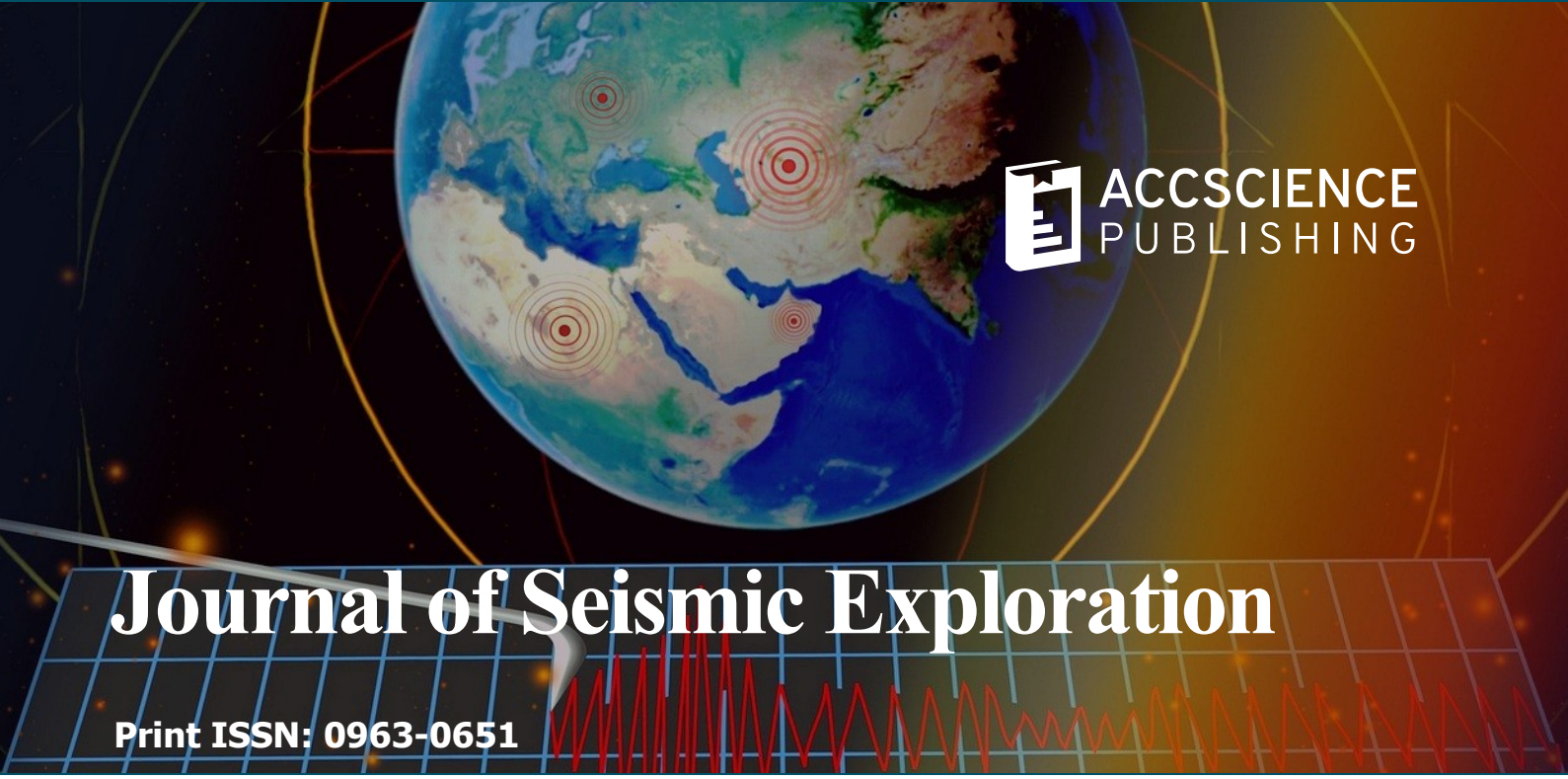
Journal of Seismic Exploration

ISSN: 0963-0651

Volume 34 · Issue 6
2025

ACCSCIENCE PUBLISHING

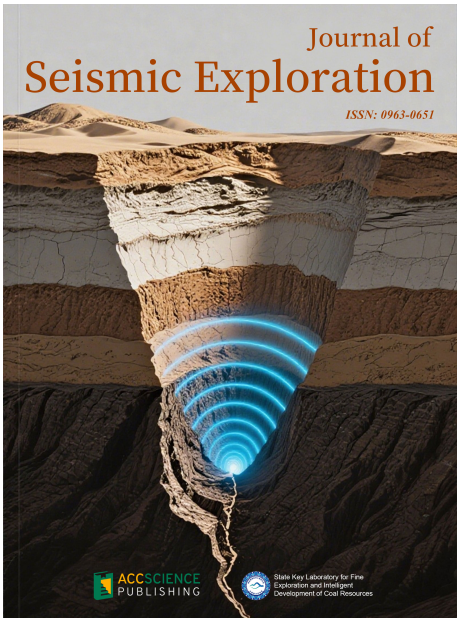
State Key Laboratory for Fine Exploration and Intelligent Development of Coal Resources



Journal of Seismic Exploration

Print ISSN: 0963-0651

The Journal of Seismic Exploration (JSE) is an international medium for the publication of research in seismic modeling, processing, inversion, interpretation, field techniques, borehole techniques, tomography, instrumentation and software. The aim and goal of JSE is to serve the interests and objectives of seismic petroleum exploration. All seismic processing methods have assumptions and prerequisites- and when they are satisfied methods are effective-and when they are violated methods become ineffective and can fail. The latter contributes to drilling dry hole exploration wells and suboptimal development wells.



About the Publisher

AccScience Publishing is a publishing company based in Singapore. We publish a range of high-quality, open-access, peer-reviewed journals and books from a broad spectrum of disciplines.

Contact Us

Managing Editor
jse.office@accscience.sg

AccScience Publishing
9 Raffles Place, Republic Plaza 1 #06-00 Singapore 048619.

