per pair

4.3. Limitations

First, thin CO_2 layers are below the resolution of post-stack data, and the inverted time-lapse changes show side lobes of alternating sign (Figure 17c). The sign of ΔI at a single sample should therefore be read with care; the plume is better outlined with the envelope or with interval averages. Second, at the well the inversion follows the trend of the log but not its thin beds, and most of the low-frequency agreement comes from the initial model, which is built from the same log.

5. Conclusions

We propose a 3D neural operator framework for time-lapse post-stack impedance inversion. The operator is trained on synthetic impedance models and their convolutional seismic responses, with a smoothed copy of the model as an additional input channel. During inversion, the operator is frozen, the smooth input is fixed to the initial model, and the impedance is updated by automatic differentiation. On the Kimberlina synthetic data, the inversion reduces the error of the initial model for every vintage and places the CO_2-induced change at the location of the true change. On the Sleipner field data, the same settings were used for four vintages, without a wavelet estimate for each vintage. The inverted changes are concentrated inside a plume mask derived from the observed data, and the traces at a control well outside the plume are stable. The operator trained on $128 \times 128 \times 300$ volumes was applied to the full $1001 \times 468 \times 249$ Sleipner grid without retraining. The main cost is a one-off training stage, shared by all vintages, plus fine-tuning for a new source wavelet. Extending the approach to pre-stack or wave-equation operators is left for future work.

Author Contributions: Conceptualization, X.M. and T.A.; methodology, X.M. and T.A.; software, X.M.; validation, X.M.; formal analysis, X.M.; investigation, X.M.; data curation, X.M.; writing—original draft preparation, X.M.; writing—review and editing, X.M. and T.A.; visualization, X.M.; supervision, T.A.; project administration, T.A.; funding acquisition, T.A. All authors have read and agreed to the published version of the manuscript.

Funding: This publication is based on work supported by King Abdullah University of Science and Technology (KAUST).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data and accompanying codes that support the findings of this study are openly available at https://github.com/DeepWave-KAUST/Time_Lapse_inversion-dev (accessed on 16 October 2026). The Sleipner 4D seismic data set and the Kimberlina synthetic data set are publicly available from their respective providers.

Acknowledgments: The authors thank the DeepWave sponsors for their support, and KAUST for providing the computational resources used in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

4D	Four-dimensional (time-lapse)
AI	Artificial intelligence
CNN	Convolutional neural network
CPU	Central processing unit
DD	Double difference
DIP	Deep image prior
FNO	Fourier neural operator
FWI	Full-waveform inversion
GPU	Graphics processing unit
INR	Implicit neural representation
MLP	Multilayer perceptron
NO	Neural operator
NRMS	Normalized root-mean-square difference
PDE	Partial differential equation

PnP	Plug and play
RMSE	Root-mean-square error
TV	Total variation
U-NO	U-shaped neural operator

