cigChannel: A ~~massive-scale~~ large-scale 3D seismic dataset with labeled paleochannels for advancing deep learning in seismic interpretation

Guangyu Wang[1, 2], Xinming Wu[1, 2], and Wen Zhang[1, 2]

[1]School of Earth and Space Sciences, University of Science and Technology of China, Hefei 230026, China
[2]Mengcheng National Geophysical Observatory, University of Science and Technology of China, Hefei 230026, China

Correspondence: Xinming Wu (xinmwu@ustc.edu.cn)

Abstract. Identifying ~~buried channels~~ paleochannels in 3D seismic volumes is essential for ~~characterizing hydrocarbon reservoirs~~ georesource development and offering insights into paleoclimate conditions~~, yet~~. However, it remains a labor-intensive and time-consuming task. ~~The data-driven deep learning methods are highly promising to automate the~~ Deep learning has shown great promise in automating seismic channel interpretation with high efficiency and accuracy, as ~~they have already achieved significant success~~ demonstrated in similar image segmentation tasks ~~within the field of~~ in computer vision (CV). ~~However~~Yet, unlike the CV domain, ~~the field of~~ seismic exploration lacks a comprehensive ~~benchmark dataset for channels, severely limiting~~ labeled dataset for paleochannels, significantly hindering the development, application, and evaluation of deep learning ~~approaches in seismic channel interpretation. Manually labeling~~ models in this field. Manual labeling of 3D ~~channels in field seismic volumes can be a~~ paleochannels is tedious and subjective~~work and most importantly, many field seismic volumes are proprietary and not accessible to most of the researchers. To overcome these limitations~~, potentially leading to mislabeling that degrades model performance. To address this, we propose a ~~comprehensive workflow of geological channel simulation and geophysical forward modeling to create a massive-scale~~ workflow to generate a synthetic seismic dataset~~containing~~, *cigChannel*, consisting of 1,~~200~~ 600 256×256×256 seismic volumes with ~~labels of more than~~ over 10,000 ~~diverse channels and their associated sedimentary facies~~labeled paleochannels. It is ~~by far the most comprehensive dataset for channel identification, providing realistic and~~ the largest dataset to date for seismic paleochannel interpretation, featuring geologically reasonable seismic volumes with ~~meandering~~accurately labeled meandering channels, tributary channel networks, ~~distributary,~~ and submarine ~~channels. Trained with this synthetic dataset, a~~ canyons. A convolutional neural network (simplified from ~~the~~ U-Net) ~~model performs well in identifying various types of channels in~~ trained on this dataset achieves F1 scores of 0.52, 0.73, and 0.63 in detecting meandering channels, tributary channel networks, and submarine canyons in three field seismic volumes, ~~which indicates the diversity and representativeness of~~ respectively. However, the synthetic seismic volumes in *cigChannel* still lack the variability and realism of field seismic data, potentially affecting the deep learning model's generalizability. To facilitate further research, we publicly release the ~~dataset. We have made the~~ dataset, ~~codes generating the data , and trained model publicly available for facilitating further research and validation of~~ data generation codes, and the trained model, aiming to advance deep learning approaches for seismic channel interpretation.

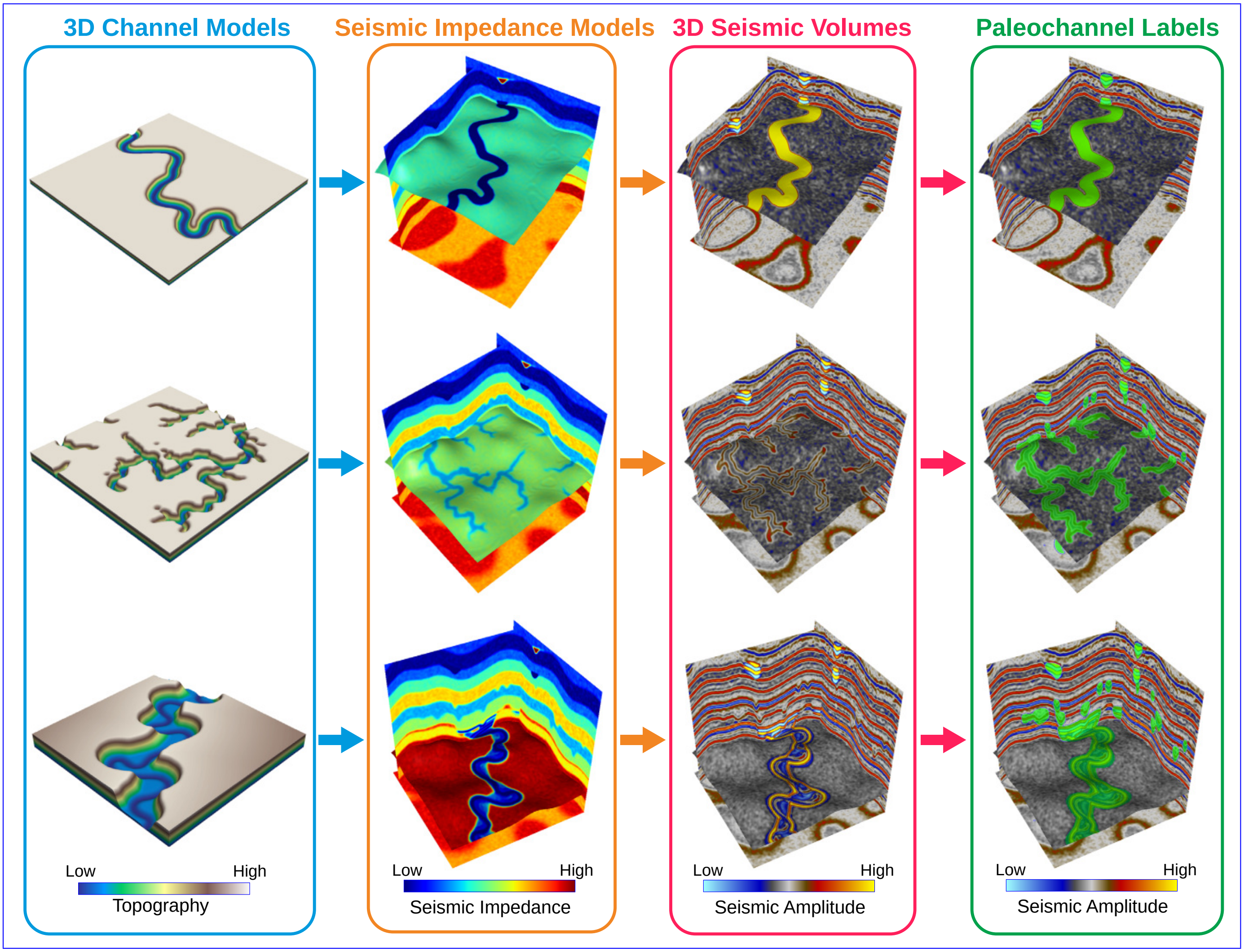


Figure 1. Workflow for generating the *cigChannel* dataset. First, we create ~~topographic~~ 3D models of three types of paleochannels: meandering channels, ~~distributary~~ tributary channel networks and submarine ~~channels~~canyons. Second, we build 3D seismic impedance models with ~~layered structure~~ multiple layers and place ~~the~~ these channels at layer boundaries as impedance anomalies. Third, the impedance models are used to calculate seismic reflection coefficients, which are subsequently convolved with ~~the~~ Ricker ~~wavelet~~ wavelets to create synthetic seismic volumes. Finally, seismic reflections of ~~the~~ paleochannels are automatically labeled. Note that both the channel models, seismic impedance models and seismic volumes are in depth domain.

1 Introduction

Paleochannels are buried river channels that have been preserved in the geological record. They can ~~serve as reservoirs for hydrocarbons (Clark and Pickering, 1996; Bridge et al., 2000; Hein and Cotterill, 2006) and~~ not only provide insights into

paleoclimate conditions ~~(Leigh and Feeney, 1995; Nordfjord et al., 2005; Sylvia and Galloway, 2006)~~(e.g., Leigh and Feeney, 1995; Nord
, but also serve as reservoirs for groundwater (e.g., Revil et al., 2005; Samadder et al., 2011), geothermal energy (e.g., Crooijmans et al., 20
, ore deposits (e.g., Heim et al., 2006; Oraby et al., 2019) and hydrocarbons (e.g., Clark and Pickering, 1996; Bridge et al., 2000) . Paleochannels can be identified in seismic volumes by their distinct shapes and sedimentary structures that differ from the surrounding rock formations. Although paleochannels are considered as geobodies, interpreters are limited to view them slice-by-slice in seismic volumes. This limitation significantly increases the complexity and time of interpreting paleochannel bodies in large seismic volumes. Moreover, the historical tectonic movement may introduce deformations such as foldings to the paleochannels, making them even more difficult to recognize.

To address those issues, automatic paleochannel identification methods based on 3D convolutional neural networks (CNNs) (Pham et al., 2019; Gao et al., 2021) ~~are~~ have been developed. The 3D CNNs are designed to capture volumetric features by performing 3D convolutions (Ji et al., 2012). They ~~treat paleochannels as bodies rather than slices as human interpreters typically see, which gives them the advantage in identifying paleochannels~~ have the advantage of handling paleochannels according to their 3D nature, as opposed to the slice-by-slice visual investigation of a human interpreter. This advantage is particularly significant when the paleochannels have been deformed by historical tectonic movements (e.g., folding and faulting), which disrupt their continuity and makes them more challenging to track in a slice-wise view. Another notable advantage is their efficiency. Once trained, the network can rapidly identify paleochannels in a large seismic volume. However, the main limitation of applying CNNs for paleochannel identification is the lack of labeled paleochannel samples for training. Unlike deep learning for computer vision, which benefits from numerous large datasets with labeled images such as ImageNet (Deng et al., 2009) and COCO (Lin et al., 2014), ~~currently~~ there is no publicly available dataset of field seismic volumes with labeled paleochannels. To create such a dataset, one needs to ~~have access to~~ access a large amount of field seismic volumes and correctly label the paleochannels. However, ~~many field seismic volumes are proprietary and not available to most of the researchers (Vizeu et al., 2022). Besides, the~~ labeling paleochannels can be challenging due to the complexity of field seismic volumes~~adds difficulty to correctly labeling the paleochannels~~, and human bias may introduce uncertainty to the labels (Bond et al., 2007). The label noise produced by mislabeling will deteriorate the performance of supervised learning (Pechenizkiy et al., 2006; Nettleton et al., 2010). Additionally, the labeling process will be time-consuming and labor-intensive.

While ~~training the networks with a large amount of labeled~~ creating a dataset by labeling paleochannels in field seismic volumes is ~~currently not an option~~expensive, an alternative solution is to use ~~the~~ synthetic seismic volumes, which are generated ~~by~~ through a series of simulation processes in order to mimic ~~the~~ field seismic volumes. Although lacking in sophisticated features, the synthetic seismic volumes are controllable, allowing us to tailor the ~~objectives that we want the network to learn~~ features that our network will learn to segment. Moreover, mislabeling can be avoided in synthetic seismic volumes since the locations of objectives are known during the simulation process. Synthetic seismic volumes have been proven effective as training data for networks to identify various objectives in field seismic volumes, such as faults (Wu et al., 2019; Zheng et al., 2019), seismic horizons (Bi et al., 2021; Vizeu et al., 2022), paleokarsts (Wu et al., 2020b; Zhang et al., 2024) and paleochannels (Pham et al., 2019; Gao et al., 2021). As for paleochannel identification, the synthetic seismic datasets created by Pham et al.

(2019) and Gao et al. (2021) only simulate ~~stacked and single~~ meandering channels, ~~respectively,~~ while the frequently observed ~~distributary (Payenberg and Lang, 2003; Li et al., 2016) and submarine (Deptuck et al., 2007; Gee et al., 2007) channels~~ tributary channel networks (e.g., Nordfjord et al., 2005; García et al., 2006; Darmadi et al., 2007) and submarine canyons (e.g. Deptuck et al., 2007; are not included. Considering the diversity of paleochannels in field seismic volumes, creating a dataset with various types of paleochannels is necessary for enhancing the networks' generalizability.

In this paper, we propose a ~~comprehensive~~ workflow (Figure 1) for generating ~~a massive-scale dataset of~~ synthetic seismic volumes ~~and labels of diverse paleochannels . In this workflow, we~~ with three types of paleochannels and their labels. We first build numerous 3D models of meandering ~~, distributary and submarine channelsfollowing the modeling methods developed by Howard and Knutson (1984), ? and Sylvester et al. (2011), respectively~~channels, tributary channel networks and submarine canyons. Parameters that control the modeling process are randomized within ~~a reasonable range~~ reasonable ranges in order to increase the diversity of channel models. Second, we build seismic impedance models with ~~layered structure~~ multiple layers and place the channels at layer boundaries as impedance anomalies. Third, the impedance models are used to calculate seismic reflection coefficients, which are subsequently convolved with ~~the Ricker wavelet~~ Ricker wavelets to create synthetic seismic volumes. Finally, channels in the seismic volume can be automatically labeled since their positions are already known. Using this workflow, we have created a ~~benchmark~~ dataset named *cigChannel* for deep learning-based seismic paleochannel interpretation. This dataset ~~, to our best knowledge, is by far the largest one that~~ contains 1,~~200~~ 600 256×256×256 seismic volumes and labels of more than 10,000 ~~diverse~~ paleochannels. The effectiveness of this dataset ~~is~~ has been validated by training a CNN to identify ~~various types of paleochannels in~~ meandering channels, tributary channel networks and submarine canyons in three field seismic volumes, respectively. It should be noted that although we have significantly improved the diversity of paleochannels compared with previous datasets, there is no guarantee that this dataset covers every form of paleochannel in field seismic volumes. Therefore, a Python package of the dataset generation workflow (https://github.com/wanggy-1/cigChannel, see Appendix C for ~~illustrative~~ demonstration codes) is also provided for customizing ~~the~~ paleochannels and facilitating further development.

2 Dataset generation workflow

In this section, we will ~~elaborate~~ outline the dataset generation workflow~~to explain details of the geological and geophysical modeling in generating the dataset. First, we will describe~~ , covering the steps for constructing 3D channel models and synthesizing seismic volumes. We will begin by describing the modeling process for meandering channels, tributary channel networks and submarine canyons. Following that, ~~distributary and submarine channels. Then,~~ we will explain how to ~~create synthetic seismic volumes~~ build seismic impedance models based on these channel models ~~, including designing folded impedance models with channels and simulating realistic~~ and use the impedance models to generate synthetic seismic volumes.

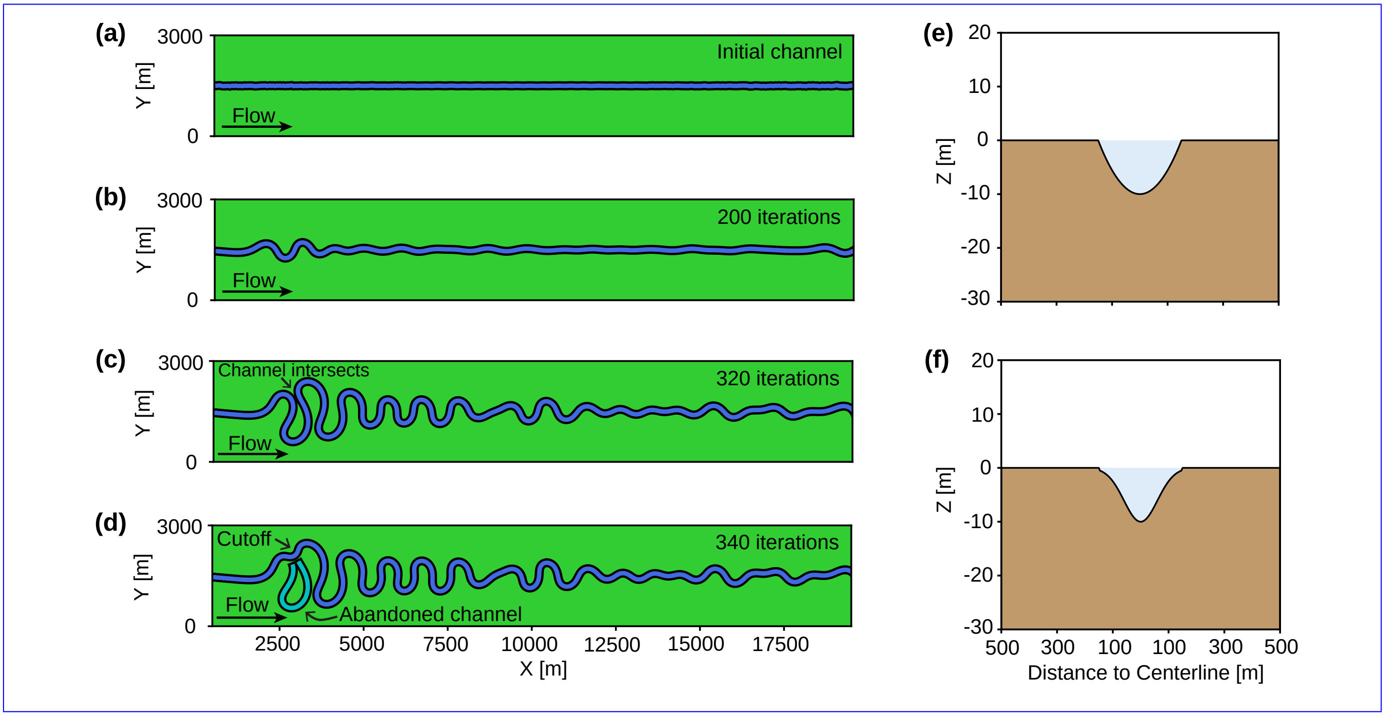


Figure 2. Meandering channel modeling process based on the open-source Python package *meanderpy* ~~(?)~~(Sylvester, 2021). First, we create (a) a straight channel with some minor perturbations. Then, (b) the channel begins to migrate, leading to the formation of multiple meanders. (c) The channel curvature increases as the migration continues, eventually causing a channel intersection, where (d) the channel cutoff will occur, forming the oxbow lake. Lastly, (e) the U- and (f) V-shaped channel cross-sections are used to define the channel topography.