Volume 34 • Issue 6 • 2025
ISSN 0963-0651 (print)

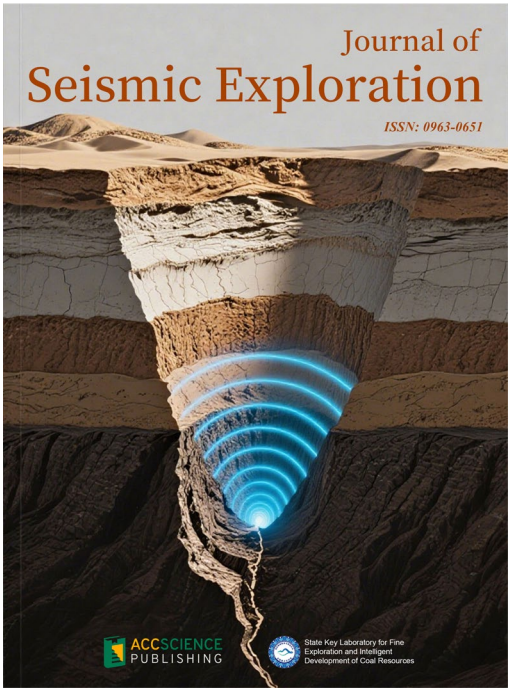
Journal of Seismic Exploration

ACCSCIENCE PUBLISHING

Access Science Without Barriers

JOURNAL OF SEISMIC EXPLORATION

ISSN: 0963-0651 (print)



Editorial and Production Credits

Publisher: AccScience Publishing
Managing Editor: Gary Liu
Production Editor: Lim Mingyuan
Article Layout and Typeset: Sinjore Technologies (India)

Supplementary file

Supplementary files of articles can be obtained at https://accscience.com/journal/JSE/34/6.

Disclaimer

AccScience Publishing is not liable to the statements, perspectives, and opinions contained in the publications. The appearance of advertisements in the journal shall not be construed as a warranty, endorsement, or approval of the products or services advertised and/or the safety thereof. AccScience Publishing disclaims responsibility for any injury to persons or property resulting from any ideas or products referred to in the publications or advertisements. AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Seismic Exploration

Editorial Board

ACCSCIENCE PUBLISHING

Volume 34 | Issue 6 | 2025 | Page 1 – 96 |

Journal of Seismic Exploration

https://accscience.com/journal/JSE

CONTENTS

ARTICLES

ACCSCIENCE PUBLISHING

Journal of Seismic Exploration

ARTICLE

A precise picking method for seismic first arrivals based on the residual long short-term memory network driven by time-frequency dual domain data

Ziyu Qin[1], Xianju Zheng[1]*, and Wenhua Wang[2]

[1]Department of Software Engineering, School of Computer Engineering, Chengdu Technological University, Chengdu, Sichuan, China

[2]Department of Intelligence Science and Technology, School of Computer Science, Chengdu Normal University, Chengdu, Sichuan, China

Abstract

First-arrival picking of seismic data is one of the key steps in seismic data processing. When seismic data have low signal-to-noise ratio (SNR) and weak first-arrival energy, accurately and efficiently picking first arrivals remain a critical challenge for most automatic picking methods. To address this issue, this paper proposes a Multi-perspective Residual Long Short-Term Memory (M-Res-LSTM) network. This network integrates the spatial feature extraction advantage of Residual Networks and the temporal dynamic modeling capability of LSTM networks, while introducing a coordinate attention mechanism. Through multi-perspective learning in both time and frequency domains, it effectively improves the reliability of automatic first-arrival picking. First, this paper elaborates on the core principle of the M-Res-LSTM network for automatic first-arrival picking: the amplitude, frequency, and phase features of seismic data are used as network inputs, and the accurately picked first arrivals manually serve as network outputs. After training the network using a supervised learning approach, the well-trained model is applied to perform automatic first-arrival picking. Second, an analysis of the network's hyperparameters is conducted to determine the optimal parameter configuration. Finally, automatic first-arrival picking tests are carried out on seismic datasets with different characteristics, and the picking results are compared with those obtained by the energy ratio method, single-input Res-LSTM, and Swin-Transformer. The results demonstrate that the proposed M-Res-LSTM method maintains good stability and accuracy even in complex scenarios with low first-arrival energy and poor SNR.

Keywords: Automatic first-arrival picking; Time-frequency dual domain; Multi-perspective learning; Res-LSTM; Attention mechanism

***Corresponding author:**
Xianju Zheng
(zhengxju@cdtu.edu.cn)

Citation: Qin Z, Zheng X, Wang W. A precise picking method for seismic first arrivals based on the residual long short-term memory network driven by time-frequency dual domain data. *J Seismic Explor*. 2025;34(6):1-15.
doi: 10.36922/JSE025330058

Received: August 16, 2025

1st revised: September 1, 2025

2nd revised: September 8, 2025

Accepted: October 29, 2025

Published online: November 18, 2025

1. Introduction

First-arrival waves refer to the seismic waves that propagate through subsurface media and reach geophones first, typically existing in the form of direct waves or refracted waves. In the seismic data processing workflow, the travel time of first-arrival waves is

of crucial significance, as it can provide core foundational data for near-surface inversion work. Although the method of manual picking of the first arrival has high accuracy, it is time-consuming and labor-intensive. Moreover, the accuracy of the picking is also affected by the experience of the interpreters.[1] To pick the first arrivals efficiently and accurately and reduce the workload of data processing personnel, scholars have proposed different semi-automatic or automatic first-arrival picking methods.

Given the similar characteristics of adjacent traces, existing studies have proposed a method to determine the first-arrival time through cross-correlation operations between adjacent traces.[2-4] The selection of the standard trace has a great influence on the first-arrival picking results of this type of method, and its effect needs to be further improved when the signal-to-noise ratio (SNR) of the data is low. The algorithm based on energy characteristics possesses robust anti-noise performance and achieves favorable automatic picking results, and has also achieved good results in the processing of actual data.[5-8] Since this method is greatly affected by the window length, many picking methods based on multiple time windows have also been developed.[9,10] Another common method is the Akaike Information Criterion (AIC), which discriminates the first arrivals using the difference in AIC values between the seismic signal and the noise.[11-13] However, the picking results are not satisfactory under a low SNR. The fractal dimension algorithm[14-17] has relatively good anti-noise ability, but it does not take into account the similarity between seismic traces. Clustering-based methods identify first arrivals in an unsupervised manner based on the characteristic differences between first arrivals and noise. However, the inherent temporal connections between different subsequences are not considered, making it difficult to distinguish low-amplitude signals from noise under low SNR conditions.[18,19]

Transforming seismic signals into other domains or spaces can further highlight the differences between first-arrival waves and noises. Performing τ-p transformation,[20] wavelet transformation,[21] shearlet transformation,[22-24] etc., are also commonly used means to enhance the precision of first-arrival picking; Beyond the common shot gather, picking first arrivals on common offset gathers and common receiver gathers also offers distinct advantages.[5,25] In addition to time-space domain signals, neural networks incorporate attributes such as amplitude, frequency, time-frequency characteristics, short-term average/long-term average (STA/LTA) ratios, and data distribution skewness as inputs,[26-31] thereby facilitating more effective capture of relevant features by the networks.