References

1. Veeken, P.C.H.; Da Silva, M. Seismic inversion methods and some of their constraints. *First Break* **2004**, *22*, 47–70.
2. Ma, W.; Lu, J.; Zheng, G.; Ahmed, S.F.; Liu, G. Machine learning fosters geophysical inversion for quantified geothermal parameters. *Geothermics* **2025**, *133*, 103477.
3. Leong, Z.X.; Zhu, T.; Sun, A.Y. Time-lapse seismic inversion for CO_2 saturation with SeisCO2Net: An application to Frio-II site. *Int. J. Greenh. Gas Control* **2024**, *132*, 104058.
4. Rosa, D.R.; Schiozer, D.J.; Davolio, A. Enhancing vertical resolution with 4D seismic inversion. *J. Pet. Sci. Eng.* **2022**, *212*, 110291.
5. Maharramov, M.; Biondi, B.L.; Meadows, M.A. Time-lapse inverse theory with applications. *Geophysics* **2016**, *81*, R485–R501.
6. Kamei, R.; Lumley, D. Full waveform inversion of repeating seismic events to estimate time-lapse velocity changes. *Geophys. J. Int.* **2017**, *209*, 1239–1264.
7. Bretaudeau, F.; Dubois, F.; Bissavetsy Kassa, S.; Coppo, N.; Wawrzyniak, P.; Darnet, M. Time-lapse resistivity imaging: CSEM-data 3-D double-difference inversion and application to the Reykjanes geothermal field. *Geophys. J. Int.* **2021**, *226*, 1764–1782.
8. Asnaashari, A.; Brossier, R.; Garambois, S.; Audebert, F.; Thore, P.; Virieux, J. Time-lapse seismic imaging using regularized full-waveform inversion with a prior model: Which strategy? *Geophys. Prospect.* **2015**, *63*, 78–98.
9. Tarantola, A. Inversion of seismic reflection data in the acoustic approximation. *Geophysics* **1984**, *49*, 1259–1266.
10. Martinez, R.; Vinje, V.; Moore, H.; Hollingworth, S.; Ringrose, P.; Stovas, A. Unraveling multilayer CO_2 plumes using the entire wavefield: Case study from the Sleipner storage site. *Interpretation* **2026**, *14*, T163–T183.
11. Choi, Y.; Alkhalifah, T. Source-independent time-domain waveform inversion using convolved wavefields: Application to the encoded multisource waveform inversion. *Geophysics* **2011**, *76*, R125–R134.
12. Li, M.; Yan, X.S.; Zhang, M.Z. A comprehensive review of seismic inversion based on neural networks. *Earth Sci. Inform.* **2023**, *16*, 2991–3021.
13. Mustafa, A.; Alfarraj, M.; AlRegib, G. Estimation of acoustic impedance from seismic data using temporal convolutional network. In *SEG Technical Program Expanded Abstracts 2019*; Society of Exploration Geophysicists: Tulsa, OK, USA, 2019; pp. 2554–2558.
14. Puzyrev, V.; Egorov, A.; Pirogova, A.; Elders, C.; Otto, C. Deep neural networks for 1D impedance inversion. *ASEG Ext. Abstr.* **2019**, *2019*, 1–4.
15. Wang, K.; Bandura, L.; Bevc, D.; Cheng, S.; DiSiena, J.; Halpert, A.; Osypov, K.; Power, B.; Xu, E. End-to-end deep neural network for seismic inversion. In *SEG Technical Program Expanded Abstracts 2019*; Society of Exploration Geophysicists: Tulsa, OK, USA, 2019; pp. 4982–4986.
16. Biswas, R.; Sen, M.K.; Das, V.; Mukerji, T. Prestack and poststack inversion using a physics-guided convolutional neural network. *Interpretation* **2019**, *7*, SE161–SE174.
17. Das, V.; Pollack, A.; Wollner, U.; Mukerji, T. Convolutional neural network for seismic impedance inversion. *Geophysics* **2019**, *84*, R869–R880.
18. Wu, X.; Yan, S.; Bi, Z.; Zhang, S.; Si, H. Deep learning for multidimensional seismic impedance inversion. *Geophysics* **2021**, *86*, R735–R745.
19. Feng, R.; Mejer Hansen, T.; Grana, D.; Balling, N. An unsupervised deep-learning method for porosity estimation based on poststack seismic data. *Geophysics* **2020**, *85*, M97–M105.
20. Lin, Y.; Theiler, J.; Wohlberg, B. Physics-guided data-driven seismic inversion: Recent progress and future opportunities in full-waveform inversion. *IEEE Signal Process. Mag.* **2023**, *40*, 115–133.
21. Celaya, A.; Araya-Polo, M. Joint Inversion of Time-Lapse Surface Gravity and Seismic Data for Monitoring of 3D CO_2 Plumes via Deep Learning. *arXiv* **2023**, arXiv:2310.04430. https://doi.org/10.48550/arXiv.2310.04430.
22. Lipari, V.; Picetti, F.; Bestagini, P.; Tubaro, S. Post-Stack Inversion with Uncertainty Estimation through Bayesian Deep Image Prior. In *Proceedings of the 82nd EAGE Annual Conference & Exhibition*; European Association of Geoscientists & Engineers: Bunnik, The Netherlands, 2021; Volume 2021, pp. 1–5.
23. Romero, J.; Corrales, M.; Luiken, N.; Ravasi, M. Plug and play post-stack seismic inversion with CNN-based denoisers. In *Proceedings of the Second EAGE Subsurface Intelligence Workshop*; European Association of Geoscientists & Engineers: Bunnik, The Netherlands, 2022; Volume 2022, pp. 1–5.
24. Izzatullah, M.; Alkhalifah, T.; Romero, J.; Corrales, M.; Luiken, N.; Ravasi, M. Plug-and-Play Stein variational gradient descent for Bayesian post-stack seismic inversion. In *Proceedings of the 84th EAGE Annual Conference & Exhibition*; European Association of Geoscientists & Engineers: Bunnik, The Netherlands, 2023; Volume 2023, pp. 1–5.