2.1 Meandering channel modeling

Meandering channels are ~~among the most frequently observed river channels~~ a common type of river channels that can be found in many seismic volumes (e.g., Noah et al., 1992; Carter, 2003; Wood, 2007; Wang et al., 2012; Alqahtani et al., 2017). They are distinguished by their sinuous paths. The continuous interaction between water and the riverbed can lead to erosion on the outer bank and deposition on the inner bank, causing the channel to migrate over time and increasing its curvature. The key to create a realistic meandering channel is to simulate its migration. We use the open-source ~~Python~~ package *meanderpy* ~~(?)~~ (Sylvester, 2021) for this purpose, which employs a kinematic simulation method that computes the river migration rate as a weighted sum of upstream curvatures (Howard and Knutson, 1984; Sylvester et al., 2019). This simple kinematic model focuses on the influence of upstream curvatures on river migration and cannot capture complex processes such as compound meander formation without cutoffs (Frascati and Lanzoni, 2009). However, it remains sufficient for generating morphologically realistic meandering channels. The meandering channel simulation process is demonstrated in Figure 2. We start with a straight channel with some minor perturbations, which provide initial curvatures for channel migration (Figure 2a). The channel migrates over time and forms meanders at its upstream (Figure 2b). As the migration continues, curvature of the meander increases and eventually leads to channel intersection (Figure 2c), where the channel cutoff will occur~~and form the oxbow lake~~ , resulting in an abandoned channel (Figure 2d). The channel migration ends when it reaches the maximum number of iteration. We neglect the ~~oxbow lake~~ abandoned channel and extract the centerline from a random segment of the most recent meandering channel, which has to be long enough to span a 256×256 square grid with a cell size of 25 m after arbitrary rotation.

The centerline is randomly placed on the grid and rotated by a ~~random~~ certain angle between 0°and 360°. ~~We define the channel topography using the simplified~~ Since meandering channels in field seismic volumes typically exhibit U- ~~and~~ or V-shaped ~~channel~~ cross-sections ~~(~~(e.g., Zhuo et al., 2015; Alqahtani et al., 2017; Zeng et al., 2020; Manshor and Amir Hassan, 2023) , we use simplified U- or V-shaped profiles to define the channel topography, as shown in Figures 2e and 2f~~). The U-shaped channel is typically found in gentle terrain, formed mainly by lateral erosion. On the contrary, the V-shaped channel usually appears in areas with steep gradients, shaped primarily by vertical erosion. The~~ . The simplified U-shaped channel is defined as a parabolic function:

$$Z(x) = \begin{cases} 4D_c(x/W_c)^2 - D_c, & x \leq W_c \\ 0, & x > W_c \end{cases}, \tag{1}$$

where x is the Euclidean distance from the centerline to any point on the grid, D_c is the maximum depth of the channel (which will be denoted as channel depth hereafter for simplicity) and W_c is the channel width. The simplified V-shaped channel is defined as a combination of Gaussian and parabolic functions:

$$Z(x) = \begin{cases} \min\left[p(x), g(x)\right], & x \leq W_c \\ 0, & x > W_c \end{cases}, \tag{2}$$

where $p(x)$ is the parabolic function in Equation (1) and

$$g(x) = -D_c e^{-\frac{x^2}{2(W_c/4)^2}}. \tag{3}$$

Although these simplified channel cross-sections may not precisely represent the real ones, they can capture ~~the~~ their main features at a low computational cost. We create diverse topographic models of the meandering channel by randomizing the modeling parameters within ~~a reasonable range~~ reasonable ranges (see Table A1). Some examples are demonstrated in Figure 5a, showing various meandering channels with different widths, depths and meander wavelengths.

2.2 ~~Distributary channel modeling~~

It should be noted that in this study, we focus on identifying the most recent meandering channels in their migration histories. Therefore, all the meandering channel models only include the last channel form of the migration process. The corresponding sedimentary facies formed during the channel migration process, such as point bars, natural levees and abandoned channels (or oxbow lakes), are not included. It is also worth noting that the width and maximum depth of each channel are fixed, while in nature they generally exhibit certain degree of variability.

~~Distributary channel modeling process based on the open-source C++ package *soillib* (?). First, we generate (a) a map of normalized water discharge using the *soillib* package. Second, we create (b) the river mask by binarizing the normalized water discharge with a threshold value of 0.4, where values greater than this threshold are considered as rivers. Third, we compute (c) the Euclidean distance to rivers and (d) the normalized width of the nearest river, which are subsequently used as parameters in a parabolic function to define (e) the channel topography. Finally, to avoid abrupt topographic shifts, a Gaussian filter is applied to create (f) a smoothed channel topography.~~

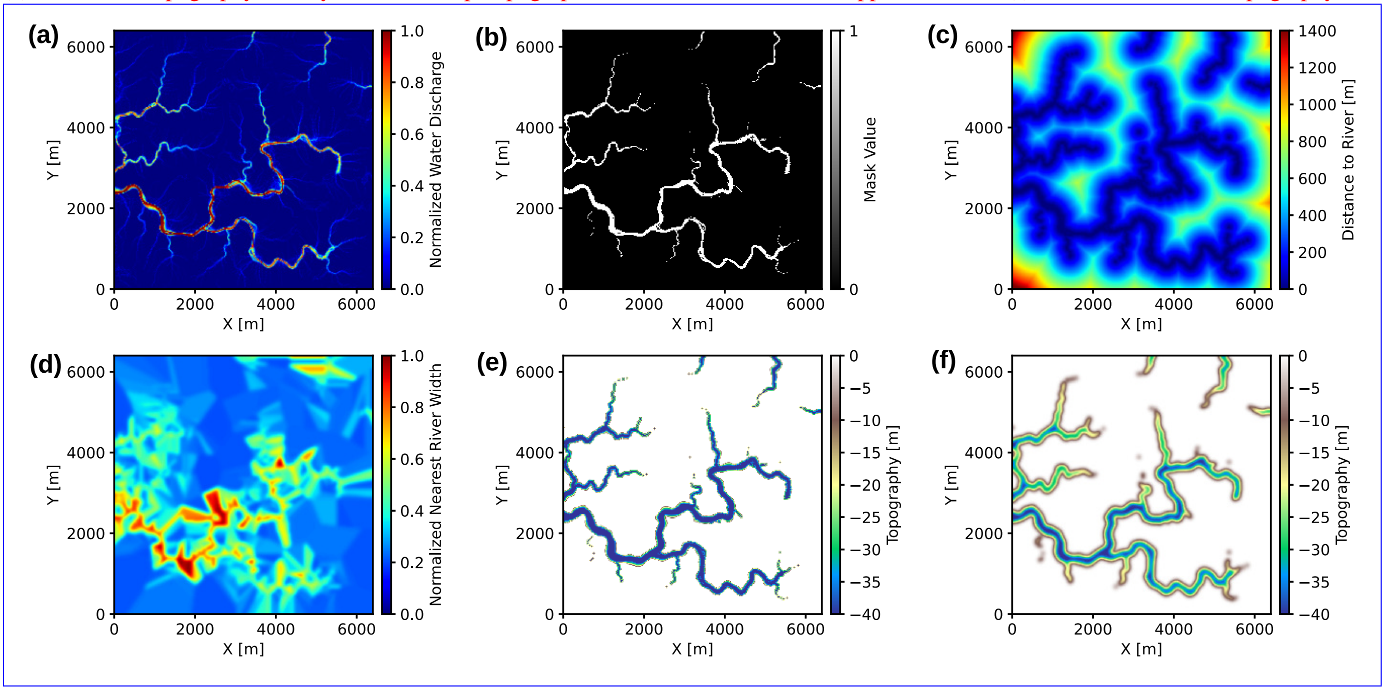


Figure 3. Tributary channel network modeling process based on the open-source package *soillib* (McDonald, 2020b). First, we generate (a) a map of normalized water discharge using the *soillib* package. Second, we create (b) the river mask by binarizing the normalized water discharge with a threshold value of 0.4, where values greater than this threshold are considered as rivers. Third, we compute (c) the Euclidean distance to rivers and (d) the normalized width of the nearest river, which are subsequently used as parameters in a parabolic function to define (e) the channel topography. Finally, to avoid abrupt topographic shifts, a Gaussian filter is applied to create (f) a smoothed channel topography.

2.2 Tributary channel network modeling

~~Distributary channels are commonly observed in river deltas, where the river channel splits into multiple smaller channels as it approaches the river mouth and spreads out into the sea or lake. Numerous numerical modeling methods based on hydrodynamics and morphodynamics have been proposed to simulate river deltas and the associated distributary channels (Seybold et al., 2007; Edmonds and Slingerland, 2007; Geleynse et al., 2011; Liang et al., 2015). However, these methods are time-consuming since they are designed to simulate detailed fluid dynamics.~~ A tributary channel network is a result of smaller rivers (tributaries) flowing into a large main river. It generally exhibit a branching or tree-like structure. To efficiently generate

~~a large number of distributary channel models~~extensive tributary channel networks that are morphologically reasonable, we adopt the open-source ~~C++~~ package *soillib* ~~(?), which is~~ (McDonald, 2020b), which offers a fast implementation of particle-based hydraulic erosion that can create a morphologically resonable tributary river network in about 10 to 20 seconds (McDonald, 2020a).

The *soillib* package is programmed to spawn hundreds of thousands of water particles at random positions on a mountainous terrain generated by ~~layered~~ random Perlin noise. The water particles move across the terrain following classical mechanics and engage in mass transfer with the surface, eventually forming ~~the distributary rivers~~a tributary river network. Figure 3a shows the normalized water discharge map of a ~~distributary~~ tributary river network generated by the *soillib* package on a 256×256 square grid with a cell size of 25m. To define the river channel topography, we first binarize the water discharge by a threshold (e.g., 0.4), where values greater than this threshold are considered as rivers (Figure 3b). Next, we compute the Euclidean distance from the river to each point on the grid (Figure 3c) and the normalized width of the nearest river (Figure 3d), which is represented by the normalized water discharge. We then define the channel topography using a parabolic function similar to that in Equation (1):

$$Z_i(x_i) = \min[4D_c(\frac{x_i}{W_c\alpha_i})^2 - D_c, 0], \tag{4}$$

where the subscript i denotes the i-th point on the grid, x is the distance to river, D_c is the maximum channel depth, W_c is the maximum channel width and α is the normalized width of the nearest river. The main modification is replacing the constant channel width W_c with a point-wise channel width $W_c\alpha_i$. By doing so, we are able to create channels with varying widths, as demonstrated in Figure 3e. The variation in channel width is controlled by α, where the mainstream is wider and the ~~distributaries~~ tributaries are narrower. However, the channel topography demonstrated in Figure 3e exhibits abrupt shifts at the channel edge due to the inherent width of the river mask. Therefore, we subsequently apply a Gaussian filter to smooth it and the final channel topography is shown in Figure 3f. When implementing the particle-based hydraulic erosion, randomness in the initial terrain and positions of water particles ensure the diversity of ~~distributary~~ tributary channels, which is demonstrated in Figure 5b. Diversity of the channel topographic models can be further increased by ~~using random~~ randomizing maximum channel widths and depths within ~~a reasonable range~~ reasonable ranges (see Table A1). Similar to the meandering channel models, our models of tributary channel networks are also designed for training deep learning models to identify the final form of the tributary channel networks. Therefore, they do not include any sedimentary process during the formation of the tributary channel network. As a result, our workflow only generates morphologically reasonable meandering channels and tributary channel networks. They lack stratigraphic components compared to those generated by stratigraphic models (e.g., Flumy (Cojan et al., 2005) and Sedsim (Wild et al., 2019)), which are more geologically realistic.

2.3 Submarine ~~channel~~ canyon modeling

~~The submarine channel is a type of underwater channel formed on the ocean floor, particularly on the margin of continental shelf~~ Submarine canyons are steep-sided valleys cut into the continental shelf at the shelf/slope break (Normark et al., 1993). They are similar to river canyons on land but are formed by the movement of turbidity currents. The pathway of the turbidity

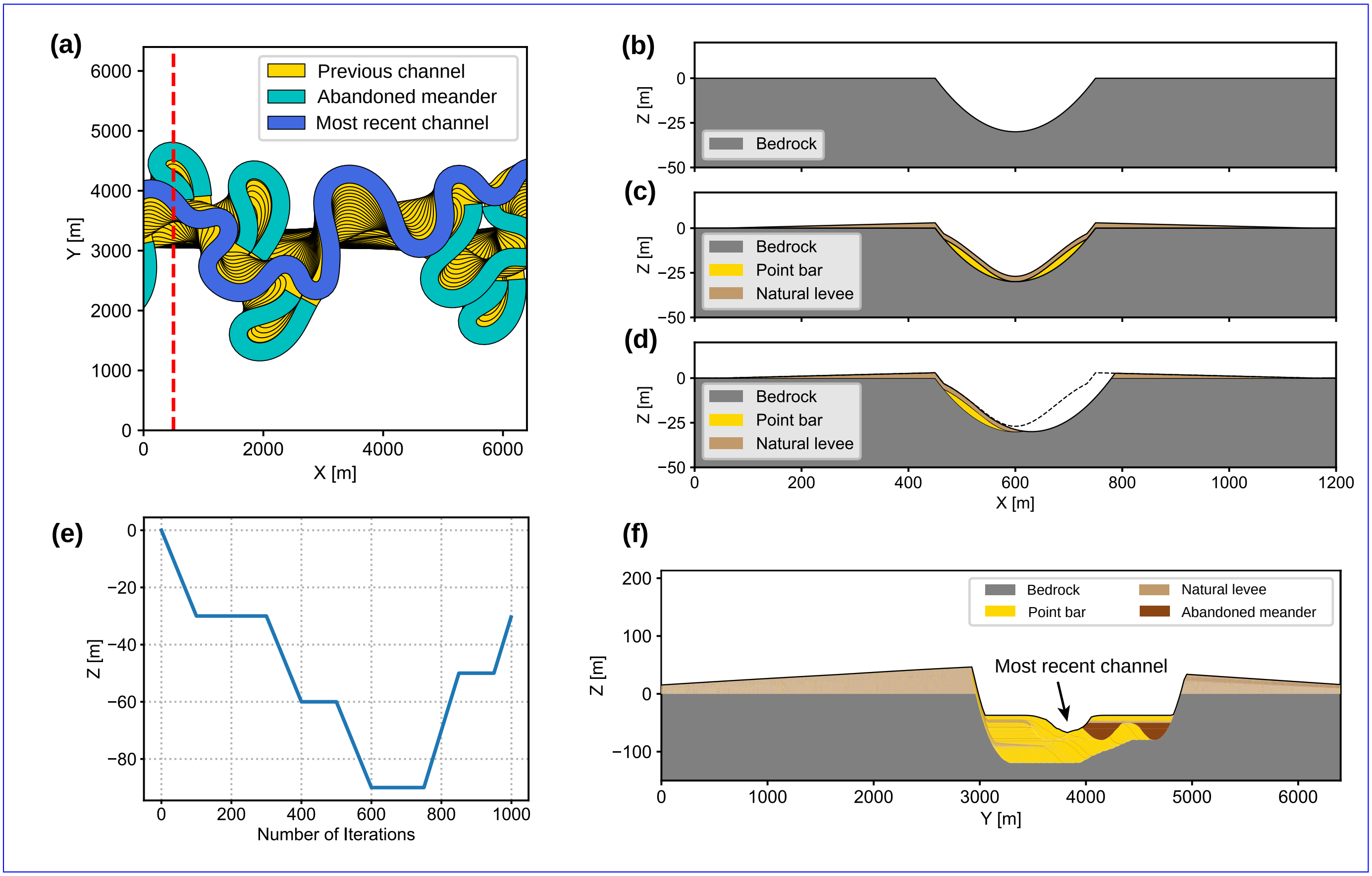


Figure 4. Submarine ~~channel~~ canyon modeling using the open-source ~~Python~~ package *meanderpy* ~~(?)~~(Sylvester, 2021). (a) Lateral migration of ~~the~~ a submarine channel within the submarine canyon. (b) Channel ~~erosional surface.~~ erosion (c) Deposition of point bars and natural levees. (d) The channel migrates towards the outer bend and erode parts of the sediments. (e) Vertical component of the channel trajectory during the migration process, which is modified from Sylvester et al. (2011), showing an initial channel incision and a later aggradation. (f) The ~~channel~~ cross-section of the submarine canyon at the red dashed line in (a), showing a large-scale erosional surface, a layered structure within the channel and a wedge-like natural levee after 1,000 ~~times~~ iterations of channel migration.

current is referred to as a submarine channel. In this work, we aim at modeling a specific type of submarine canyon related to the submarine channel-levee system (Deptuck et al., 2003; Kane et al., 2007; Catterall et al., 2010), assuming the turbidity current carries enough fine-grained sediments to form natural levees. ~~These channels are primarily carved out by turbidity currents, which carry loads of sediment from shallow coastal areas and move downslope to deeper parts of the ocean under the influence of gravity.~~ Similar to ~~terrestrial river channels which~~ a terrestrial river channel which can meander across the floodplain ~~, submarine channels also exhibit meandering patterns on the ocean floor, especially in areas of gentle slope. The meanders of submarine channels~~ on land, a submarine channel can also migrate laterally ~~and undergo cutoffs~~on the seabed. However, a ~~major difference~~ key distinction between terrestrial and submarine channels lies in the ~~significant~~ pronounced vertical incision and aggradation of submarine channels, which are driven by the ~~powerful erosive and depositional~~ erosion and deposition processes associated with ~~turbidity currents~~the turbidity current. As a result, submarine ~~channels generally pocess~~ canyons generally exhibit a large-scale ~~erosional surface and a layered structure within the channel~~erosion surface and layered sediments within the canyon, which are discernible in high-resolution seismic profile (Kolla et al., 2007).