Deep learning can efficiently extract the internal laws of data, construct multi-level data representations, and is more friendly to massive data. Leveraging the capabilities of generative adversarial networks,[32] convolutional neural networks,[1,33] UNet,[34-37] recurrent neural networks, meta-learning,[38,39] transformers,[40-42] transfer learning,[43] and various hybrid networks[38,44] in extracting complex features from seismic data, deep learning algorithms have emerged as a critical force in the picking of seismic first arrivals. Similar to conventional methods, to further improve the picking accuracy, scholars have attempted to input data with different attributes into the network, such as time-frequency domain data and STA/LTA feature maps.[44,45] The rich data features provide more information references for the model.

Studies indicate that current first-arrival picking operations are typically performed exclusively within either the time-space domain or a single transformed domain. The energy features and time-frequency features of seismic data serve as valid criteria for first-arrival identification, and the integration of multiple sets of feature data can, further, enhance the precision of first-arrival extraction. Building on this insight, this study comprehensively leverages the data features from both the time domain and frequency domain, proposes a Multi-perspective Residual Long Short-Term Memory (M-Res-LSTM) network tailored for first-arrival picking, and elaborates in detail on the complete workflow of automatic first-arrival picking. Finally, experiments on automatic first-arrival picking were conducted on real seismic datasets with distinct characteristics using this network, thereby verifying the effectiveness of the proposed method.

2. Methods

2.1. Network architecture

To fully leverage the spatiotemporal and time-frequency characteristics of seismic signals, this study proposes the M-Res-LSTM network for automatic first-arrival picking. Figure 1 shows its structural schematic with key features.

2.1.1. Branch structure

The network comprises four parallel branch modules, each consisting of m residual modules and n residual LSTM modules. These branches process four types of two-dimensional (2D) input data, namely, time-space domain seismic signals, instantaneous amplitude, frequency, and phase features, thereby enabling multi-perspective extraction of seismic information. A parameter-sharing mechanism between LSTM branches enhances training efficiency and generalization.

Residual modules retain spatial information through shortcut paths, mitigating degradation in deep networks through direct gradient flow.[46] For seismic data, this

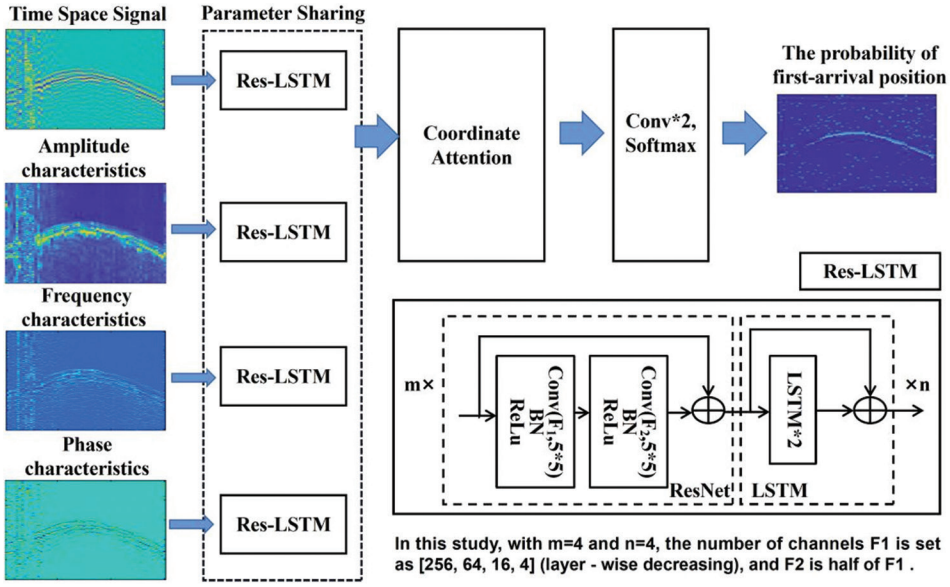


Figure 1. Schematic diagram of the M-Res-LSTM network model structure. "m" and "n" represent the numbers of Residual Networks and LSTM modules, respectively.
Abbreviation: LSTM: Long Short-Term Memory; M-Res-LSTM: Multi-perspective Residual Long Short-Term Memory; Res-LSTM: Residual Long Short-Term Memory.

preserves inter-trace correlations and nonlinear spatial features. For input data $X_k \in R_{H\times W}, k=1,2,3,4$, the output of the i-th residual layer is:

$$Z_i^{res} = X_k + F_i\left(X_k, \theta_{res}^i\right), i=1,2\cdots m \quad \text{(I)}$$

Where F_i denotes nonlinear transformations (convolution, batch normalization, and activation) with parameters θ_{res}^i .

Following residual processing, outputs are converted to sequences for residual LSTM modules. LSTM's gating mechanisms (input, forget, and output) capture temporal dynamics. For n residual LSTM layers, the recurrence relation is:

$$h_j^t = x_{j-1}^t + LSTM\left(x_t^j, x_j^t, \phi_j\right), j=1,2\cdots n \quad \text{(II)}$$

where h_j^t is the hidden state of the j-th layer at time t, ϕ_j are layer parameters, and the n-th layer output h_n^t represents temporal features, which are reshaped to 2D for fusion.

2.1.2. Coordinate attention module

To adapt to the requirements of the first-arrival picking task, multi-domain features extracted from the four branches are fused through the coordinate attention mechanism.[47] Unlike the standard coordinate attention, the proposed method first extracts multi-dimensional visual features tailored to the task characteristics, then assigns adaptive weights to enhance critical information. Meanwhile, it not only additionally designs a convolutional gating structure specifically for suppressing high-amplitude noise in seismic data but also further integrates a residual connection to prevent first-arrival signals from being excessively suppressed. The detailed process is as follows:

First, global pooling is performed on the concatenated feature $F \in R_{H\times W\times C}$ along the width (W) and height (H) directions:

$$z_h(k) = \frac{1}{W}\sum_{0\le j<W} F(h,i,k), z_w(\omega) = \frac{1}{H}\sum_{0\le j<H} F(j,\omega,k) \quad \text{(III)}$$

After aligning the dimensions of x_w with x_h through a transposition operation, the two are concatenated along the height direction. The concatenated result is processed by a custom activation function and 1×1 convolution to obtain the feature m. The processed feature m is split into height-related and width-related components. Meanwhile, a convolutional gating structure is designed based on the original branch features to generate a screening mask g_i. Finally, branch attention weights are generated and fused:

$$branch\_att = \frac{a_h + a_w}{2} \quad g_i \quad \text{(IV)}$$

Here, a_h and a_w are the branch attention weights, respectively; and e denotes element-wise multiplication.