25. Romero, J.; Heidrich, W.; Luiken, N.; Ravasi, M. Seismic reservoir characterization with implicit neural representations. *J. Geophys. Res. Mach. Learn. Comput.* **2024**, *1*, e2024JH000275.
26. Yang, Y.; Gao, A.F.; Castellanos, J.C.; Ross, Z.E.; Azizzadenesheli, K.; Clayton, R.W. Seismic wave propagation and inversion with neural operators. *Seism. Rec.* **2021**, *1*, 126–134.
27. Zou, C.; Azizzadenesheli, K.; Ross, Z.E.; Clayton, R.W. Deep neural Helmholtz operators for 3-D elastic wave propagation and inversion. *Geophys. J. Int.* **2024**, *239*, 1469–1484.
28. Li, B.; Wang, H.; Feng, S.; Yang, X.; Lin, Y. Solving seismic wave equations on variable velocity models with Fourier neural operator. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4508618.
29. Huang, X.; Alkhalifah, T. Learned frequency-domain scattered wavefield solutions using neural operators. *Geophys. J. Int.* **2025**, *241*, 1467–1478.
30. Kong, Q.; Rodgers, A. *Feasibility of Using Fourier Neural Operators for 3D Elastic Seismic Simulations*; Technical Report; Lawrence Livermore National Laboratory (LLNL): Livermore, CA, USA, 2023.
31. Rahman, M.A.; Ross, Z.E.; Azizzadenesheli, K. U-NO: U-shaped neural operators. *arXiv* **2022**, arXiv:2204.11127.
32. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2015; pp. 234–241.
33. Wu, X.; Liang, L.; Shi, Y.; Fomel, S. FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. *Geophysics* **2019**, *84*, IM35–IM45.
34. Ovcharenko, O.; Kazei, V.; Alkhalifah, T.A.; Peter, D.B. Multi-task learning for low-frequency extrapolation and elastic model building from seismic data. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4510717.
35. Ma, X.; Wang, S.; Alkhalifah, T. High-resolution velocity model estimation with a hybrid neural operator and the time-shift imaging condition. *SSRN* **2026**. https://doi.org/10.2139/ssrn.6486770.
36. Ma, X.; Taufik, M.H.; Alkhalifah, T. Diffusion priors enhanced velocity model building from time-lag images using a neural operator. *J. Geophys. Res. Mach. Learn. Comput.* **2026**, *3*, e2025JH001223.
37. Alumbaugh, D.; Gasperikova, E.; Crandall, D.; Commer, M.; Feng, S.; Harbert, W.; Li, Y.; Lin, Y.; Samarasinghe, S. The Kimberlina synthetic multiphysics dataset for CO_2 monitoring investigations. *Geosci. Data J.* **2024**, *11*, 216–234.
38. Ravasi, M.; Vasconcelos, I. PyLops—A linear-operator Python library for scalable algebra and optimization. *SoftwareX* **2020**, *11*, 100361.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.