To model the large-scale erosional surface and layered ~~structure within the channel~~sediments within the submarine canyon, we adopt ~~the~~ a modeling method based on submarine channel trajectories (Sylvester et al., 2011), which is also implemented in ~~the~~ *meanderpy*. The modeling process is illustrated in Figure 4. ~~We first simulate~~ It first simulates the lateral migration of ~~the~~ a submarine channel (Figure 4a) ~~using the same algorithm to simulate that of the meandering channel.~~ At each iteration during the migration process, a parabolic function shown in Equation (1) is used to define the surface of channel erosion (Figure 4b), which is followed by ~~the~~ deposition of point bars and natural levees (Figure 4c). Point bars are accumulated sediments on the inner bends of the channel where the flow velocity is lower. Their top surface is defined using a combination of parabolic and Gaussian function as shown in Equation (2) and (3). For modeling convenience, point bars are created on both inner and outer bends, with those on the outer bends will be subsequently eroded. Natural levees are structures that form along the sides of ~~submarine channels~~ a submarine channel when the turbidity ~~currents~~ current overflow the channel banks. They typically exhibit a wedge-like shape because the turbidity ~~currents~~ current lose energy and sediments as ~~they~~ it move away from the channel margins. The natural levee thickness is defined as follows:

$$T(x) = \begin{cases} \frac{T_{\max}}{W_l}(x - \frac{W_c - W_l}{2}), & x \geq W_c \\ T_{\max}, & x < W_c \end{cases}, \tag{5}$$

where x denotes the distance to channel centerline, $T_{\max}$ is the maximum levee thickness, W_l is the levee width on one side of the channel and W_c is the channel width. After the deposition of point bars and natural levees, the channel will migrate towards its outer bends and erode parts of these sediments (Figure 4d). The erosion and deposition processes repeat until the channel migration ends. In the meantime of lateral migration, the channel also experience vertical incision and aggradation (Figure 4e). At the end of migration, the ~~submarine channel will exhibit~~ movement of submarine channel and deposition of sediments will create a large-scale ~~erosional surface,~~ submarine canyon with a wedge-like ~~natural levee , and a layered structure within the channel, which consists of oxbow lake sediments and~~ outer levee and layered sediments within the canyon. Sediments within the canyon consist of interbedded layers of sandy point bars and ~~natural levees~~muddy inner levees, as well as muds of abandoned meanders (Figure 4f). To create diverse ~~submarine channels~~forms of submarine canyons, we use a random set of modeling parameters within ~~a reasonable range~~ reasonable ranges (see Table A1), and some of the ~~resulting topographic models are demonstrated~~ submarine canyon models are shown in Figure 5c.

2.4 Seismic volume simulation

After constructing over 10,000 channel topographic models covering meandering channels, tributary channel networks and submarine canyons, ~~distributary and submarine channels,~~ we proceed to create synthetic seismic volumes based on these models. The first step is to define the seismic impedance, which is a crucial parameter for simulating seismic events. In

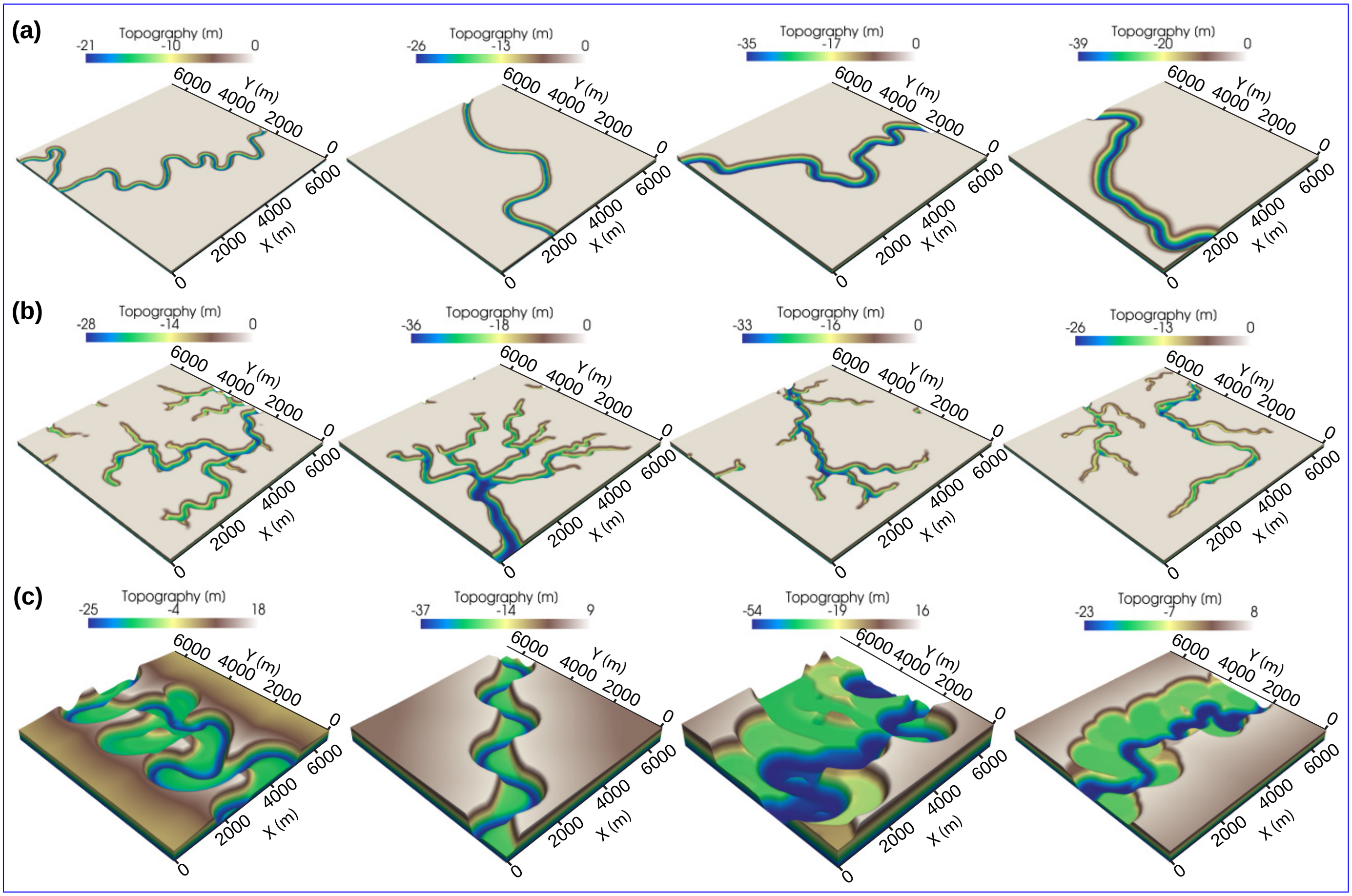


Figure 5. Diverse topographic models of (a) meandering channels, (b) ~~distributary~~ tributary channel networks and (c) submarine ~~channels~~canyons.

seismic exploration, seismic waves from an artificial source travel through the subsurface rock mass, and part of the waves will be reflected back to the surface at the boundaries of two geological layers with a contrast in seismic impedance. The reflected seismic waves will form ~~the~~ seismic events, which are considered ~~to be~~ as representatives of layer boundaries, and their amplitudes are related to the contrast in seismic impedance. We start by generating 3D seismic impedance models with horizontal layers. In each layer, we add some minor random perturbations to the impedance to make it more realistic. Details about the configuration of the impedance model are listed in Table D1. The channel topographic models are then placed at the layer boundaries, and the seismic impedance of the channel is defined according to the channel type.

Within meandering ~~and distributary~~ channels and tributary channels, we fill them with ~~uniform impedance. The impedance value~~ relatively uniform impedance with some minor perturbations (approximately 100 m/s.g/cm^3). The average impedance value of the channel is determined by a parameter ε, which is defined as the impedance contrast between the channel and its

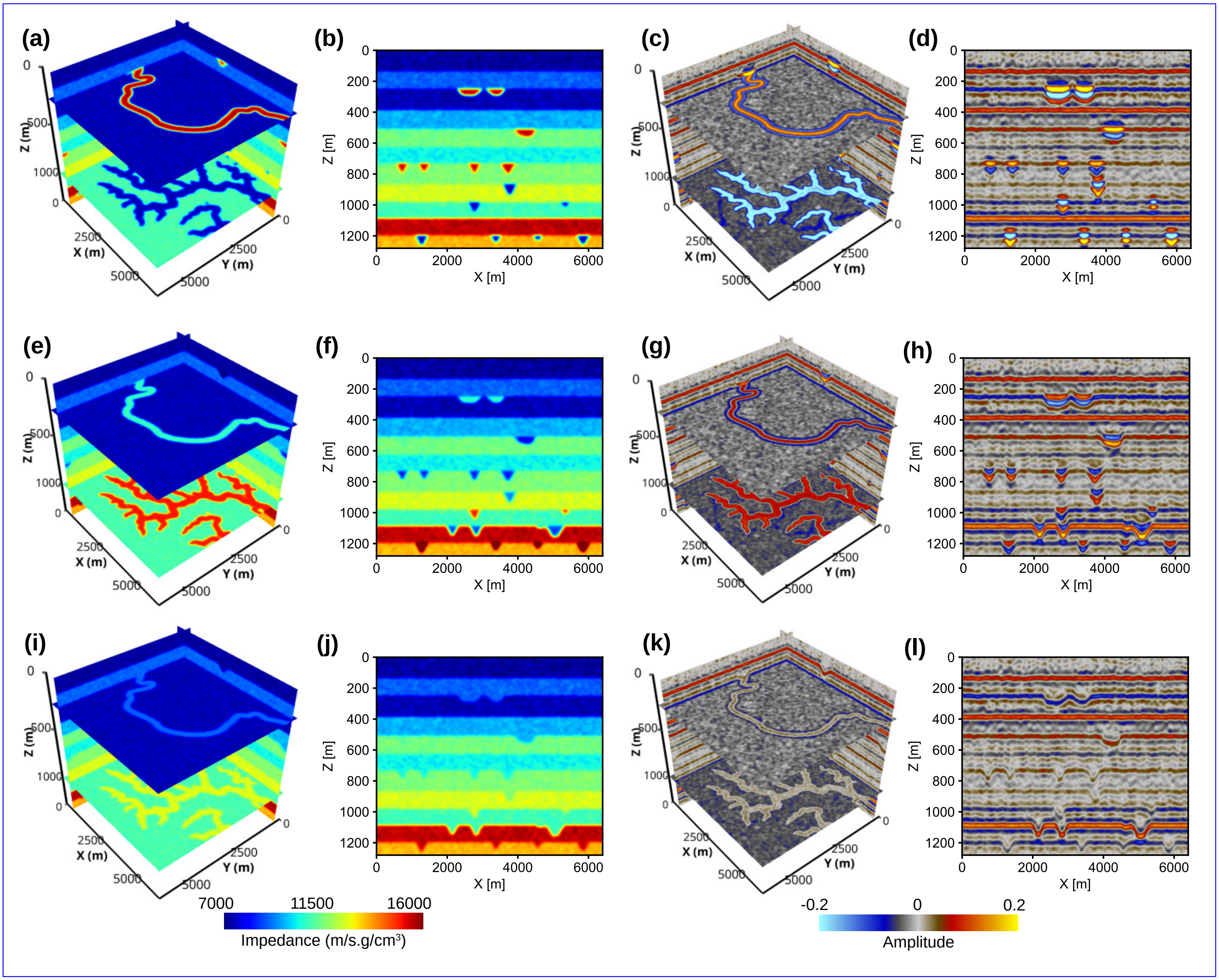


Figure 6. Seismic impedance and amplitude volumes ~~of~~ containing meandering ~~and distributary~~ channels and tributary channel networks, showing different levels of impedance contrast between ~~the channel~~ channels and ~~its~~ their covering ~~layer~~layers. (a) to (d) correspond to channels with high impedance contrast, (f) to (h) correspond to channels with low impedance contrast, and (i) to (l) correspond to channels with no impedance contrast with their covering layers.

covering layer:

$$\varepsilon = \frac{|Z_f - Z_u|}{Z_u}, \tag{6}$$

where Z_f denotes the impedance filling in ~~channels~~the channel, and Z_u denotes the impedance of the covering layer~~of the channel~~. The value of ε varies between zero and one, with the value of one indicating the highest impedance constrast between the channel and its covering layer, and the value of zero indicating the impedance of channel is the same as that of its covering layer. Figures 6a and 6b demonstrate the horizontal and vertical slices of a 3D impedance model, which consists of meandering and ~~distributary~~ tributary channels with high impedance contrast ($\varepsilon = 1$). The impedance model is then used for computing the seismic reflectivity as follows:

$$R_i = \frac{Z_{i+1} - Z_i}{Z_{i+1} + Z_i}, i = 1, 2, ..., N-1, \tag{7}$$

where the subscript i denotes the i-th point in the vertical direction of the model, and N denotes the total number of points in the vertical direction. The reflectivity model is subsequently convolved with ~~the~~ a Ricker wavelet (see Figures 8a and 8b for examples), which is commonly used to create synthetic seismic data. The mathematical expression of ~~the~~ a Ricker wavelet in the depth-domain is:

$$f(s) = (1 - 2\pi^2 k_m^2 s^2) e^{-\pi^2 k_m^2 s^2}, \tag{8}$$

where s denotes the distance and k_m denotes the peak wavenumber of the wavelet. Figure 6c shows the synthetic seismic volume corresponding to a high impedance contrast between the channel and its covering layer. We can observe that the channels have strong seismic amplitudes, appearing as bright spots on the vertical slice of the seismic volume (Figure 6d). As the value of ε decreases to 0.2, the impedance contrast between the channel and its covering layer becomes lower, as shown in Figures 6e and 6f. The corresponding seismic volume (Figure 6g) also indicates a reduction in seismic amplitude of the channels, which exhibit an infilling feature on the vertical slice of the seismic volume (Figure 6h). When the value of ε is set to zero, the impedance of channel will be the same as that of its covering layer (Figures 6i and 6j). As a result, the channels show no seismic response except at their erosion boundaries (Figure 6k), and an incision feature can be observed on the vertical slice of the seismic volume (Figure 6l).

The impedance of submarine ~~channels~~ canyons is determined based on their sedimentary facies, which include point bars, natural levees and ~~oxbow lakes~~abandoned meanders. Figure 7a shows the sedimentary facies of a submarine ~~channel~~canyon, which is primarily filled with layers of point bars as a result of ~~continuous channel migration~~ the continuous migration of a submarine channel. Additionally, the ~~channel~~ canyon is also filled with ~~oxbow lake sediments~~ sediments of abandoned meanders and inner natural levees. As shown in Figure 7b, the sediments of point bars are assigned lower impedance because they generally consist of sand, whereas the sediments of natural levees and ~~oxbow lake~~ abandoned meanders sediments are assigned higher impedance due to their muddy composition. The ~~reference impedance ranges of~~ impedance ranges we assigned for the point bars, natural levees and ~~oxbow lakes~~ abandoned meanders are listed in Table D1. It should be ~~noticed~~ noted that an impedance discrepancy exists between the neighboring layers of point bars, such that the ~~channel~~ canyon will exhibit a layered feature on the vertical slice of seismic volume and a meander belt on the horizontal slice, as shown in Figure 7c. Minor impedance perturbations ($\pm$100 m/s.g/cm^3) also exists within each sedimentary facies.