Each branch feature is multiplied by its corresponding attention weight, and the weighted branch features are summed to obtain a fused representation:

$$F_{fused} = \sum_{i=1}^{4} \left(b_i \bullet branch\_att_i \right) \quad \text{(V)}$$

Where b_i is the feature of the i-th branch, and $branch\_att_i$ is its corresponding attention weight.

This process not only retains the feature advantages of each branch in specific domains but also mitigates the interference of redundant information and noise through weight modulation.

To further focus on regions with concentrated first-arrival wave energy, coordinate attention computation is re-applied to F_{fused}: first, global average pooling is performed along the width and height directions, with the pooling formulas as follows:

$$z'_h(k) = \frac{1}{W} \sum_{0 \le j < W} F_{fused}(h,i,k), z'_w(\omega) = \frac{1}{H} \sum_{0 \le j < H} F_{fused}(j,\omega,k) \quad \text{(VI)}$$

After dimension alignment, feature concatenation, channel compression, and component splitting, the base spatial attention weights c_h and c_w are obtained. Concurrently, a convolutional gating structure is designed based on F_{fused} to generate the spatial screening mask g_s. The spatially attentive weights with enhanced noise robustness are derived through the following formulas:

$$\alpha = c_h \quad g_s, \beta = c_w \quad g_s \quad \text{(VII)}$$

Finally, the spatial attention weights are combined with F_{fused} through residual connection to enhance the signals in key regions. The calculation formula for the final output feature $\hat{F}$ is as follows:

$$\check{F}(h,\omega,k) = F_{fused}(h,i,k) + F_{fused}(h,i,k) \cdot \alpha(h,k) \cdot \beta(\omega,k) \quad \text{(VIII)}$$

By generating screening masks with the same dimension as the attention weights through convolutional gating, this method can effectively suppress high-amplitude noise in seismic data, accurately focus on regions with significant first-arrival wave energy variations, and remarkably improve the processing performance of seismic data with low SNR.

2.1.3. Output layer

Fused features $\hat{F}$ are passed through two convolution layers and a Softmax activation to predict first-arrival positions:

$$P = Soft\max\left(Conv_2\left(Conv_1\left(\hat{F}, \omega_1 \right), \omega_2 \right) \right) \quad \text{(IX)}$$

Where ω_1, ω_2 are convolution parameters.

The network takes the amplitude, frequency, and phase characteristics of seismic data as inputs and incorporates a coordinate attention mechanism to achieve feature fusion across different branches. By assigning distinct weights to multiple features, it allocates varying levels of attention to them, thereby enhancing task-critical features while suppressing those irrelevant to the current task. This mechanism effectively mitigates the mispicking of first arrivals.

In recent studies, a multitude of innovative networks have been proposed for seismic first-arrival picking, including those based on Transformer,[40] meta-learning,[38] and multi-stage network[45] architectures. Table 1 presents a comparison between the method proposed in this study and the aforementioned methods, focusing on their characteristics including network architecture, input, advantages, and dependency conditions.

Compared with other networks, M-Res-LSTM still possesses unique characteristics and advantages:

(i) In terms of feature input types, M-Res-LSTM innovatively incorporates amplitude, frequency, and phase information, providing more comprehensive feature support for first-arrival picking

(ii) M-Res-LSTM introduces the coordinate attention mechanism, which includes a time-frequency domain weight allocation layer. By calculating the weights of time-domain and frequency-domain features, it can more accurately capture key information in the time-frequency domain and improve the ability to identify weak first-arrival signals and the like

(iii) In terms of feature fusion methods, M-Res-LSTM employs parallel branches and adaptive weight fusion, enabling collaborative optimization of multi-domain features such as time and frequency, thus enhancing the effectiveness and flexibility of feature fusion.

2.2. Dataset and training

To apply the M-Res-LSTM network for automatic first-arrival picking, seismic data are first transformed to generate profiles containing amplitude, phase, and frequency features. Subsequently, these feature profiles are partitioned into three subsets: the training set, validation set, and test set, with the respective proportions accounting for 80%, 10%, and 10% of the total dataset, respectively. Finally, the network is trained using the training set, the optimal generalization of the model is achieved on the

Table 1. Comparison of different network features

Comparison	M-Res-LSTM	Res-LSTM	Swin-Transformer[40]	Meta-learning[38]	MSSPN[45]
Architecture	Multi-branch parallel architecture and realizing dynamic fusion of time-frequency dual-domain features via the coordinate attention mechanism	Single-branch cascaded architecture	U-shaped hierarchical self-attention architecture	Dual-loop meta-training framework	Four-stage cascaded segmentation architecture
Input	Spatiotemporal signals, amplitude, frequency, phase	Spatiotemporal domain signals	Spatiotemporal domain signals	Spatiotemporal domain signals	Spatiotemporal domain signals and STA/LTA feature maps
Advantages	Multi-input supports the attention mechanism for accurate noise suppression; residual-temporal modeling adapts to seismic wave propagation	Concise architecture, low computational overhead, and easy reproduction	SW-MSA adapts to drastic changes in local first arrivals; Dilated convolution expands the receptive field	Only 5–20% of manual labels are required, reducing costs; weight matrix filters label noise	VCTE effectively narrows the first-arrival range; mixed loss enhances the continuity of first arrivals
Dependency Conditions	High-quality labels	High-quality labels	High-quality labels	Allowing partial low-quality labels	Requiring prior information and high-quality labels

Abbreviations: LTA: Long-term average; LSTM: Long-Short-Term Memory; M-Res-LSTM: Multi-perspective Residual Long-Short-Term Memory; MSSPN: Multistage segmentation picking network; Res-LSTM: Residual Long-Short Term Memory; SW-MSA: Shifted window-multihead self-attention; STA: Short-term average; VCTE: Velocity-constrained trend estimation.

validation set, and the first-arrival picking test is carried out on the test set. In this study, common shot gather records are selected as the original seismic data. Theoretically, common receiver gather records, common offset gather records, and common midpoint gather records are all optional. The main steps of the model training process are described as follows.