By far, all the channels and layers in the impedance model are horizontal. However, the channels and layers in practice often undergo structural deformations, such as ~~inclination and folding , which can be observed~~ folding and faulting, which are common in many field seismic volume. To increase the diversity and realism of synthetic seismic volumes, we introduce

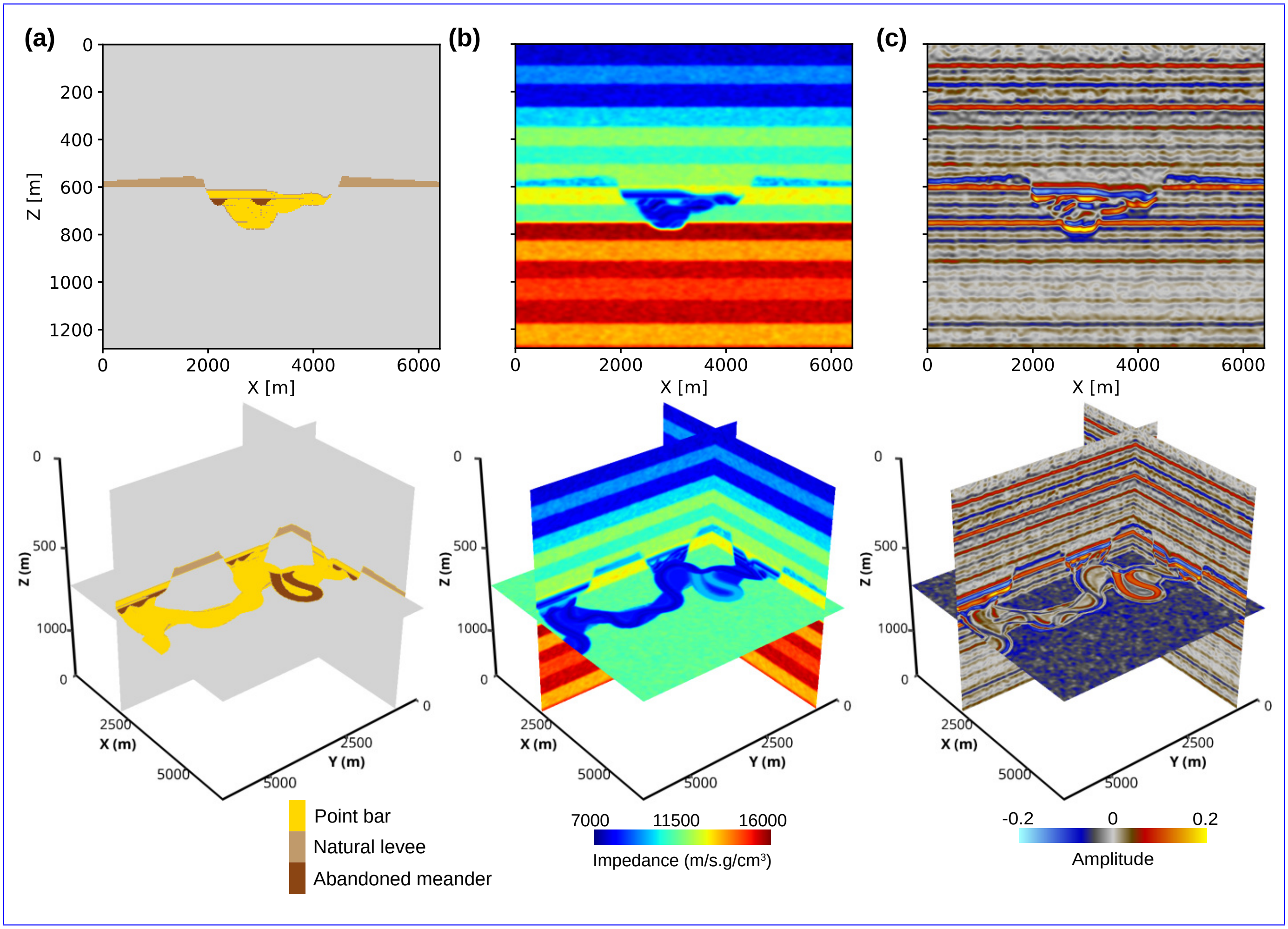


Figure 7. Illustration of creating a seismic impedance model and seismic amplitude volume containing a submarine channel according to its sedimentary facies, showing vertical and orthogonal slices of the (a) sedimentary facies, (b) seismic impedance and (c) amplitude volume.

inclination~~and folding~~ , folds and faults into the impedance model following the workflow proposed by Wu et al. (2020a). An example of the resulting impedance model with ~~inclined and folded layers~~ structural deformation is shown in Figure 8c. Another way to increase the diversity of synthetic seismic volumes is to use wavelets with various peak wavenumbers. This is also necessary because the peak wavenumber of seismic waves reflected by the channel can be diverse in field seismic volumes. It depends on various factors, such as the absorption effect of subsurface media and the characteristics of the seismic source. Figure 8 shows two synthetic seismic profiles with different wavelets computed from the same impedance model. Using a wavelet with small peak wavenumber (Figure 8a) will generate a low-resolution seismic profile with thick seismic events (Figure 8d), where some thin layers within the submarine ~~channel~~ canyon at the bottom part of the profile is hard to distinguish. On the contrary, using a large-wavenumber wavelet (Figure 8b) will create a high-resolution seismic profile

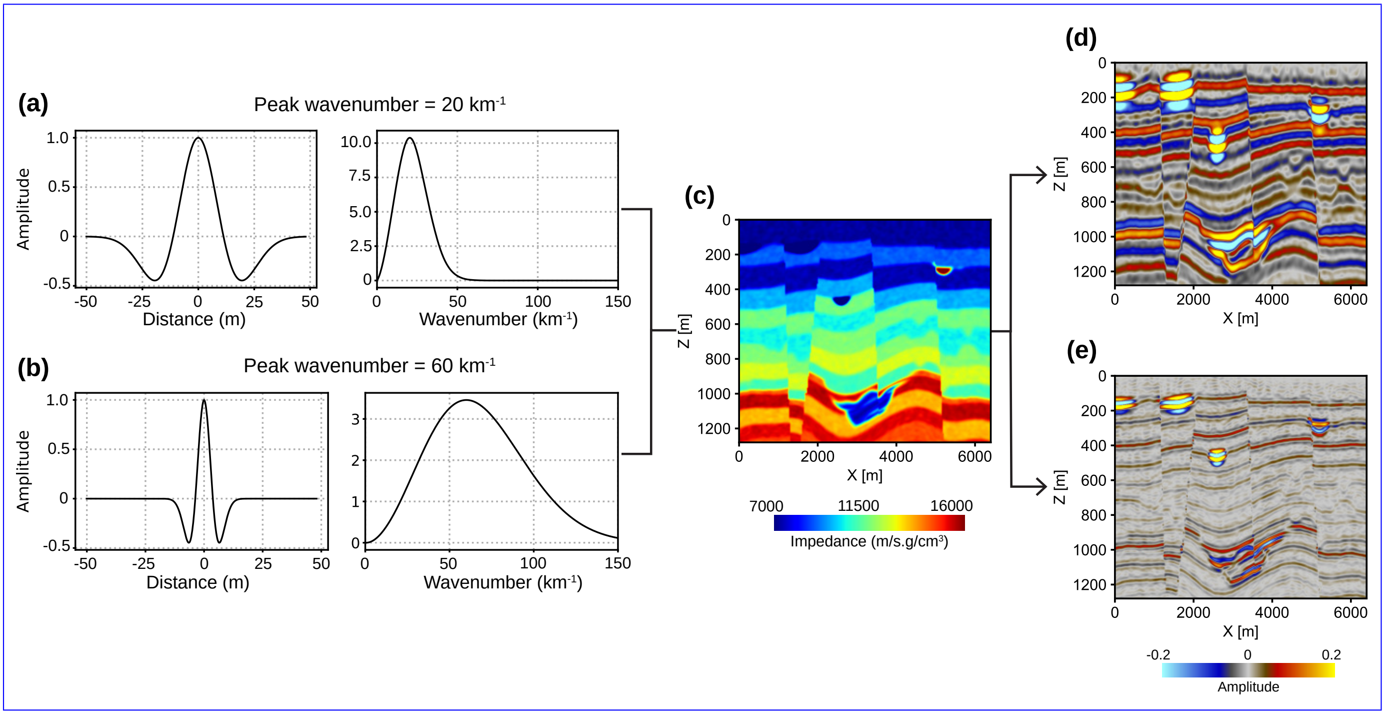


Figure 8. Synthetic seismic profile with different wavelets computed from the same seismic impedance model. (a) A small-wavenumber Ricker wavelet with a peak wavenumber of 20 km^{-1} (corresponding to a peak frequency of 20 Hz) in depth-domain and wavenumber-domain. (b) A large-wavenumber Ricker wavelet with a peak frequency of 60 km^{-1} (corresponding to a peak frequency of 60 Hz) in depth-domain and wavenumber-domain. (c) Seismic impedance model with inclined and folded structure. (d) Low-resolution seismic profile generated by using the small-wavenumber wavelet. (e) High-resolution seismic profile generated by using the large-wavenumber wavelet.

(Figure 8e), where those thin layers within the submarine ~~channel~~ canyon become discernible. The peak wavenumber range of the Ricker wavelet that we used to generate ~~the synthetic seismic volume~~ synthetic seismic volumes is listed in Table D1.

3 Results

~~Using the aforementioned~~

Using the proposed workflow, we ~~create~~ construct the *cigChannel* dataset~~containing~~, which consists of 1,~~200~~ 600 synthetic seismic volumes ~~with more than~~ containing over 10,000 labeled paleochannels. Each seismic volume has a size of 256×256×256. ~~Four~~ The dataset is organized into four task-specific subsets~~are included in the~~: meandering channels, tributary channel networks, submarine canyons, and assorted channels. In addition to the seismic volumes, the *cigChannel* ~~dataset, namely the meandering , distributary, submarine and assorted channel subsets, whose detailed components can be found~~ dataset includes the corresponding seismic impedance models. Furthermore, the submarine canyon subset provides sedimentary facies volume associated with submarine canyons. A detail breakdown of the dataset's components is presented in Table B1.

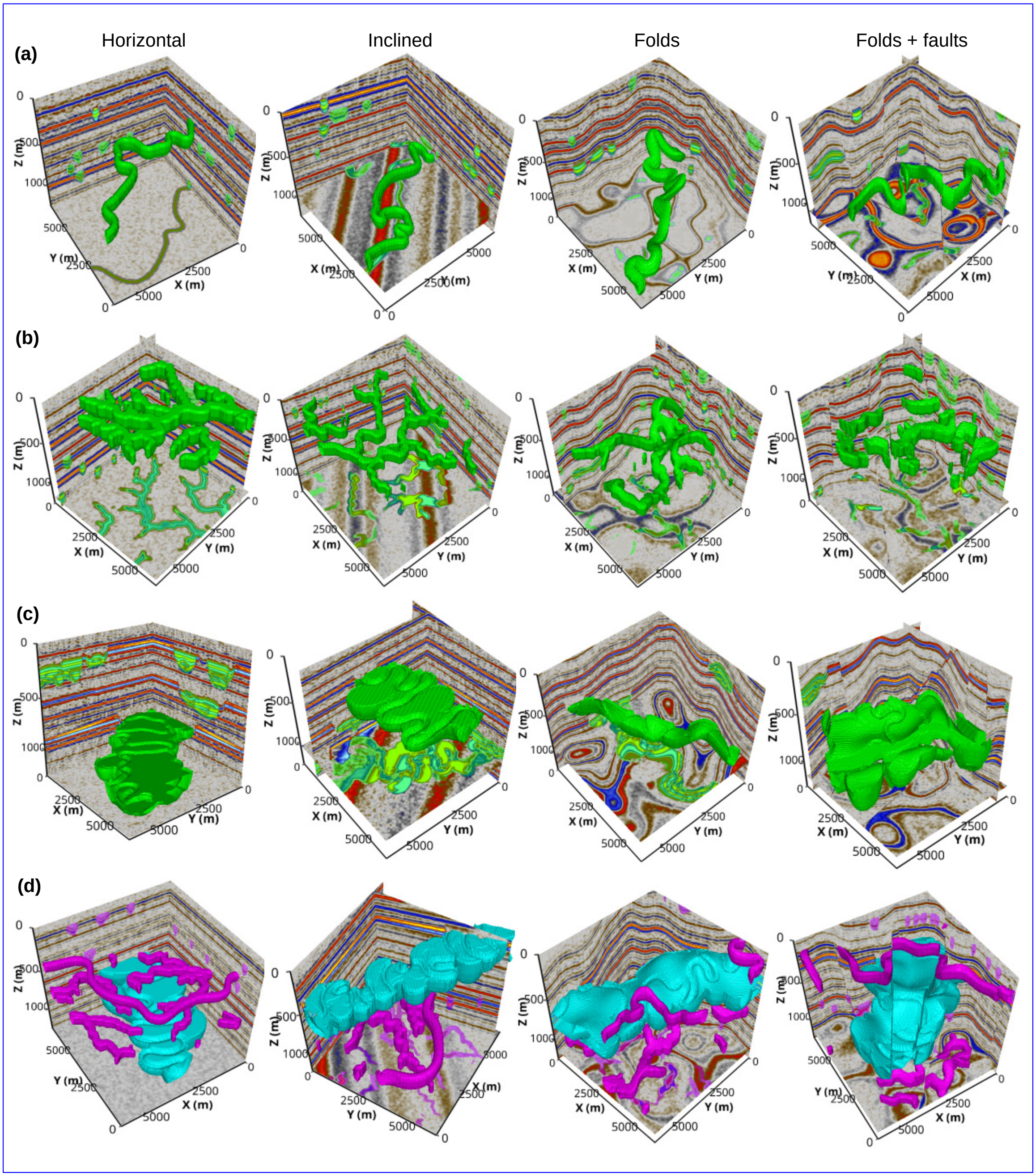


Figure 9. Synthetic seismic volumes and paleochannel labels from the (a) meandering channel, (b) tributary channel network, (c) submarine canyon and (d) assorted channel subsets of the *cigChannel* dataset, showing various types of structures. The first three subsets provides binary class labels to distinguish between channels and the background (i.e. the non-channel areas), while the assorted channel subset provides multi-class labels to distinguish between terrestrial channels, submarine canyons and the background.

Aiming to train deep learning models to identify specific types of channels, ~~each of the meandering, distributary and submarine channel subsets provides 300~~ the subsets of meandering channels, tributary channel networks and submarine canyons each provides 400 seismic volumes containing only the corresponding type of channel. Binary class labels are provided in these subsets, ~~which are designed solely to distinguish between channels and the background (i.e. the~~ with 0 denoting non-channel areas ~~)~~and 1 denoting channels. As shown in Figure 9, each subset contains seismic volumes featuring horizontal, inclined~~and folded~~ , folded and faulted structures, serving as training data for deep learning models to identify channels with various types of structures. These structures ~~. The inclined and folded structures~~ are randomly generated to introduce variability in the seismic volumes. ~~The number of channels in each individual seismic volume varies according to the size of channel. A single seismic volume may contain multiple meandering or distributary channelsyet no more than three submarine channels~~Since submarine canyons are generally wider and deeper than terrestrial channels (Normark et al., 2003; Kolla et al., 2007; Covaul , we honor this nature in the *cigChannel* dataset by generating submarine canyons larger than meandering and tributary channels.

~~Synthetic seismic volumes and paleochannel labels (visualized as coloured masks and bodies) from (a) the meandering, (b) distributary, (c) submarine and (d) assorted channel subsets of the *cigChannel* dataset, showing horizontal, inclined and folded structures. Each of the meandering, distributary and submarine channel subsets provides binary class labels to distinguish between channels and the background (i.e. the non-channel areas), while the assorted channel subset provides multi-class labels to distinguish between terrestrial, submarine channels and the background.~~

The assorted channel subset ~~contains 300 seismic volumeswith~~ also has 400 seismic volumes. Each seismic volume contains multiple terrestrial channels (including meandering channels and tributary channel networks) and a submarine canyon. This subset provides multi-class ~~channel labels. It is designed to train deep learning models not only to identify but also to distinguish terrestrial and submarine channels in seismic volumes, which is important because they are indicators for different environments. As~~ labels of non-channel areas, terrestrial channels and submarine canyons, as shown in Figure 9d~~, the terrestrial channels, which are represented by meandering and distributary channels in this dataset, have different characteristics from those of submarine channels. The most apparent one is their difference in size. Submarine channels are generally larger than terrestrial channels for many reasons. For instance, the turbidity currents that form the submarine channels are denser than their terrestrial counterparts, and the absence of vegetation on the ocean bottom eliminates a main limitation on channel erosion and sediment transport. Regarding the potential problems of the class imbalance problem and the size discrepancy between terrestrial and submarine channels, we simulate~~ . They are denoted by 0, 1, and 2 in the label volume, respectively. The reason of simulating multiple terrestrial channels but only one submarine ~~channel~~ canyon in a single seismic volume ~~in order to make~~ is to balance their voxel amounts~~as balanced as possible~~, since a model trained on an imbalanced dataset perform poorly on the minority class (i.e., the class-imbalance problem). However, there is still a huge gap in voxel amounts between channels and ~~the background. Therefore, it is suggested~~ non-channel areas. This gap exists in all four subsets. Therefore, we suggest to adopt strategies for addressing the ~~class imbalance~~ class-imbalance problem when using the *cigChannel* dataset to train a deep learning model, such as employing the ~~weighted loss functions~~class-balanced cross-entropy loss function (Xie and Tu, 2015).

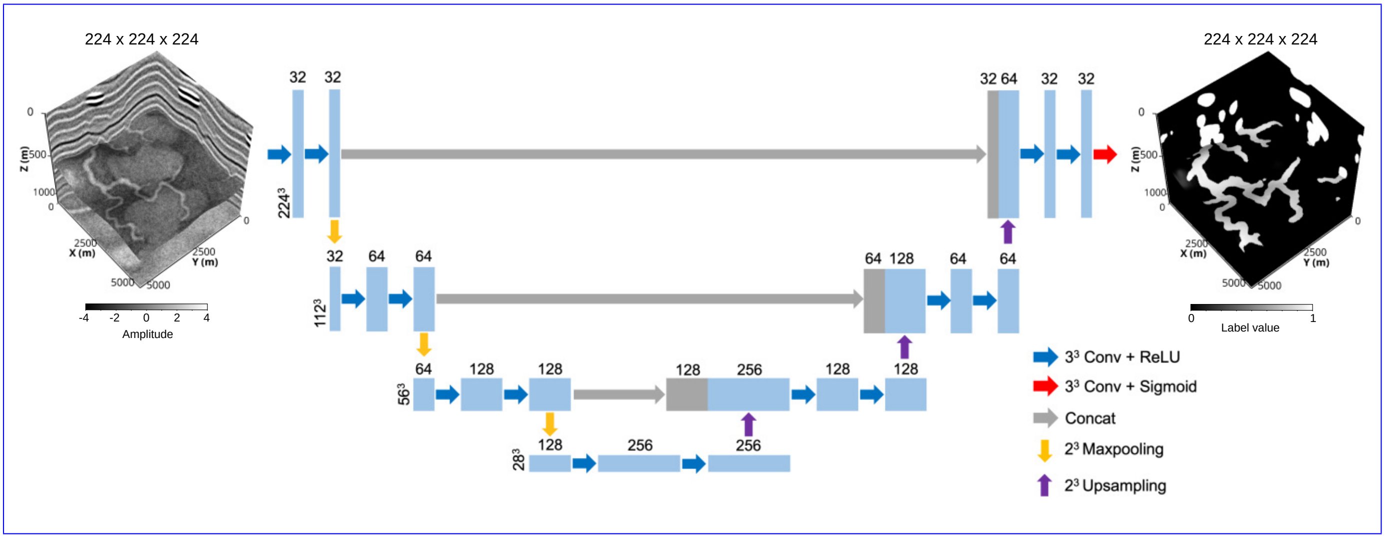


Figure 10. A simplified U-Net ~~for paleochannel identification~~ used to identify paleochannels in seismic volumes. The inputs of the U-Net are seismic volumes and the outputs are channel probabilities between 0 and 1.

4 Applications

~~(a) Field seismic volume from the Parihaka seismic survey (courtesy of New Zealand Crown Minerals), showing multiple meandering channels (indicated by the yellow arrows), their river mouth sediments (indicated by the white arrow) and a nearby fault (indicated by the red arrow). (b) The channel identification result of the U-Net trained by the *cigChannel* dataset.~~

~~(a) Field seismic volume acquired in the Tarim basin (courtesy of China National Petroleum Corporation), showing several distributary channels (indicated by the yellow arrows) with a V-shaped cross-section (indicated by the red arrow). (b) The channel identification result of the U-Net trained by the *cigChannel* dataset.~~

~~We use the *cigChannel* dataset to train a simplified U-Net and apply it~~ Three U-Nets are trained on the subsets of meandering channels, tributary channel networks and submarine canyons, respectively, which are then applied to identify paleochannels in three field seismic volumes. ~~This is a preliminary test mainly to verify the effectiveness of the dataset for training a deep learning model to distinguish between channels and the background in a field seismic volume. Therefore, the multi-class labels in the assorted channel subset are converted into binary labels like those in the other subsets. Architecture of the simplified~~ The U-Net architecture is demonstrated in Figure ~~11, which has fewer~~ 10, which is reduced in convolutional layers and feature maps ~~than its original architecture proposed by~~ compared to the original architecture in Ronneberger et al. (2015) to save memory and computational costs. The network's input is a 224×224×224 seismic volume, which is cropped from the original 256×256×256 ~~seismic volume .~~ volume due to the memory limit of GPU. Each seismic volume is normalized using the mean-variance normalization method, and Gaussian random noise is added to the synthetic seismic volume to make the training process more robust and reduce the tendency towards overfitting. The noise is zero-mean and its standard deviation is determined according to the expected signal-to-noise ratio (SNR) of the noisy seismic volume. We

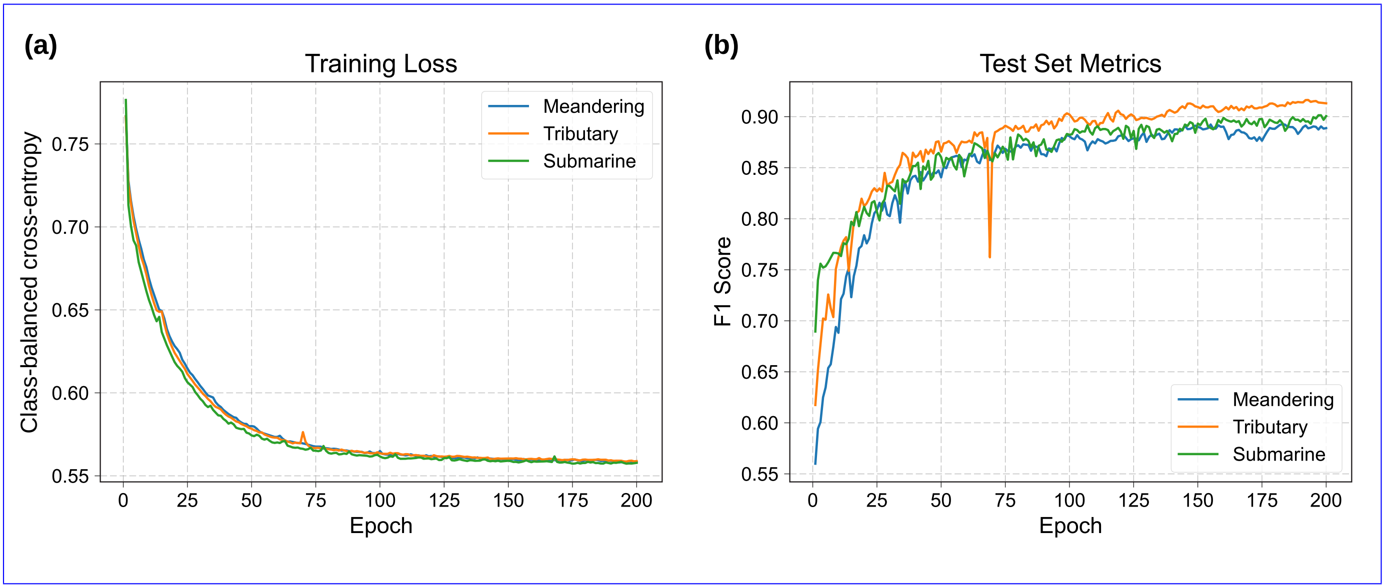


Figure 11. Training progress of the U-Net on the subsets of meandering channels, tributary channel networks, and submarine canyons, showing (a) training loss (class-balanced cross-entropy) and (b) F1 score on the test set over epochs.

set the SNR of each seismic volume to vary between 5 dB and 10 dB, which is a reasonable range for field seismic volumes (Zhang et al., 2017; Wu et al., 2021). The noisy seismic volume goes through the contracting ~~path~~ and expansive path of the U-Net for feature extraction. The final output layer ~~is a 1~~of the network is a 3×~~1~~3×~~1~~ 3 convolutional layer followed by a sigmoid activation~~to map the feature vector into channel probability values. Regarding the~~ , which maps the extracted feature into channel probabilities between 0 and 1. We binarize the channel probability values using a threshold of 0.5 in order to compare with human-made channel interpretation.

To evaluate the training performance, each subset is divided into training and test set. The training and test set contain 70% and 30% of the total samples, respectively. The class-balanced cross-entropy is used as loss function regarding the huge gap in voxel amounts between channels and ~~the background, we use the balanced cross-entropy as the loss function for networktraining.~~ non-channel areas. The F1 score is used as a metric to evaluate the network's performance on the test set. We use the Adam method (Kingma, 2014) to optimize the network's parameters and set the learning rate to be 0.0001. As shown in Figure 11, the training loss of each network converges after 200 epochs, and the F1 scores of the test sets gradually increase to around 0.9. The networks from the last epoch are used to identify paleochannels in field seismic volumes.

~~We first use the trained U-Net to identify meandering channels in a seismic~~

The U-Net is trained on the meandering channel subset and applied to a volume from the Parihaka seismic survey ~~, which is a publicly available dataset provided by the New Zealand Crown Minerals. As demonstrated in Figure 11~~(https://wiki.seg.org/wiki/Parihaka-3D). As shown in Figure 12a, the seismic volume ~~shows~~ reveals several meandering channels ~~and the sediments of their river mouths where they enter the ocean. Channel~~ feeding into a larger channel (may be a submarine canyon). The

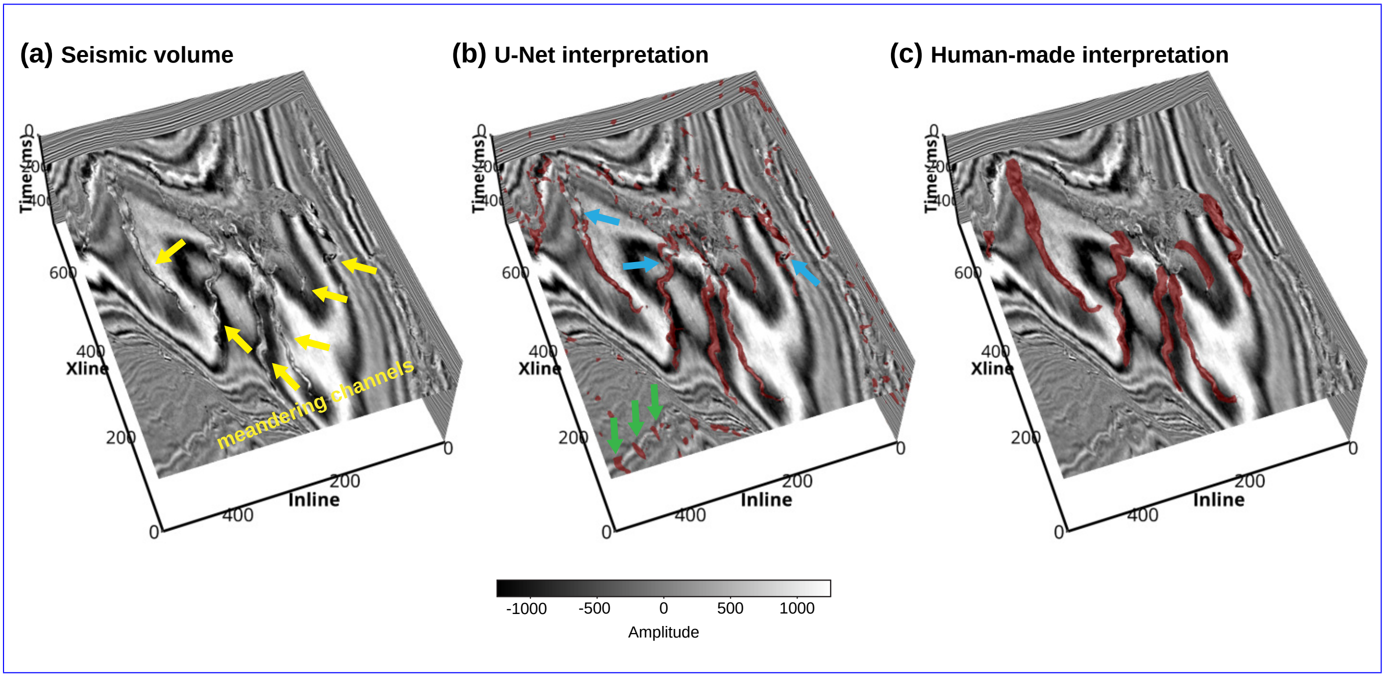


Figure 12. (a) A field seismic volume from the Parihaka seismic survey (courtesy of New Zealand Crown Minerals), showing multiple meandering channels (indicated by the yellow arrows). (b) Channel interpretation result of the U-Net trained on the subset of meandering channels. The blue arrows indicate channel areas that fail to be identified, and the green arrows indicate false positive channel identification results. (c) Human-made channel interpretation result.

channel identification result of the U-Net is shown in Figure ~~11b, where the meandering channels are all mapped with moderate to high probability. However, there is a mistaken identification of a fault at the bottom left corner of the image~~12b. which has a F1 score of 0.52 when compared to the human-made channel interpretation (Figure 12c). Some channel areas with significant variations in seismic amplitude or where the channel width suddenly increase are not correctly identified, as indicated by the blue arrows in Figure 12b. This is ~~probably because the layers are dragged downward by the normal faulting, making them exhibit an incised feature like the channels. Other noisy clusters with high channel probability may indicate segments of channels which are separated by folds or faults~~likely due to that each meandering channel in the training set has a fixed channel width, and the seismic amplitude within each channel is relatively uniform. Moreover, there are many false positive channel identification results, as indicated by the green arrows in Figure 12b, which might be local structural deformations that resemble the feature of a U- or V-shaped channel.

~~In the second example, the network is applied to identify distributary channels in a seismic volume acquired in the Tarim basin, which is provided by China National Petroleum Corporation~~

The second U-Net is trained on the tributary channel network subset and applied to a volume from an anonymous seismic survey (denoted as NW seismic survey hereafter). As demonstrated in Figure ~~12~~13a, this seismic volume shows several

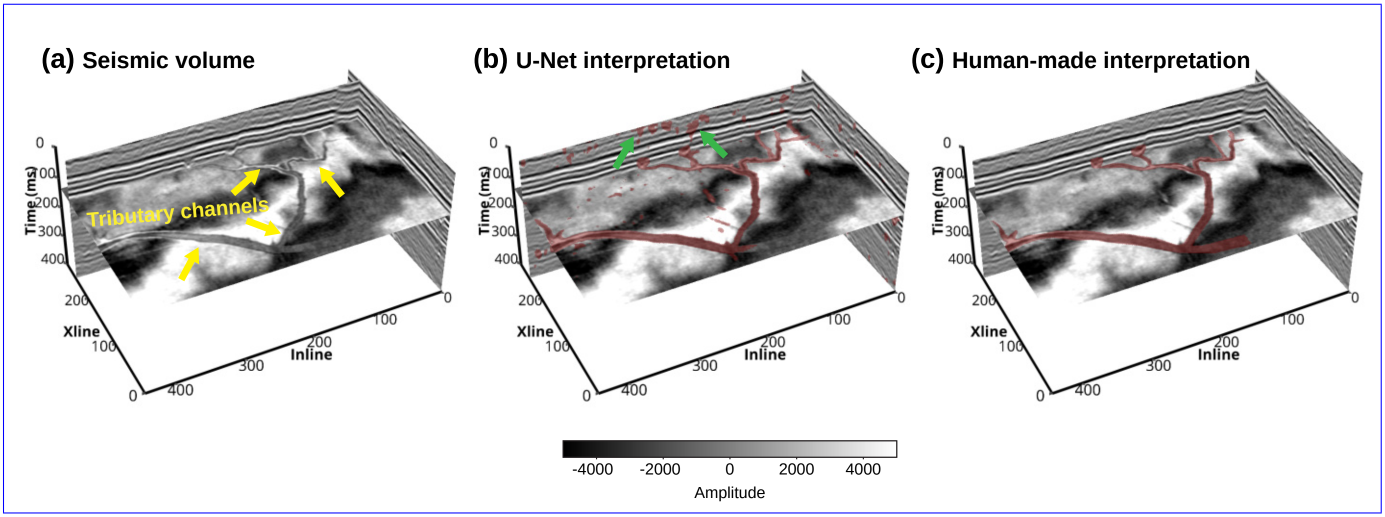


Figure 13. (a) A field seismic volume from an anonymous seismic survey (denoted as NW seismic survey), showing a tributary channel network (indicated by the yellow arrows) with V-shaped cross-sections. (b) Channel interpretation result of the U-Net trained on the subset of tributary channel networks. Some false positive channel interpretation results are indicated by the green arrows. (c) Human-made channel interpretation result.

~~distributary~~ tributary channels with a V-shaped cross-section. Seismic amplitudes within the channel are relatively homogeneous, indicating a relatively uniform seismic impedance within the channel as we designed in our ~~dataset~~ data generation workflow. The channel identification result of the U-Net is demonstrated in Figure ~~12~~13b, showing that most of the channels are correctly identified~~except some extremely narrow branches.~~

~~In the last example, we identified a submarine channel in the seismic volume from the Parihaka seismic survey, which is pointed out by the yellow~~ . However, there are still a number of small-scale structural deformations that are mistakenly identified as channels, as indicated by the green arrows in Figure 13~~a. Its large-scale erosional surface can be seen on the vertical slice of the seismic volume , which is distinct from the small-scale erosional surface of terrestrial channels, such as the one indicated in Figure 13a. This submarine channel has a medium to~~ b. The F1 score between the U-Net and human-made interpretation result (Figure 13c) is 0.73.

The last U-Net is trained on the submarine canyon subset and applied to another volume from the Parihaka seismic survey. As demonstrated in Figure 14a, a submarine canyon with a large erosion surface can be observed in this seismic volume. It has a relatively low seismic amplitude compared with that of its surrounding ~~layer, which indicates~~ layers, indicating a low discrepancy in seismic impedance within the ~~channel~~canyon. However, ~~a layered structure is~~ some layered structures are still visible within the ~~channel. The horizontal slice that intersects this submarine channel shows its meander belt with a notable boundary. Figure 13~~canyon. Figure 14b demonstrates the channel identification result of the U-Net. ~~We observe that most~~ Most areas of the submarine ~~channel are correctly mapped with high probability.~~

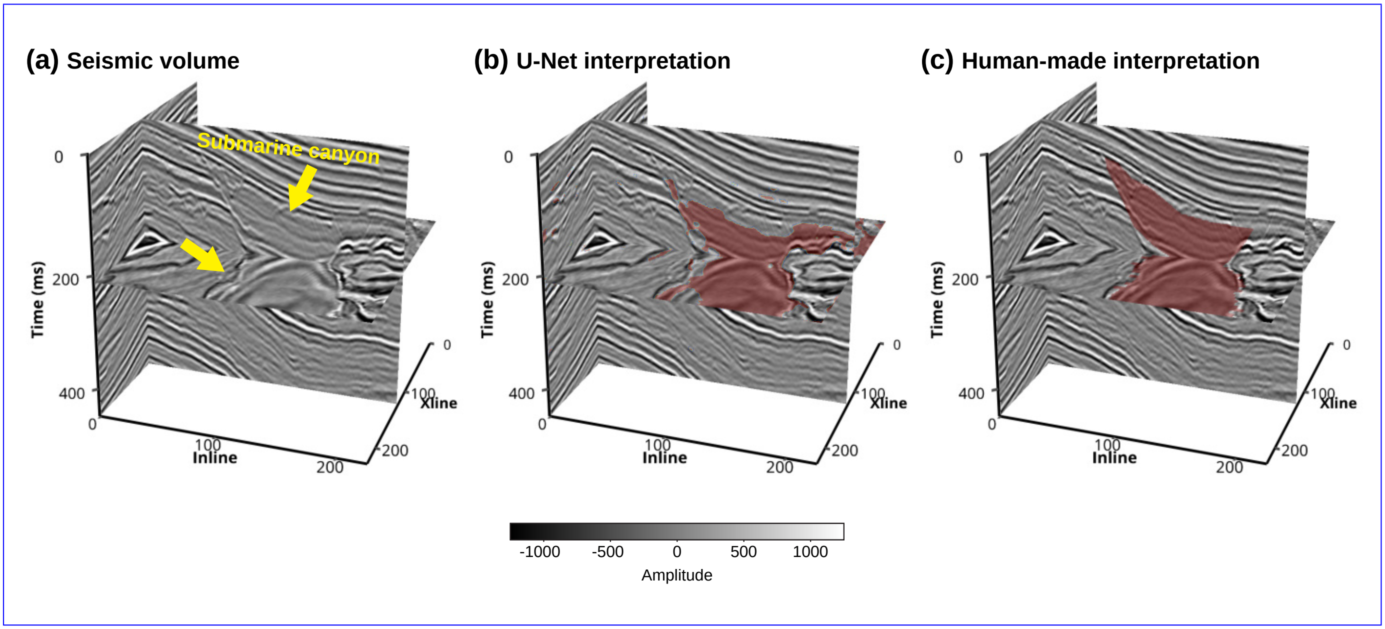


Figure 14. (a) A field seismic volume from the Parihaka seismic survey (courtesy of New Zealand Crown Minerals), showing a submarine canyon (indicated by the yellow arrows). (b) Channel interpretation result of the U-Net trained on the subset of submarine canyon. (c) Human-made channel interpretation result.