2.2.1. Generation of feature profiles

The three instantaneous profiles calculated using the Hilbert transform can reflect various characteristics of seismic data. The analytic signal of a seismic signal can be expressed as:[48]

$$\tilde{x}(t) = x(t) + ix^{*}(t) \tag{X}$$

Where $x(t)$ denotes the real component of the complex trace, corresponding to the actual seismic record trace; $x^{*}(t)$ represents the imaginary component of the complex trace, which is orthogonal to the real component.

Instantaneous amplitude reflects the energy intensity of a seismic signal at a certain moment, which is usually related to the reflection coefficient of the stratum. A high instantaneous amplitude may indicate lithologic abrupt changes or the presence of fluids.[49] Instantaneous amplitude can be expressed as:

$$A(t) = \sqrt{x^{2}(t) + x^{*2}(t)} \tag{XI}$$

Instantaneous phase describes the phase state of a signal, which can identify stratal continuity and structural features. Abrupt changes may indicate faults or unconformities:

$$\theta(t) = tg^{-1}\left[x^{*}(t) / x(t)\right] \tag{XII}$$

Instantaneous frequency is the time rate of change of the phase and can assist in identifying lithology and predicting reservoirs:

$$\omega(t) = d\theta(t) / dt \tag{XIII}$$

In the first-arrival picking process, the energy mutation points of instantaneous amplitude facilitate the identification of the first-arrival wavefront; the jump characteristics of instantaneous phase can enhance the identification of interfaces at the first-arrival time; and the high-frequency concentration characteristics of instantaneous frequency help distinguish signal differences between first-arrival waves and subsequent reflected waves, thereby providing abundant information for improving the accuracy of first-arrival picking.

Since first arrivals correspond to the high-amplitude signals first received by geophones in seismic records, amplitude serves as their primary characteristic. However, seismic data may contain noise with frequency and phase similar to those of effective signals. To suppress such noise, it is necessary to constrain the instantaneous frequency and instantaneous phase through amplitude (*i.e.*, seismic data) to obtain frequency and phase characteristic data.

2.2.2. Data normalization and cropping

Seismic data and feature data exhibit differences in dimensions and numerical magnitudes, so it is necessary to normalize all types of data to enable the network to better capture the first-arrival features. In this study, the maximum absolute value normalization method is adopted, defined as follows:

$$x = x / \max(abs(x)) \quad \text{(XIV)}$$

Before the network training, the seismic data are segmented into a size of 256 × 256, so as to speed up the training process and eliminate unnecessary information.

2.2.3. Network training

Typically, identifying the actual location of the first arrivals (i.e., the onset) poses a challenge. Therefore, in practical processing, a fixed phase (such as a wave crest or a wave trough) is selected as the first-arrival position.[5] In this study, the network treats first-arrival picking as a binary segmentation task. During sample preparation, the first peak value of the first-arrival wave is taken as the first-arrival position and labeled as 1 (the first category), while positions corresponding to all other time points are labeled as 0 (the second category). Accordingly, the network employs a binary cross-entropy loss function to quantify the discrepancy between the network output and the labels for classification purposes. The loss is defined as the sum of the losses of all pixels in the training samples within a mini-batch, expressed as follows:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N} y_i \log\left(p\left(y_i\right)\right) + \left(1-y\right)\log\left(1-p\left(y_i\right)\right) \quad \text{(XV)}$$

Where y stands for the binary label (either 0 or 1), while $p(y)$ denotes the probability that the output corresponds to label y.

The network undergoes training through the Adaptive Moment Estimation (Adam) algorithm,[50] combined with the back-propagation approach. When the validation set loss decreases and stays stable over a certain number of iteration cycles, training halts, and the weights are saved as training results. The saved optimal weights are then applied to predict the first arrivals in the test set according to a specific formula:

$$y_{test} = \Phi\left(\theta_{opt}, x_{test}\right) \quad \text{(XVI)}$$

With Φ representing the function of the network.

2.2.4. Performance evaluation

Pixel accuracy serves as a widely used indicator in semantic segmentation,[51] defined as the proportion of correctly labeled pixels relative to the total number of pixels:

$$PA = \sum_{i=0}^{k} p_{ii} / \sum_{i=0}^{k}\sum_{j=0}^{k} p_{ij} \quad \text{(XVII)}$$

Where p_{ii} represents the quantity of pixels for which class i is inferred as class i, and $\sum_{i=0}^{k}\sum_{j=0}^{k} p_{ij}$ represents all the pixel points.

After processing by the deep learning network, the probability of a point being a first arrival ranges from 0 to 1 (with a maximum of 1 and a minimum of 0). To determine the first-arrival position, thresholding is first applied to all seismic traces: if a trace contains no points with probability exceeding the threshold, the entire trace is discarded. For traces containing points with probability exceeding the threshold, the position corresponding to the original maximum probability is designated as the first-arrival position. Verified through tests on multiple datasets with distinct characteristics, the model achieves optimal overall performance when the threshold is set to 0.4.

2.2.5. Hyperparameter analysis

The first-arrival picking results of the M-Res-LSTM network are affected by hyperparameters. Therefore, during the process of training the model, we conducted experiments on some parameters, including the learning rate, the size of the kernel matrix, batch size, and the network depth. During the experiments, only the parameter being tested was changed while other parameters remained the same, and the optimal parameters were determined according to the pixel accuracy value of the validation set. The test results are shown in Figure 2.

Through the experiments, it can be seen that an overly large learning rate will make it difficult for the network to converge, and the phenomenon of back-and-forth oscillation will occur. For this data, a learning rate of 0.005 has the best effect, as shown in Figure 2A; as illustrated in Figure 2B, when the kernel matrix size is 5 × 5 of Residual Networks (ResNet), the model achieves the highest accuracy along with a fast convergence rate; Figure 2C shows that a larger batch size results in better generalization performance, though it accordingly demands more computation time and memory capacity. Using the network structures of 4×ResNet + 4×lstm and 5×ResNet + 5×lstm can both achieve relatively good accuracy, but a deeper network means that more memory will be occupied (Figure 2D). Table 2 shows the