~~These applications also reveal some limitations of this dataset. As indicated in Figure 11, the network trained by our dataset cannot discriminate faults and channels, which is likely due to that faults are not included in the seismic volumes in this dataset~~canyon are correctly identified but the U-Net cannot delineate the canyon boundary accurately. The F1 score between the U-Net and human-made interpretation result (Figure 14c) is 0.63. ~~Therefore, adding faults to the seismic volume~~

5 Discussion

5.1 Plausibility of the synthetic seismic volumes

While the *cigChannel* dataset provides various samples for training deep learning models to identify paleochannels in seismic volumes, the plausibility of the synthetic seismic volume remains uncertain. Several simplifications are applied to reduce computational costs during the generation of synthetic seismic volumes. For instance, the configuration of seismic impedance models ignores the variability within layers and channel facies. However, this variability is ubiquitous in the subsurface. Moreover, the forward seismic modeling uses the simplest 1D convolution between seismic (P-wave) impedance and Ricker wavelet. It disregards many aspects of wave propagation in the subsurface, including the contribution of shear waves, separate contributions from P-wave velocity and density, and ~~labeling them as the background would help reduce the~~ multi-path

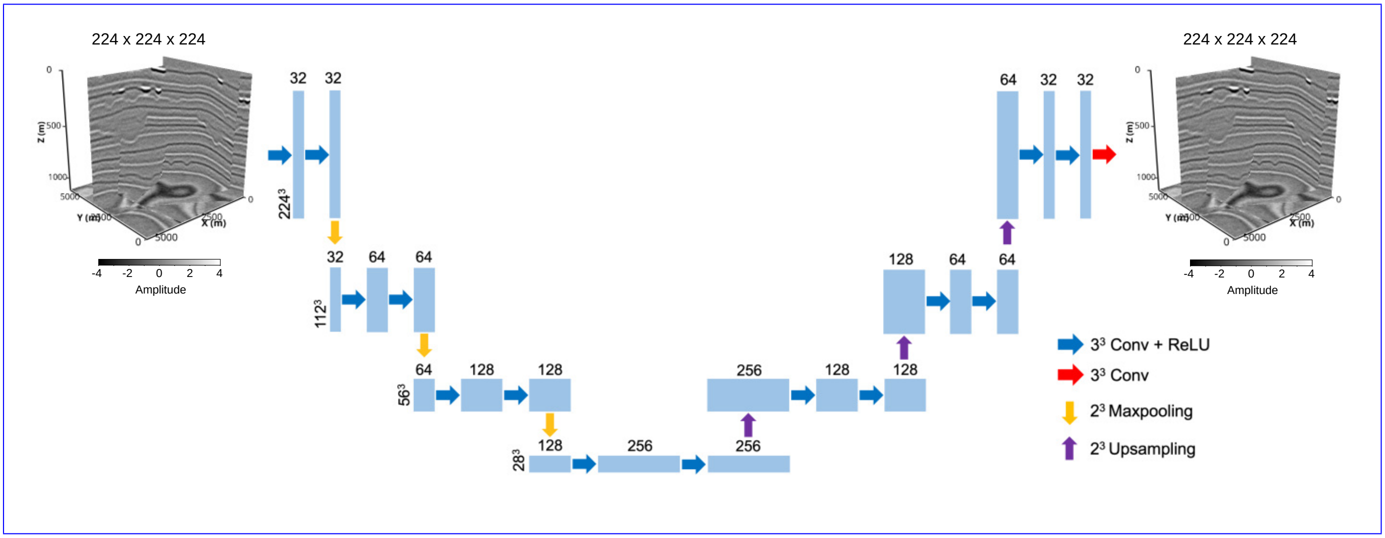


Figure 15. U-Net-based autoencoder architecture for reconstructing seismic volumes. Compared to the U-Net architecture used for paleochannel identification, the skip connections are removed, and the final layer is a $3\times3\times3$ convolutional layer without sigmoid activation. The inputs of the autoencoder are the original seismic volumes and the outputs are their reconstruction results.

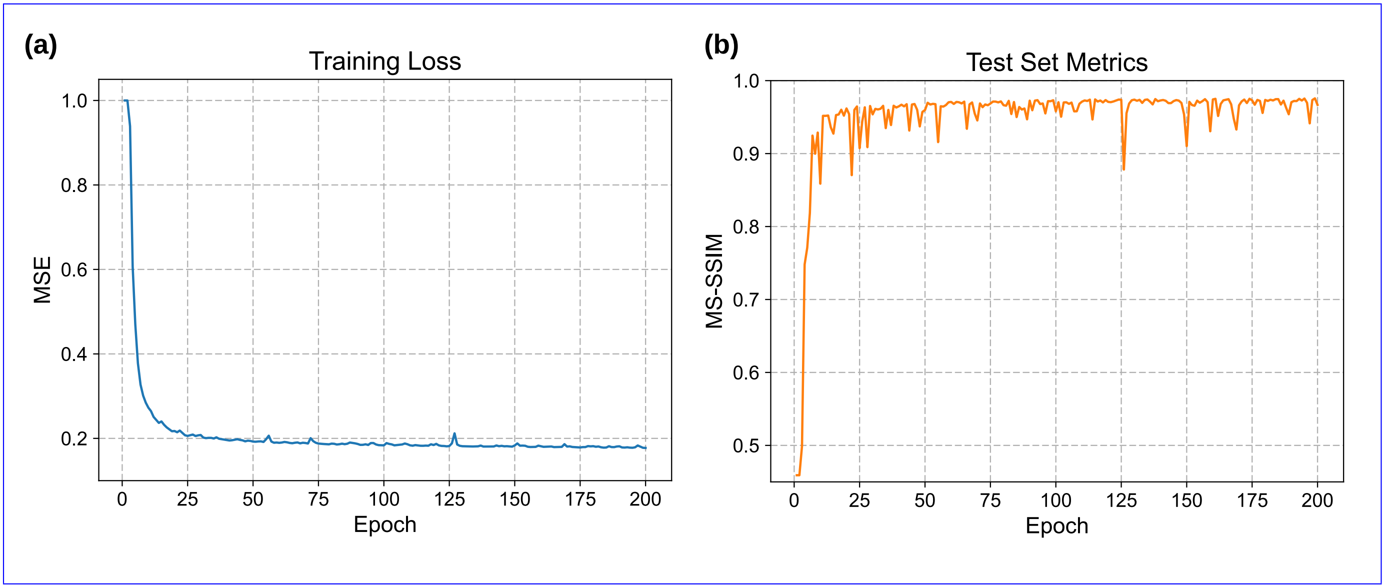


Figure 16. Training progress of the U-Net-based autoencoder, showing (a) training loss (mean squared error) and (b) multi-scale structural similarity (MS-SSIM) on the test set over epochs.

reflection. These simplifications reduce the realism of synthetic seismic volumes. It is questionable whether the synthetic seismic volumes can capture the patterns in the field seismic volumes.

To answer this question quantitatively, we use the synthetic seismic volumes in the *cigChannel* dataset to train an autoencoder to reconstruct seismic volumes. If this autoencoder can reconstruct the field seismic volumes as well as the synthetic ones, it means that the synthetic seismic volumes are plausible and representative enough of field seismic volumes. Otherwise, it indicates room for improvement. To construct training and test set, we randomly choose 70 samples for training and 30 samples for testing from each subsets. That makes a total number of 280 training samples and 120 test samples. The architecture of the autoencoder is adapted from the U-Net used for identifying paleochannels. As shown in Figure 15, we remove all the skip connections from the U-Net and the sigmoid activation from the final convolutional layer. Each synthetic seismic volume is cropped into a size of 224×224×224. They will serve as both inputs and labels to train the autoencoder. The seismic volumes (both synthetic and field ones) will be normalized and zero-mean Gaussian random noise will be added to the synthetic seismic volume. The standard deviation of the noise is determined according to the expected SNR of the noisy seismic volume, which is set to vary between 5 dB and 10 dB. During the training process, the mean squared error (MSE) between the original and reconstructed seismic volumes will be calculated as the training loss, and the multi-scale structural similarity (MS-SSIM) will be used as metrics to evaluate the network's ~~tendency to mis-identification between faultsand channels. It can also be seen that~~ generalization performance on the test set.

Figure 16 shows the evolution of training loss and test set metrics over the training epochs. The training loss decreases rapidly in the first 25 epochs, and reaches fully convergence after 200 epochs. Meanwhile, the reconstruction of seismic volumes in the test set achieves an average MS-SSIM of 0.96, in spite of some minor fluctuations. The reconstruction of a synthetic seismic volume from the test set is demonstrated in Figure 17a. Seismic events, including those related to the paleochannels (indicated by the yellow arrows) are mostly reconstructed. However, as shown in the residual volume, random noise, artifacts related to faults, and some weak seismic reflections within geologic layers (i.e., between seismic events) are not fully recovered. The reconstruction results of the three field seismic volumes with meandering, tributary channels and submarine canyons are respectively demonstrated in Figure 17b, c, and d. The general patterns (i.e., geometries, relative seismic amplitudes) of the seismic events and paleochannels have been successfully reconstructed. However, we can see from the residual volumes that many detailed seismic reflections related to the geologic layers and paleochannels have not been recovered, especially for the seismic volumes from the Parihaka survey (Figure 17b and c). Table 1 lists the metrics of the autoencoder for reconstructing synthetic and field seismic volumes shown in Figure 17. The reconstruction of the Parihaka seismic volumes (Figure 17b and d) is less accurate compared to that of the synthetic seismic volume (Figure 17a). However, the ~~channel identification result in Figure 11 is not as good as that in Figure 12, where the distributary channels in Figure 12 are mapped with uniformly high probability while some parts of the meandering channels in Figure 11 are mapped with moderate probability. It is probably because that the distributary channels in Figure 12 are filled with uniform seismic amplitude as we designed in our dataset , while the meandering channels in Figure 11 are filled with heterogenous seismic amplitude, which is an exceptional case for our dataset. Therefore, the identification performance of channels with heterogeneous seismic amplitude would be improved if meandering and distributary channels with heterogeneous seismic amplitude can be included in this dataset. As we mentioned, these are preliminary tests mainly to find out whether this dataset can help the network discriminate channels and~~ autoencoder is capable of reconstructing the NW seismic volume (Figure 17c) with a quality comparable to that of the synthetic seismic volume.

Table 1. Metrics of the autoencoder for reconstructing synthetic and field seismic volumes.

Seismic/channel type	Source	MS-SSIM* ↑	MSE* ↓
Synthetic/assorted (Figure 17a)	cigChannel Dataset	0.93	0.17
Field/meandering (Figure 17b)	Parihaka survey	0.86	0.23
Field/tributary (Figure 17c)	NW survey	0.95	0.04
Field/submarine (Figure 17d)	Parihaka survey	0.79	0.23

* MS-SSIM: Multi-scale structural similarity

* MSE: Mean squared error

The difference in reconstruction performance on field seismic volumes is likely related to the variability in seismic data. Compared with the two Parihaka seismic volumes (Figure 17b and d), the NW seismic volume (Figure 17c) has less variations in seismic amplitude along seismic events, and the seismic amplitude within paleochannels is relatively uniform. These characteristics are similar to the synthetic seismic volumes, and therefore the autoencoder can reconstruct the NW seismic volume as effectively as the synthetic ones. In conclusion, the synthetic seismic volumes have captured the general patterns in field seismic data, such as the geometries of structures and paleochannels. However, they cannot capture the detailed variations in seismic data that are related to wave propagation and changes in rock properties. This may lead to generalization issues for deep learning models trained on this dataset when applied to field seismic volumes with significant variability. Applying more realistic seismic forward modeling methods such as full-waveform modeling and considering the variations in rock properties within geologic layers and paleochannel facies could help improve the plausibility of the synthetic seismic volumes.

5.2 Limitations of the dataset

Although the application of the *cigChannel* dataset has shown its capability of training deep learning models to identify paleochannels in field seismic volumes, there are several limitations of this dataset that users should be aware of. The first one lies in the diversity of terrestrial channel and submarine canyon models. The widths of terrestrial channels are set to be relatively small ($\leq$ 500 m) in order to be more distinguishable with submarine canyons. However, much wider terrestrial channel systems (e.g., $\geq$ 1 km) have also been reported (Gibling, 2006), which could be comparable in size with a relatively narrow submarine canyon such as the La Jolla canyon (Paull et al., 2013). Therefore, if the aim is to train a deep learning model to differentiate between terrestrial and submarine channel systems, then the model trained on the assorted channel subset may face challenges when distinguishing small submarine canyons from large terrestrial channels. Moreover, as we mentioned, our modeling of submarine canyons aims to replicate the characteristics of the submarine channel-levee system, which requires enough fine-grained sediments to form levees. Relatively coarse grained sediments (e.g., conglomeratic channel lag deposits) that correspond to a sandier depositional environment are not captured in our submarine canyon models. Consequently, deep learning models trained on the submarine canyon subset may struggle to accurately identify submarine canyons that contain a significant amount of coarse-grained sediments.

The second limitation concerns the realism of seismic impedance within channels. We assign a relatively uniform seismic impedance to terrestrial channels, introducing slight random perturbations to capture natural variability. The seismic impedance of these channels is determined based on a predefined contrast with the surrounding layers. However, these simplifications reduce the realism of the impedance representation. In reality, terrestrial channel fills exhibit variations in facies and lithologies (Miall, 2014; Mueller and Pitlick, 2013), which can result in considerable seismic impedance heterogeneity. Although under certain circumstances this heterogeneity could be diminished due to the relatively small size of terrestrial channels and the inherent limitations of seismic resolution, assigning a relatively uniform impedance to terrestrial channels limits the comprehensiveness of their seismic response. As a result, deep learning models trained on the meandering or tributary channel subset may face challenges to accurately identify channels that exhibit heterogeneous seismic amplitudes, such as the example shown in Figure 12. Additionally, for submarine canyons, seismic impedance variations related to grain size distribution within sedimentary facies are not accounted for. The spatial transition from coarse-grained sediments in the channel thalweg to fine-grained sediments along the channel margins (Jobe et al., 2017) is not represented in our impedance models, which further limits the diversity and realism of the synthetic seismic volumes. Consequently, deep learning models trained on the submarine canyon subset may face generalization challenges when applied to identify submarine canyons in field seismic volumes.

The third limitation relates to the realism of non-channel areas ~~. Future work could involve using this dataset to train a network to classify terrestrial and submarine channels,~~ in the synthetic seismic volumes. In addition to not fully capturing various characteristics of wave propagation due to the use of 1D convolution for seismic synthesis, the synthetic seismic volumes also lack structural diversity and stratigraphic variability. While folds and faults are included, their scales are enlarged to be comparable to the horizontal extent of the seismic volumes (i.e., 6.4 km). Small-scale (e.g., hundreds of meters) structural deformations, particularly those forming localized U- or V-shaped geometries, are not incorporated, despite their common occurrence in field seismic volumes. Consequently, deep learning models trained on our dataset may struggle to distinguish between small-scale concave structures and U- or ~~to interpret the sedimentary facies of the submarine channels.~~ V-shaped channels, which could lead to false positive results. Moreover, each layer in the seismic impedance model is assigned a uniform thickness and a relatively consistent seismic impedance, resulting in a lack of stratigraphic variability in the synthetic seismic volumes. Given this limitation, it is not surprising that a deep learning model trained on our dataset may infer that the primary distinction between channel and non-channel areas is the presence of stratigraphic variability. This inference arises because, in the synthetic seismic volumes, channels—particularly submarine canyons—are the only structures exhibiting such variability. However, in field seismic volumes, stratigraphic variability is widespread among non-channel areas. Consequently, deep learning models trained on our dataset may produce false positives in non-channel areas with significant stratigraphic variability.

~~(a) Field seismic volume from the Parihaka seismic survey (courtesy of New Zealand Crown Minerals), showing a large-scale submarine channel (indicated by the yellow arrows). (b) The channel identification result of the U-Net trained by the *cigChannel* dataset.~~

6 Conclusions

~~The~~ In this paper, we present a workflow for generating a large number of 3D synthetic seismic volumes containing paleochannels along with their corresponding segmentation labels. Using this approach, we construct the *cigChannel* dataset~~is dedicated to overcome the shortage~~ , which comprises 1,600 seismic volumes featuring three distinct types of paleochannels. This dataset is designed to address the scarcity of training data for deep learning-based paleochannel identification in seismic volumes. ~~It provides a more~~ Compared to previously used datasets (Pham et al. (2019) and Gao et al. (2021)), the *cigChannel* dataset offers a more diverse and comprehensive collection of paleochannels~~than its predecessors. Workflow for generating this dataset was designed to produce synthetic seismic volumes with realistic characteristics of paleochannels, which exhibit large variability due to the randomization of many parameters that control the workflow.~~ . The effectiveness of this dataset is demonstrated

~~by its application on several~~ through its application to three field seismic volumes, ~~which shows that even~~ where a simplified U-Net~~works well in identifying paleochannels after being trained with our dataset.~~

~~Other than providing training data~~ , trained on the *cigChannel* dataset, successfully identifies paleochannels with promising results. This highlights the feasibility of using synthetic data to train deep learning models for paleochannel identifications, bridging the gap between limited field seismic volume annotations and the need for efficient and robust seismic paleochannel interpretation. Beyond providing a rich source of training samples for deep learning models~~to identify paleochannels in seismic volumes, this dataset can also serve as a publicly available benchmark dataset for validating the performance of various deep learning models and training strategies. This dataset can be further improved by incorporating new elements into the dataset generation workflow, such as adding faults to create a more complex structure and introducing heterogeneous seismic amplitude to the meandering and distributary channels . As the codes corresponding to the dataset generation workflow are also made publicly available, users can customize the controlling parameters and create datasets that used to identify specific forms of paleochannels~~, the *cigChannel* dataset and its generation workflow hold potential for advancing seismic modeling techniques and supporting educational applications. For example, rock physics models incorporating fluvial or turbiditic facies could be developed to evaluate new seismic modeling approaches, while the synthetic seismic volumes could serve as effective tools for demonstrating the influence of geological heterogeneities on seismic data. However, synthetic seismic volumes in the *cigChannel* dataset still lack the diversity and realism of field seismic volumes, primarily due to the simplifications of channel modeling, seismic impedance representation, and the synthesis of seismic volumes.