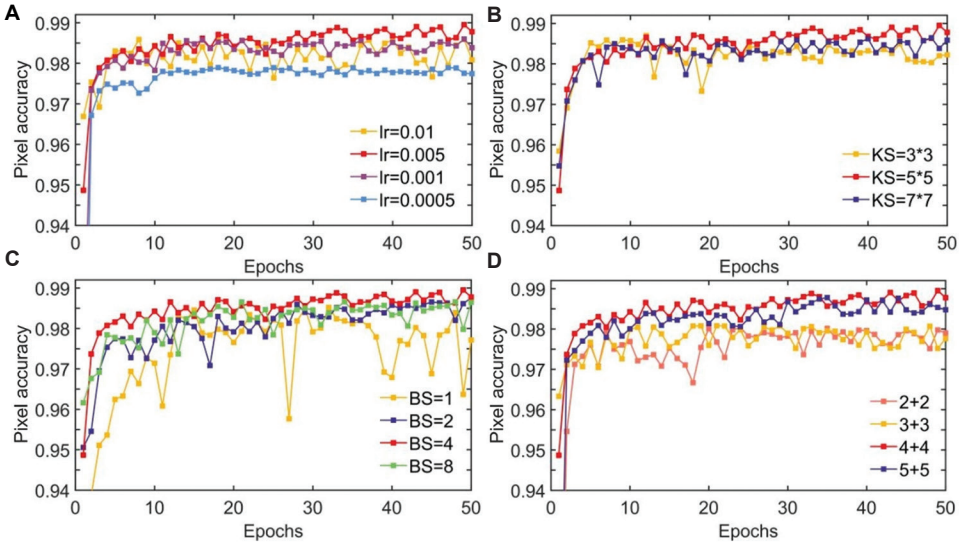


Figure 2. Pixel accuracy curves corresponding to various hyperparameters. (A) Curves of accuracy under different learning rates. (B) Accuracy curves with varying kernel matrix sizes. (C) Accuracy curves for different batch sizes. (D) Accuracy curves across different model depths.

Table 2. Experimental results of hyperparameters at the 50th epoch

Hyperparameters	Variants	PA
Learning rate	0.01	0.981
	0.005	**0.987**
	0.001	0.984
	0.0005	0.977
Batch size	1	0.976
	2	0.987
	4	**0.987**
	8	0.986
Kernel size	(3,3)	0.982
	(5,5)	**0.987**
	(7,7)	0.985
Network depth	2+2	0.979
	3+3	0.976
	4+4	**0.987**
	5+5	0.984

Notes: Values in boldface represent the values corresponding to the optimal performance for each hyperparameter.
Abbreviation: PA: Pixel accuracy.

pixel accuracy values obtained with different parameters when 50 epochs of iteration are carried out. Considering comprehensively the changing trend of the pixel accuracy value with the epoch (Figure 2) and the final accuracy (Table 2), the network will be trained with the parameters of kernel size = 5×5, learning rate = 0.005, batch size = 4, and 4×ResNet + 4×lstm. This parameter combination is expected to obtain the optimal convergence effect.

Table 3 elaborates on the parameter configurations of each module within the M-Res-LSTM network, including the input/output dimensions, channel numbers, and kernel sizes for submodules such as ResNet, LSTM, and coordinate attention. It provides a detailed technical blueprint for the network's architecture, with a total parameter quantity of 5.73 million, thus enabling the network to efficiently tackle seismic data processing tasks.

3. Results and discussion

We utilized the M-Res-LSTM network to perform first-arrival picking on three sets of real seismic data with distinct characteristics, and compared its prediction results with those from manual picking and the traditional energy ratio method. All the training was conducted on a single NVIDIA GeForce GTX 1080 Ti GPU, using the TensorFlow framework.

3.1. Data 1

Data 1 consists of small-scale 3D seismic data acquired in a plain area using dynamite sources. For each shot, 10 receiver arrays were designed, with 60 geophones deployed in each array, and the maximum offset is 1200 m. Each trace of the acquired data contains 501 sampling points, with a sampling interval set to 4 ms, and the effective recording duration of each trace is 2 s. The work area features hilly terrain, with surface elevation varying in the range of 92–160 m. Due to the limited coverage range of a single shot, the impact of topographic relief is relatively minor, and the first arrivals of seismic waves exhibit an overall smooth characteristic, providing a favorable foundation for first-arrival picking. It should be specifically noted

that the acquired seismic data contain strong industrial electrical interference and mechanical interference, and such interference signals have exerted a significant impact on first-arrival picking for some seismic traces.

The results of the energy ratio method (blue circles), manual picking (green triangles), and M-Res-LSTM (red triangles) are displayed on a representative single shot record, as shown in Figure 3. When the seismic trace contains clear first arrivals that can be identified manually, the results of M-Res-LSTM are consistent with those of manual picking. When the first arrivals are indistinguishable even to human interpreters (often due to strong abnormal noise), the model fails to pick them. This is because training labels cannot provide corresponding first-arrival positions for such unidentifiable traces. The energy ratio method attempts to pick every seismic trace, resulting in messy outputs for traces with unrecognizable first arrivals. It is reasonable to abandon picking for traces where first arrivals are unidentifiable (even manually) than to generate incorrect picks, as erroneous first arrivals significantly impact velocity modeling, while the absence of a small number of picks has minimal effect on subsequent processing.

Figure 4 shows the projections of the first-arrival time on the seismic data, instantaneous amplitude, frequency characteristic profile, and phase characteristic profile. After zooming in on the data in the red box, it can be seen that the first arrivals are located at the position of the first continuous strong amplitude and has similar phase and frequency characteristics. The method proposed in this paper can accurately pick the first arrivals through these characteristics. Figure 5 shows the absolute error of each seismic trace relative to the result of manual picking (only comparing the picked seismic traces). It can be seen from the absolute error that the picking effect of the energy ratio method is not as good as that of M-Res-LSTM.

3.2. Data 2

Data 2 used in this study is 2D seismic data acquired in a loess tableland area with dynamite sources. Each shot has 800 receiver channels, and the maximum offset is 8000 m. Each trace contains 751 sampling points with a sampling interval of 2 ms. From the perspective of the work area's geological conditions and data characteristics, the thickness of the loess layer in the work area varies significantly, with surface elevation ranging from 1200 to 1800 m and a maximum elevation difference of 600 m in the region. The severe topographic relief exerts a significant impact on the propagation path of seismic waves—not only causing the first-arrival phase within a single shot to be significantly disturbed by terrain but also leading to a large first-arrival time difference between adjacent receiver channels, which increases the basic difficulty of first-arrival picking. More critically, affected by the strong scattering of the loess layer itself and the energy attenuation of seismic waves, the first-arrival energy of the acquired data is generally weak, with unobvious onset characteristics. This has become the

Table 3. Detailed parameter table of each module in the M-Res-LSTM network

Network module	Submodule	Input dimension	Output dimension	Channel	Kernel size
Branch 1/2/3/4	ResNet	256×256×1	256×256×2	[256,128,64,32,16,8,4,2]	5×5
	LSTM	256×512	256×256×2	-	-
Feature fusion	Coordinate attention	4×[256×256×2]	256×256×4	4	1×1
Output layer	Convolutional Layer1	256×256×4	256×256×2	2	3×3
	Convolutional Layer2	256×256×2	256×256×1	1	1×1

Note: Parameter quantity=5.73 million.
Abbreviation: LSTM: Long Short-Term Memory.