In the future, we aim to enhance our workflow to improve the realism and diversity of the generated seismic volumes. Terrestrial meandering and tributary channels will be modeled using stratigraphic approaches to better capture sedimentary processes, thereby enhancing geological realism. The dataset will also be expanded to include a broader range of channel types, such as braided and deltaic systems, further increasing its diversity. To improve seismic impedance modeling, we plan to account for grain size distribution and its impact on impedance variations within channel sedimentary facies. Additionally, the current simplistic 1D convolution will be replaced with 3D convolution or full-waveform modeling to better capture seismic data variability. These advancements will enhance the geological realism and diversity of our dataset, ultimately improving its effectiveness for deep learning-based seismic paleochannel interpretation.

7 Code and data availability

The *cigChannel* dataset (Wang et al., 2024) can be accessed via Zenodo. It has been organized into four subsets, whose links are provided as followed:

1. Meandering channels: https://doi.org/10.5281/zenodo.11078794;
2. Tributary channel networks: https://doi.org/10.5281/zenodo.11073030;
3. Submarine canyons: https://doi.org/10.5281/zenodo.11079950;
4. Assorted channels: https://doi.org/10.5281/zenodo.11044512.

Codes corresponding to the dataset generation workflow are provided on GitHub (https://github.com/wanggy-1/cigChannel).

The three seismic volumes demonstrated in the Application section can be downloaded from the following links:

1. Meandering channel example: https://drive.google.com/file/d/1ItOmdluWUfApzamA4mCeJNhnz_CYUZuf/view?usp=drive_link;

2. Tributary channel example: https://drive.google.com/file/d/1l4-gBRE-SEoQkx7souERjtiRLpyABrJ-/view?usp=drive_link;

3. Submarine canyon example: https://drive.google.com/file/d/1qxO8-onWFlffp7t3UHMtm-rHUmkkMvQx/view?usp=drive_link.

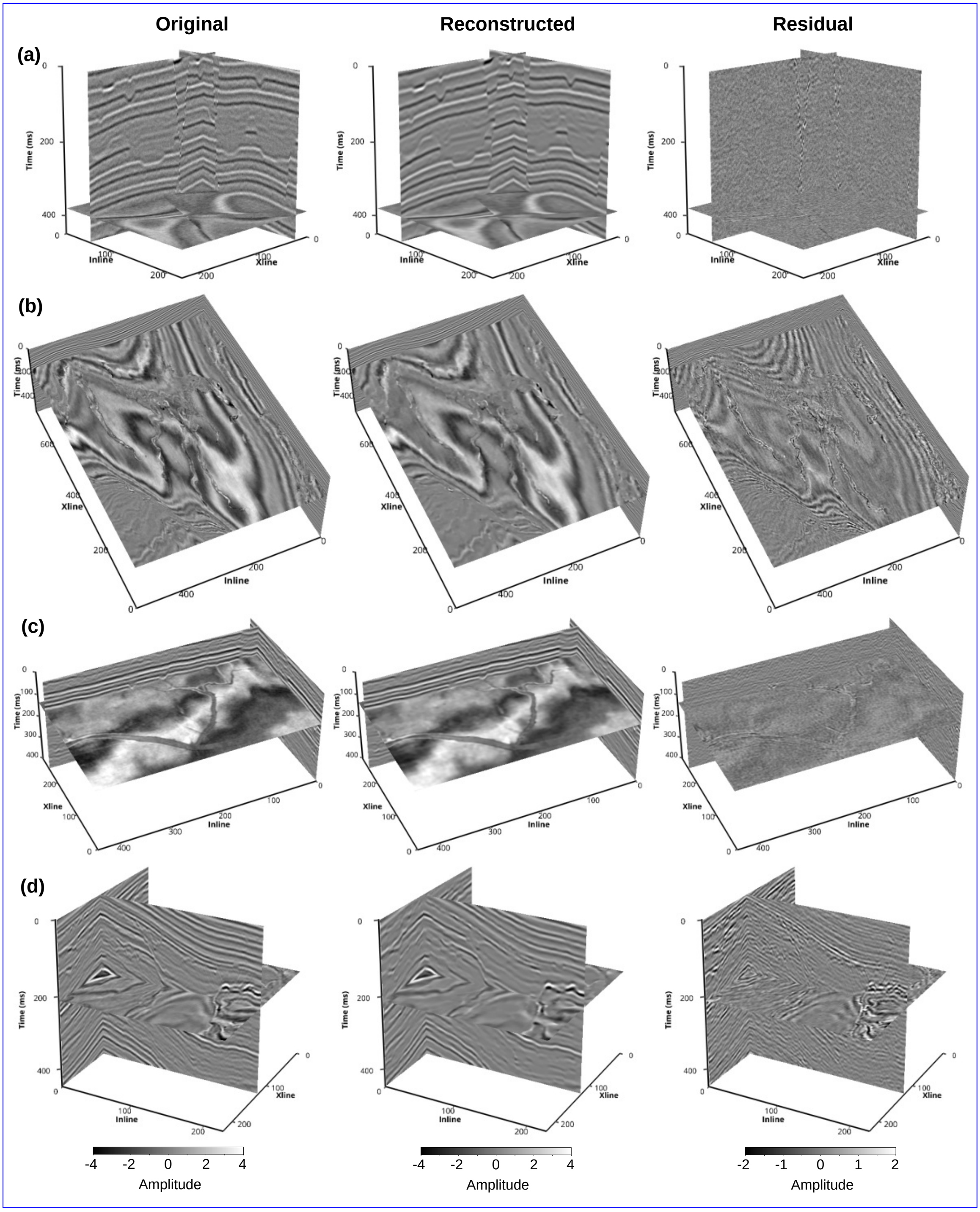


Figure 17. The original, reconstructed and residual volumes of (a) synthetic seismic data with assorted channels, and field seismic data with (b) meandering channels, (c) tributary channels and (d) submarine canyon.

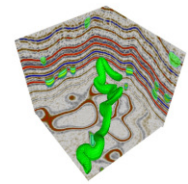
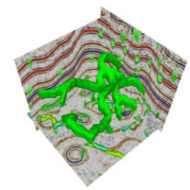
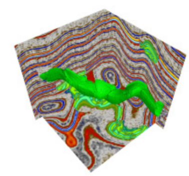
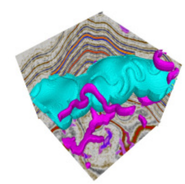
Appendix A: Channel modeling parameters

Table A1. Channel modeling parameters~~with their reference values~~.

Channel type	Parameter	Value	Reference
Meandering channel	Width	200 m - 500 m	30 m - 15 km (Gibling, 2006)
	Maximum depth	20 m - 50 m	1 m - 38 m (Gibling, 2006)
	Strike	N0°E - N360°E	
	Migration rate constant ~~*~~	40 m/~~year~~ yr - ~~50m~~50 m/~~year~~ yr	Exaggerated to accelerate simul reference range: 0.5 m/yr - 15 (Donovan et al., 2021; Schook et al., 2017
	Dimensionless Chezy's friction fractor ~~*~~	0.06 - 0.08	Exaggerated to accelerate simulation; refe range: 0.002 - 0.005 (Chow, 1988)
	Iteration time step ~~*~~	0.1 ~~year~~ yr	
	Number of iterations ~~*~~	1000 - 2000	
Tributary channel	Maximum width	~~200m~~ 200 m - 400 m	10 m - 1000 m (Trigg et al., 2012)
	Width/depth ratio	10 - 12	2 - 870 (Gibling, 2006)
	Maximum number of iterations ~~†~~	8192	
	Number of Particles for early-termination ~~†~~	0	
Submarine canyon	~~Width~~ Channel width	300 m - 400 m	195 m - 6.8 km (Shumaker et al., 2018)
	Maximum depth	30 m - 40 m	4 m - 132 m (Shumaker et al., 2018)
	Strike	N0°E - N360°E	
	Migration rate constant ~~*~~	~~50m~~50 m/~~year~~ yr - ~~60m~~60 m/~~year~~ yr	Exaggerated to accelerate simulation; refe range: 2 m/yr - 14 m/yr (Biscara et al., 20
	Dimensionless Chezy's friction factor ~~*~~	0.07 - 0.08	Exaggerated to accelerate simulation pr reference range: 0.002 - 0.005 (Chow, 198
	Iteration time step ~~*~~	0.1 ~~year~~ yr	
	Number of iterations ~~*~~	500 - 2000	
	Natural levee ~~maximum thickness*~~ deposition rate	~~0.5~~5 m/~~iteration~~ yr	Exaggerated to accelerate simulation; refe value: 0.66 m/kyr (Allen et al., 2022)
	Natural levee width	~~6000 m~~ 6 km - ~~8000 m~~ 8 km	Restrained to fit model's extension; refe range: 25 km - 40 km (Klaucke et al., 199
	~~Incision rate *~~ Channel incision rate	8 m/~~year~~ yr	Exaggerated to accelerate simulation; refe

Appendix B: Components of the *cigChannel* dataset

Table B1. Components of the *cigChannel* dataset.

Name	Sample amount	Contents	Features	Example
Meandering channel subset	400	1. Seismic volumes 2. Binary label volumes 3. Seismic impedance volumes	1. Meandering channels only 2. Horizontal, inclined, folded and faulted structures 3. Noise-free	
Tributary channel network subset	400	1. Seismic volumes 2. Binary label volumes 3. Seismic impedance volumes	1. Tributary channel network only 2. Horizontal, inclined, folded and faulted structures 3. Noise-free	
Submarine canyon subset	400	1. Seismic volumes 2. Binary label volumes 3. Seismic impedance volumes 4. Sedimentary facies volumes	1. Submarine canyons only 2. Horizontal, inclined, folded and faulted structures 3. Noise-free	
Assorted channel subset	400	1. Seismic volumes 2. Multi-class label volumes 3. Seismic impedance volumes	1. Meandering channels, tributary channel networks and submarine canyons 2. Horizontal, inclined, folded and faulted structures 3. Noise-free	

Appendix C: Illustrative codes of the dataset generation workflow

```
# Import all functions.
from functions import *

# Number of models.
n_model = 400
# Data generation.
for i in range(n_model):
    # Initialize the model.
    model = GeoModel()
    # Assign P-wave velocities.
    model.add_vp()
    # Add meandering channels.
    model.add_meandering_channel()
    # Add tributary channels.
    model.add_tributary_channel()
    # Add submarine canyons.
    model.add_submarine_canyon()
    # Add inclination.
    model.add_dipping()
    # Add folds.
    model.add_fold()
    # Add faults.
    model.add_faults()
    # Resampling model's z-coordinates.
    model.resample_z()
    # Compute P-wave impedance.
    model.compute_Ip()
    # Compute reflection coefficients.
    model.compute_rc()
    # Make synthetic seismic data.
    model.make_synseis()
    # Save data.
    model.Ip.tofile()  # Impedance volume.
    model.seismic.tofile()  # Seismic volume.
    model.seis_label.tofile()  # Channel label volume.
    model.facies.tofile()  # Sedimentary facies volume.
```

Appendix D: Parameters of the seismic impedance model and Ricker wavelet

Table D1. Parameters of the seismic impedance model, Ricker wavelet and their reference values.

	Parameter	Value
Model extension	X	0 m - 6400 m
	Y	0 m - 6400 m
	Z	0 m - 1280 m
	Grid spacing	25 m $\times$ 25 m $\times$ 5 m (X $\times$ Y $\times$ Z)
Layer	Seismic impedance	7000 m/s.g/cm^3 - 16000 m/s.g/cm^3
	Impedance perturbation	300 m/s.g/cm^3 - 500 m/s.g/cm^3
	Thickness	60 m - 150 m
Meandering channel	Impedance contrast with covering layer (ε)	0 - 1
~~Distributary~~ Tributary channel	Impedance contrast with covering layer (ε)	0 - 1
Submarine canyon	Point bar impedance	6000 m/s.g/cm^3 - 8400 m/s.g/cm^3
	Natural levee impedance	8400 m/s.g/cm^3 - 14400 m/s.g/cm^3
	~~Oxbow lake~~ Abandoned meander impedance	8400 m/s.g/cm^3 - 14400 m/s.g/cm^3
Ricker wavelet	Peak wavenumber	20 km^{-1} - 60 km^{-1}

Author contributions. Guangyu Wang wrote the manuscript and the Python package of the dataset generation workflow. Xinming Wu were involved in conceptualisation and manuscript preparation. Wen Zhang conducted the experiments on field application of this dataset and co-wrote the Application section.

Competing interests. The authors declare no competing interests.

Acknowledgements. ~~The authors thank~~ We acknowledge the USTC supercomputing center for providing computational resources for this project and Jintao Li for providing the Python package *CIGVis* (Li et al., 2025) to visualize the 3D seismic ~~images. The authors also~~ volumes. We appreciate the valuable feedback and suggestions from Andrea Rovida, Samuel Bignardi and one anonymous referee, which have greatly enhance the clarity and rigor of this work. We would also like to thank Hang Gao and Jiarun Yang for their useful suggestions on the training strategy of the U-Net.

References

Allen, C., Peakall, J., Hodgson, D. M., Bradbury, W., and Booth, A. D.: Latitudinal changes in submarine channel-levee system evolution, architecture and flow processes, Frontiers in Earth Science, 10, 976 852, https://doi.org/https://doi.org/10.3389/feart.2022.976852, 2022.

Alqahtani, F. A., Jackson, C. A.-L., Johnson, H. D., and Som, M. R. B.: Controls on the geometry and evolution of humid-tropical fluvial systems: insights from 3D seismic geomorphological analysis of the Malay Basin, Sunda Shelf, Southeast Asia, Journal of Sedimentary Research, 87, 17–40, https://doi.org/https://doi.org/10.2110/jsr.2016.88, 2017.

Bi, Z., Wu, X., Geng, Z., and Li, H.: Deep relative geologic time: A deep learning method for simultaneously interpreting 3-D seismic horizons and faults, Journal of Geophysical Research: Solid Earth, 126, e2021JB021 882, https://doi.org/10.1029/2021JB021882, 2021.

Biscara, L., Mulder, T., Hanquiez, V., Marieu, V., Crespin, J.-P., Braccini, E., and Garlan, T.: Morphological evolution of Cap Lopez Canyon (Gabon): illustration of lateral migration processes of a submarine canyon, Marine Geology, 340, 49–56, https://doi.org/https://doi.org/10.1016/j.margeo.2013.04.014, 2013.

Bond, C. E., Gibbs, A. D., Shipton, Z. K., Jones, S., et al.: What do you think this is¿'Conceptual uncertainty"in geoscience interpretation, GSA today, 17, 4, https://doi.org/10.1130/GSAT01711A.1, 2007.

Bridge, J. S., Jalfin, G. A., and Georgieff, S. M.: Geometry, lithofacies, and spatial distribution of Cretaceous fluvial sandstone bodies, San Jorge Basin, Argentina: outcrop analog for the hydrocarbon-bearing Chubut Group, Journal of Sedimentary Research, 70, 341–359, https://doi.org/10.1306/2DC40915-0E47-11D7-8643000102C1865D, 2000.

Carter, D. C.: 3-D seismic geomorphology: Insights into fluvial reservoir deposition and performance, Widuri field, Java Sea, AAPG bulletin, 87, 909–934, https://doi.org/https://doi.org/10.1306/01300300183, 2003.

Catterall, V., Redfern, J., Gawthorpe, R., Hansen, D., and Thomas, M.: Architectural style and quantification of a submarine channel–levee system located in a structurally complex area: offshore Nile Delta, Journal of Sedimentary Research, 80, 991–1017, https://doi.org/https://doi.org/10.2110/jsr.2010.084, 2010.

Chow, V. t.: Open-channel hydraulics, classical textbook reissue, https://doi.org/https://doi.org/10.1016/c2019-0-03618-7, 1988.

Clark, J. D. and Pickering, K. T.: Architectural elements and growth patterns of submarine channels: application to hydrocarbon exploration, AAPG bulletin, 80, 194–220, https://doi.org/10.1306/64ED878C-1724-11D7-8645000102C1865D, 1996.

Cojan, I., Fouché, O., Lopéz, S., and Rivoirard, J.: Process-based reservoir modelling in the example of meandering channel, Geostatistics Banff 2004, pp. 611–619, https://doi.org/https://doi.org/10.1007/978-1-4020-3610-1_62, 2005.

Covault, J. A., Sylvester, Z., Ceyhan, C., and Dunlap, D. B.: Giant meandering channel evolution, Campos deep-water salt basin, Brazil, Geosphere, 17, 1869–1889, https://doi.org/https://doi.org/10.1130/GES02420.1, 2021.

Crooijmans, R., Willems, C., Nick, H., and Bruhn, D.: The influence of facies heterogeneity on the doublet performance in low-enthalpy geothermal sedimentary reservoirs, Geothermics, 64, 209–219, https://doi.org/https://doi.org/10.1016/j.geothermics.2016.06.004, 2016.

Darmadi, Y., Willis, B., and Dorobek, S.: Three-dimensional seismic architecture of fluvial sequences on the low-gradient Sunda Shelf, offshore Indonesia, Journal of Sedimentary Research, 77, 225–238, https://doi.org/https://doi.org/10.2110/jsr.2007.024, 2007.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L.: Imagenet: A large-scale hierarchical image database, in: 2009 IEEE conference on computer vision and pattern recognition, pp. 248–255, Ieee, https://doi.org/10.1109/CVPR.2009.5206848, 2009.