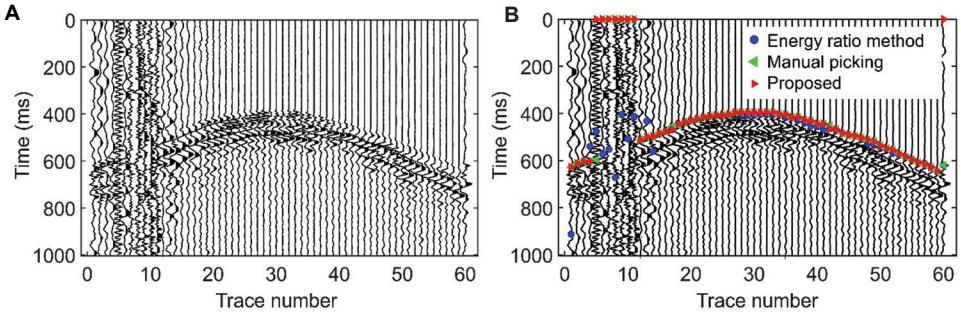


Figure 3. Data 1. (A) Original seismic record. (B) Picking results.

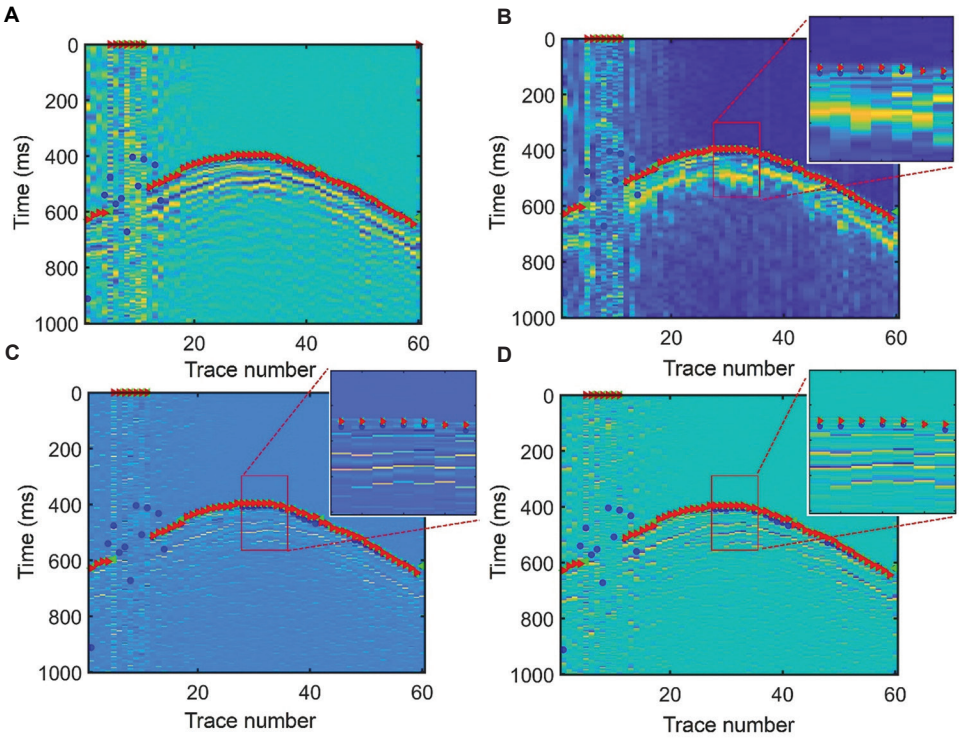


Figure 4. Projections of the seismic first arrivals on Data 1. (A) Original seismic record. (B) Instantaneous amplitude. (C) Frequency characteristic data. (D) Phase characteristic data.

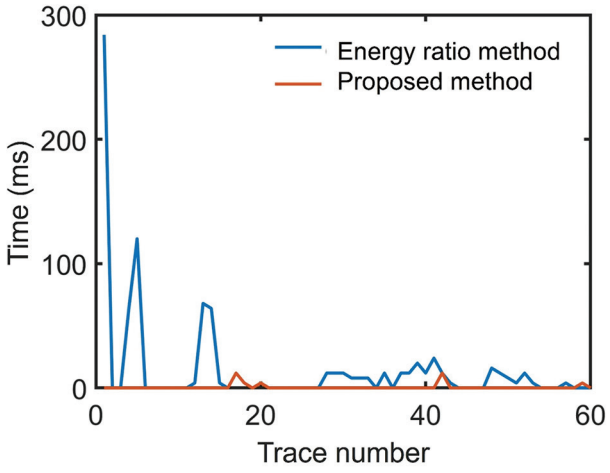


Figure 5. Absolute errors of the picking results between the energy ratio method and the method proposed in this paper (for Data 1)

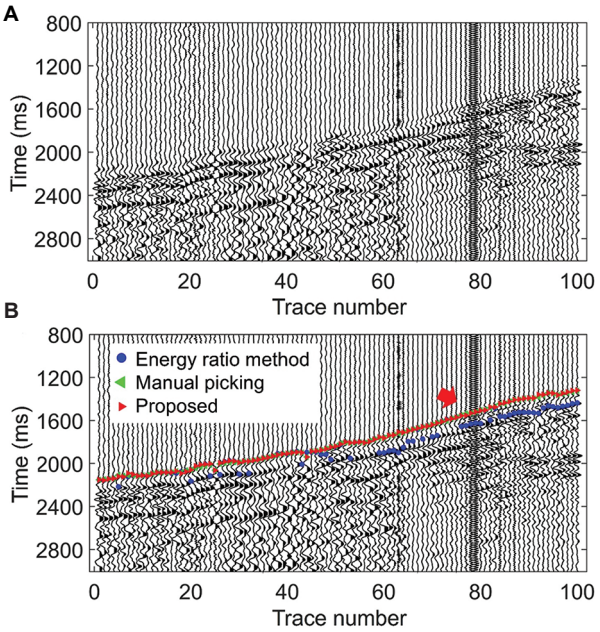


Figure 6. Data 2. (A) Original seismic record. (B) Picking results.

core difficulty in first-arrival picking for this dataset: most conventional picking methods tend to mistakenly identify subsequent phases after the first arrival as the first arrival, resulting in deviations in picking results.