Deptuck, M. E., Steffens, G. S., Barton, M., and Pirmez, C.: Architecture and evolution of upper fan channel-belts on the Niger Delta slope and in the Arabian Sea, Marine and Petroleum Geology, 20, 649–676, https://doi.org/https://doi.org/10.1016/j.marpetgeo.2003.01.004, 2003.

Deptuck, M. E., Sylvester, Z., Pirmez, C., and O'Byrne, C.: Migration–aggradation history and 3-D seismic geomorphology of submarine channels in the Pleistocene Benin-major Canyon, western Niger Delta slope, Marine and Petroleum Geology, 24, 406–433, https://doi.org/10.1016/j.marpetgeo.2007.01.005, 2007.

Donovan, M., Belmont, P., and Sylvester, Z.: Evaluating the relationship between meander-bend curvature, sediment supply, and migration rates, Journal of Geophysical Research: Earth Surface, 126, e2020JF006 058, https://doi.org/https://doi.org/10.1029/2020JF006058, 2021.

Edmonds, D. A. and Slingerland, R. L.: Mechanics of river mouth bar formation: Implications for the morphodynamics of delta distributary networks, Journal of Geophysical Research: Earth Surface, 112, https://doi.org/10.1029/2006JF000574, 2007.

Englert, R., Hubbard, S., Matthews, W., Coutts, D., and Covault, J.: The evolution of submarine slope-channel systems: Timing of incision, bypass, and aggradation in Late Cretaceous Nanaimo Group channel-system strata, British Columbia, Canada, Geosphere, 16, 281–296, https://doi.org/https://doi.org/10.1130/GES02091.1, 2020.

Frascati, A. and Lanzoni, S.: Morphodynamic regime and long-term evolution of meandering rivers, Journal of Geophysical Research: Earth Surface, 114, https://doi.org/https://doi.org/10.1029/2008JF001101, 2009.

Gao, H., Wu, X., and Liu, G.: ChannelSeg3D: Channel simulation and deep learning for channel interpretation in 3D seismic images, Geophysics, 86, IM73–IM83, https://doi.org/10.1190/geo2020-0572.1, 2021.

García, M., Alonso, B., Ercilla, G., and Gràcia, E.: The tributary valley systems of the Almeria Canyon (Alboran Sea, SW Mediterranean): sedimentary architecture, Marine Geology, 226, 207–223, https://doi.org/https://doi.org/10.1016/j.margeo.2005.10.002, 2006.

Gee, M., Gawthorpe, R. L., Bakke, K., and Friedmann, S.: Seismic geomorphology and evolution of submarine channels from the Angolan continental margin, Journal of Sedimentary Research, 77, 433–446, https://doi.org/10.2110/jsr.2007.042, 2007.

Geleynse, N., Storms, J. E., Walstra, D.-J. R., Jagers, H. A., Wang, Z. B., and Stive, M. J.: Controls on river delta formation; insights from numerical modelling, Earth and Planetary Science Letters, 302, 217–226, https://doi.org/10.1016/j.epsl.2010.12.013, 2011.

Gibling, M. R.: Width and thickness of fluvial channel bodies and valley fills in the geological record: a literature compilation and classification, Journal of sedimentary Research, 76, 731–770, https://doi.org/https://doi.org/10.2110/jsr.2006.060, 2006.

Heim, J. A., Vasconcelos, P. M., Shuster, D. L., Farley, K. A., and Broadbent, G.: Dating paleochannel iron ore by (U-Th)/He analysis of supergene goethite, Hamersley province, Australia, Geology, 34, 173–176, https://doi.org/https://doi.org/10.1130/G22003.1, 2006.

Hein, F. J. and Cotterill, D. K.: The Athabasca oil sands—a regional geological perspective, Fort McMurray area, Alberta, Canada, Natural Resources Research, 15, 85–102, https://doi.org/10.1007/s11053-006-9015-4, 2006.

Heo, J., Duc, T. A., Cho, H.-S., and Choi, S.-U.: Characterization and prediction of meandering channel migration in the GIS environment: A case study of the Sabine River in the USA, Environmental monitoring and assessment, 152, 155–165, https://doi.org/https://doi.org/10.1007/s10661-008-0304-8, 2009.

Howard, A. D. and Knutson, T. R.: Sufficient conditions for river meandering: A simulation approach, Water Resources Research, 20, 1659–1667, https://doi.org/10.1029/WR020i011p01659, 1984.

Ji, S., Xu, W., Yang, M., and Yu, K.: 3D convolutional neural networks for human action recognition, IEEE transactions on pattern analysis and machine intelligence, 35, 221–231, https://doi.org/10.1109/TPAMI.2012.59, 2012.

Jobe, Z., Sylvester, Z., Bolla Pittaluga, M., Frascati, A., Pirmez, C., Minisini, D., Howes, N., and Cantelli, A.: Facies architecture of submarine channel deposits on the western Niger Delta slope: Implications for grain-size and density stratification in turbidity currents, Journal of Geophysical Research: Earth Surface, 122, 473–491, https://doi.org/https://doi.org/10.1002/2016JF003903, 2017.

Kane, I. A., Kneller, B. C., Dykstra, M., Kassem, A., and McCaffrey, W. D.: Anatomy of a submarine channel–levee: an example from Upper Cretaceous slope sediments, Rosario Formation, Baja California, Mexico, Marine and Petroleum Geology, 24, 540–563, https://doi.org/https://doi.org/10.1016/j.marpetgeo.2007.01.003, 2007.

Kang, F., Yang, X., Wang, X., Zheng, T., Bai, T., Liu, Z., and Sui, H.: Hydrothermal features of a sandstone geothermal reservoir in the north Shandong plain, China, Lithosphere, 2021, 1675 798, https://doi.org/https://doi.org/10.2113/2022/1675798, 2022.

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

Klaucke, I., Hesse, R., and Ryan, W.: Seismic stratigraphy of the Northwest Atlantic Mid-Ocean Channel: growth pattern of a mid-ocean channel-levee complex, Marine and Petroleum Geology, 15, 575–585, https://doi.org/https://doi.org/10.1016/S0264-8172(98)00044-0, 1998.

Kolla, V., Posamentier, H., and Wood, L.: Deep-water and fluvial sinuous channels—Characteristics, similarities and dissimilarities, and modes of formation, Marine and Petroleum Geology, 24, 388–405, https://doi.org/10.1016/j.marpetgeo.2007.01.007, 2007.

Leigh, D. S. and Feeney, T. P.: Paleochannels indicating wet climate and lack of response to lower sea level, southeast Georgia, Geology, 23, 687–690, https://doi.org/10.1130/0091-7613(1995)023<0687:PIWCAL>2.3.CO;2, 1995.

Li, J., Shi, Y., and Wu, X.: CIGVis: An open-source Python tool for the real-time interactive visualization of multidimensional geophysical data, Geophysics, 90, F1–F10, https://doi.org/https://doi.org/10.1190/geo2024-0041.1, 2025.

Li, X., Chen, Q., Wu, C., Liu, H., and Fang, Y.: Application of multi-seismic attributes analysis in the study of distributary channels, Marine and Petroleum Geology, 75, 192–202, https://doi.org/10.1016/j.marpetgeo.2016.04.016, 2016.

Liang, M., Voller, V., and Paola, C.: A reduced-complexity model for river delta formation–Part 1: Modeling deltas with channel dynamics, Earth Surface Dynamics, 3, 67–86, https://doi.org/10.5194/esurf-3-67-2015, 2015.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L.: Microsoft coco: Common objects in context, in: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, pp. 740–755, Springer, https://doi.org/10.1007/978-3-319-10602-1_48, 2014.

Manshor, N. A. and Amir Hassan, M. H.: Seismic geomorphological analysis of channel types: a case study from the Miocene Malay Basin, Journal of Geophysics and Engineering, 20, 159–171, https://doi.org/https://doi.org/10.1093/jge/gxac103, 2023.

McDonald, N.: Simple Particle-based Hydraulic Erosion, https://nickmcd.me/2020/04/10/simple-particle-based-hydraulic-erosion/, 2020a.

McDonald, N.: soillib, https://github.com/erosiv/soillib, 2020b.

Miall, A.: The facies and architecture of fluvial systems, Fluvial depositional systems, pp. 9–68, https://doi.org/https://doi.org/10.1007/978-3-319-00666-6_2, 2014.

Mueller, E. R. and Pitlick, J.: Sediment supply and channel morphology in mountain river systems: 1. Relative importance of lithology, topography, and climate, Journal of Geophysical Research: Earth Surface, 118, 2325–2342, https://doi.org/https://doi.org/10.1002/2013JF002843, 2013.

Nettleton, D. F., Orriols-Puig, A., and Fornells, A.: A study of the effect of different types of noise on the precision of supervised learning techniques, Artificial intelligence review, 33, 275–306, https://doi.org/10.1007/s10462-010-9156-z, 2010.

Noah, J. T., Hofland, G., and Lemke, K.: Seismic interpretation of meander channel point-bar deposits using realistic seismic modeling techniques, The Leading Edge, 11, 13–18, https://doi.org/https://doi.org/10.1190/1.1436890, 1992.

Nordfjord, S., Goff, J. A., Austin Jr, J. A., and Sommerfield, C. K.: Seismic geomorphology of buried channel systems on the New Jersey outer shelf: assessing past environmental conditions, Marine Geology, 214, 339–364, https://doi.org/10.1016/j.margeo.2004.10.035, 2005.

Normark, W. R., Posamentier, H., and Mutti, E.: Turbidite systems: state of the art and future directions, Reviews of Geophysics, 31, 91–116, https://doi.org/https://doi.org/10.1029/93RG02832, 1993.

Normark, W. R., Carlson, P. R., Chan, M., and Archer, A.: Giant submarine canyons: Is size any clue to their importance in the rock record?, Special Papers-Geological Society of America, pp. 175–190, 2003.

Oraby, E., Eksteen, J., Karrech, A., and Attar, M.: Gold extraction from paleochannel ores using an aerated alkaline glycine lixiviant for consideration in heap and in-situ leaching applications, Minerals Engineering, 138, 112–118, https://doi.org/https://doi.org/10.1016/j.mineng.2019.04.023, 2019.

Paull, C., Caress, D., Lundsten, E., Gwiazda, R., Anderson, K., McGann, M., Conrad, J., Edwards, B., and Sumner, E.: Anatomy of the La Jolla submarine canyon system; offshore Southern California, Marine Geology, 335, 16–34, https://doi.org/https://doi.org/10.1016/j.margeo.2012.10.003, 2013.

Payenberg, T. H. D. and Lang, S. C.: Reservoir geometry of fluvial distributary channels—Implications for Northwest Shelf, Australia, deltaic successions, The APPEA Journal, 43, 325–338, https://doi.org/10.1071/AJ02017, 2003.

Pechenizkiy, M., Tsymbal, A., Puuronen, S., and Pechenizkiy, O.: Class noise and supervised learning in medical domains: The effect of feature extraction, in: 19th IEEE symposium on computer-based medical systems (CBMS'06), pp. 708–713, IEEE, https://doi.org/10.1109/CBMS.2006.65, 2006.

Pham, N., Fomel, S., and Dunlap, D.: Automatic channel detection using deep learning, Interpretation, 7, SE43–SE50, https://doi.org/10.1190/INT-2018-0202.1, 2019.

Revil, A., Cary, L., Fan, Q., Finizola, A., and Trolard, F.: Self-potential signals associated with preferential ground water flow pathways in a buried paleo-channel, Geophysical Research Letters, 32, https://doi.org/https://doi.org/10.1029/2004GL022124, 2005.

Ronneberger, O., Fischer, P., and Brox, T.: U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp. 234–241, Springer, https://doi.org/10.1007/978-3-319-24574-4_28, 2015.

Samadder, R. K., Kumar, S., and Gupta, R. P.: Paleochannels and their potential for artificial groundwater recharge in the western Ganga plains, Journal of Hydrology, 400, 154–164, https://doi.org/https://doi.org/10.1016/j.jhydrol.2011.01.039, 2011.

Schook, D. M., Rathburn, S. L., Friedman, J. M., and Wolf, J. M.: A 184-year record of river meander migration from tree rings, aerial imagery, and cross sections, Geomorphology, 293, 227–239, https://doi.org/https://doi.org/10.1016/j.geomorph.2017.06.001, 2017.

Seybold, H., Andrade Jr, J. S., and Herrmann, H. J.: Modeling river delta formation, Proceedings of the National Academy of Sciences, 104, 16 804–16 809, https://doi.org/10.1073/pnas.0705265104, 2007.

Shumaker, L. E., Jobe, Z. R., Johnstone, S. A., Pettinga, L. A., Cai, D., and Moody, J. D.: Controls on submarine channel-modifying processes identified through morphometric scaling relationships, Geosphere, 14, 2171–2187, https://doi.org/https://doi.org/10.1130/GES01674.1, 2018.

Sylvester, Z.: meanderpy, https://github.com/zsylvester/meanderpy, 2021.

Sylvester, Z., Pirmez, C., and Cantelli, A.: A model of submarine channel-levee evolution based on channel trajectories: Implications for stratigraphic architecture, Marine and Petroleum Geology, 28, 716–727, https://doi.org/https://doi.org/10.1016/j.marpetgeo.2010.05.012, 2011.

Sylvester, Z., Durkin, P., and Covault, J. A.: High curvatures drive river meandering, Geology, 47, 263–266, https://doi.org/10.1130/G45608.1, 2019.

Sylvia, D. A. and Galloway, W. E.: Morphology and stratigraphy of the late Quaternary lower Brazos valley: Implications for paleo-climate, discharge and sediment delivery, Sedimentary Geology, 190, 159–175, https://doi.org/10.1016/j.sedgeo.2006.05.023, 2006.

Trigg, M. A., Bates, P. D., Wilson, M. D., Schumann, G., and Baugh, C.: Floodplain channel morphology and networks of the middle Amazon River, Water Resources Research, 48, https://doi.org/https://doi.org/10.1029/2012WR011888, 2012.

Vizeu, F., Zambrini, J., Tertois, A.-L., da Graça e Costa, B. d. A., Fernandes, A. Q., and Canning, A.: Synthetic seismic data generation for automated AI-based procedures with an example application to high-resolution interpretation, The Leading Edge, 41, 392–399, https://doi.org/10.1190/tle41060392.1, 2022.

Wang, G., Wu, X., and Zhang, W.: cigChannel: A massive-scale 3D seismic dataset with labeled paleochannels for advancing deep learning in seismic interpretation, https://doi.org/10.5281/zenodo.10791151, 2024.

Wang, Z., Yin, C., Fan, T., and Lei, X.: Seismic geomorphology of a channel reservoir in lower Minghuazhen Formation, Laizhouwan subbasin, China, Geophysics, 77, B187–B195, https://doi.org/https://doi.org/10.1190/geo2011-0209.1, 2012.

Wild, T. B., Loucks, D. P., and Annandale, G. W.: SedSim: a river basin simulation screening model for reservoir management of sediment, water, and hydropower, Journal of Open Research Software, 7, https://doi.org/https://doi.org/10.5334/jors.261, 2019.

Wood, L. J.: Quantitative seismic geomorphology of Pliocene and Miocene fluvial systems in the northern Gulf of Mexico, USA, Journal of Sedimentary Research, 77, 713–730, https://doi.org/https://doi.org/10.2110/jsr.2007.068, 2007.

Wu, J., Chen, Q., Gui, Z., and Bai, M.: Fast dictionary learning for 3D simultaneous seismic data reconstruction and denoising, Journal of Applied Geophysics, 194, 104 446, https://doi.org/https://doi.org/10.1016/j.jappgeo.2021.104446, 2021.

Wu, X., Liang, L., Shi, Y., and Fomel, S.: FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation, Geophysics, 84, IM35–IM45, https://doi.org/10.1190/geo2018-0646.1, 2019.

Wu, X., Geng, Z., Shi, Y., Pham, N., Fomel, S., and Caumon, G.: Building realistic structure models to train convolutional neural networks for seismic structural interpretation, Geophysics, 85, WA27–WA39, https://doi.org/10.1190/geo2019-0375.1, 2020a.

Wu, X., Yan, S., Qi, J., and Zeng, H.: Deep learning for characterizing paleokarst collapse features in 3-D seismic images, Journal of Geophysical Research: Solid Earth, 125, e2020JB019 685, https://doi.org/10.1029/2020JB019685, 2020b.

Xie, S. and Tu, Z.: Holistically-nested edge detection, in: Proceedings of the IEEE international conference on computer vision, pp. 1395–1403, https://doi.org/https://doi.org/10.1109/iccv.2015.164, 2015.

Zeng, H., Zhu, X., Liu, Q., Zhu, H., and Xu, C.: An alternative, seismic-assisted method of fluvial architectural-element analysis in the subsurface: Neogene, Shaleitian area, Bohai Bay Basin, China, Marine and Petroleum Geology, 118, 104 435, https://doi.org/https://doi.org/10.1016/j.marpetgeo.2020.104435, 2020.

Zhang, H., Chen, X.-H., and Zhang, L.-Y.: 3D simultaneous seismic data reconstruction and noise suppression based on the curvelet transform, Applied Geophysics, 14, 87–95, https://doi.org/https://doi.org/10.1007/s11770-017-0607-z, 2017.

Zhang, Z., Li, H., Yan, Z., Jing, J., and Gu, H.: Deep carbonate fault–karst reservoir characterization by multi-task learning, Geophysical Prospecting, 72, 1092–1106, https://doi.org/10.1111/1365-2478.13460, 2024.

Zheng, Y., Zhang, Q., Yusifov, A., and Shi, Y.: Applications of supervised deep learning for seismic interpretation and inversion, The Leading Edge, 38, 526–533, https://doi.org/10.1190/tle38070526.1, 2019.

Zhuo, H., Wang, Y., Shi, H., He, M., Chen, W., Li, H., Wang, Y., and Yan, W.: Contrasting fluvial styles across the mid-Pleistocene climate transition in the northern shelf of the South China Sea: Evidence from 3D seismic data, Quaternary Science Reviews, 129, 128–146, https://doi.org/https://doi.org/10.1016/j.quascirev.2015.10.012, 2015.