Figure 6A presents a typical original profile, while Figure 6B displays the picking results obtained through manual picking (green triangles), the energy ratio method (blue circles), and the method proposed in this paper (red triangles). As observed from the picking results, the proposed method exhibits high consistency with manual

picking, whereas the energy ratio method erroneously identifies the position of the second peak (with stronger energy) as the first arrival. In addition, the proposed method successfully picks a small number of noisy traces in the seismic profile (at the position of the red arrow). This is attributed to the adoption of multi-trace input for training, enabling the network to infer first-arrival positions based on the characteristics of adjacent traces—analogous to the logic of manual picking. Figure 7 illustrates the absolute errors of the picking results of the two methods. Since the energy ratio method regards the second peak as the first arrivals, it has a relatively large absolute error. Similarly, by projecting the first-arrival time onto the seismic data, instantaneous amplitude, frequency characteristic profile, and phase characteristic profile (Figure 8), it can be seen that the first arrivals exhibit good consistency with these profiles. On magnification (within the white square), the proposed network is shown to accurately capture the amplitude, phase, and frequency characteristics of the first arrivals. Under the joint constraints of these three aspects, the accuracy of the picking is ensured.

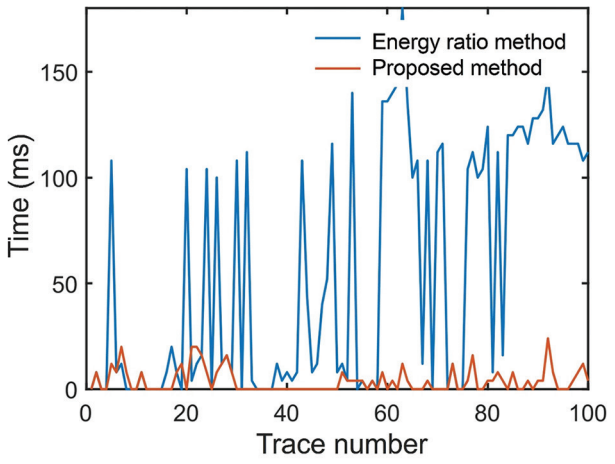


Figure 7. Absolute errors of the picking results between the energy ratio method and the method proposed in this paper (for Data 2)

3.3. Data 3

Data 3 is 3D seismic data acquired in the marginal area of a basin using a vibroseis source. For each shot, 30 receiver arrays were designed, with each array containing 170 receiver channels; the maximum offset exceeds 5000 m, enabling wide-range coverage of deep geological structures. However, due to the inherent limitation of the vibroseis source, namely, its limited excitation energy, the effective seismic wave energy received by geophones far from the source is significantly weakened, resulting in a relatively low overall SNR of the data. From the perspective of the work area's geological and topographic conditions, this region features a typical piedmont zone landscape, with extremely severe surface elevation relief: the elevation

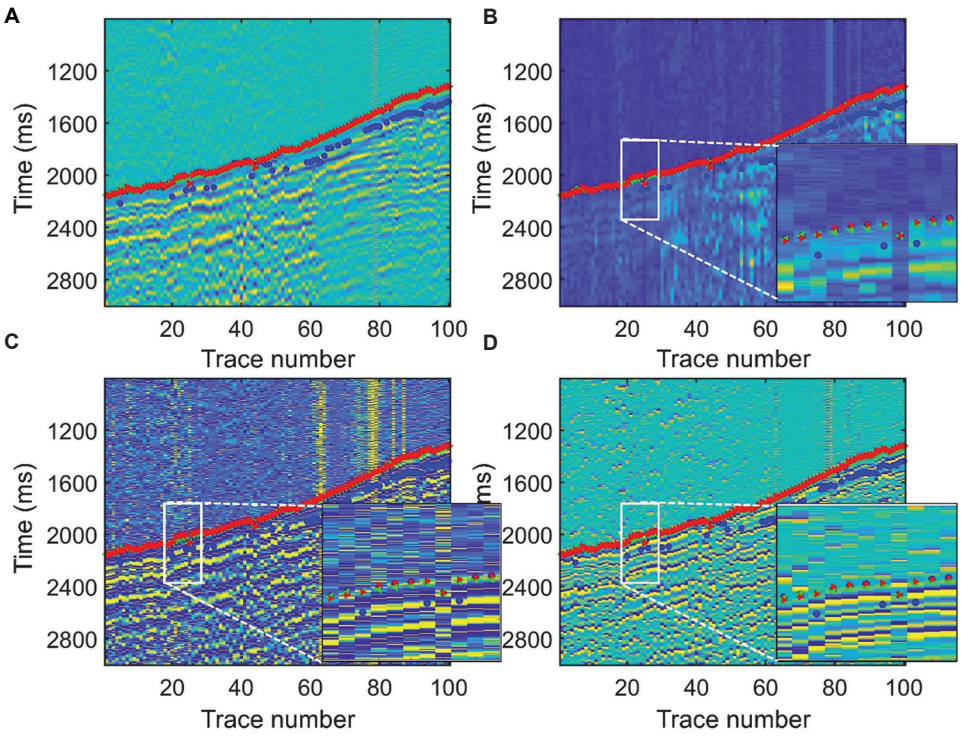


Figure 8. Projections of the seismic first arrivals on Data 2. (A) Original seismic record. (B) Instantaneous amplitude. (C) Frequency characteristic data. (D) Phase characteristic data.

ranges from 300 to 1500 m, and the maximum elevation difference in the area reaches 1200 m. Such severe topographic relief leads to a significant increase in the first-arrival time difference between adjacent receiver channels. Combined with the data's inherent issues—weak first-arrival energy and strong noise interference—this further complicates first-arrival picking.

To verify the cross-work area generalization ability of M-Res-LSTM, the model trained on Data 1 and Data 2 was directly transferred to Data 3 without any fine-tuning. Figure 9A shows a typical shot gather of Data 3, from which it can be seen that the first-arrival signals of some receiver channels are completely submerged in noise, and the energy difference between the first arrivals and background noise is small. Figure 9B compares the picking results of manual picking (yellow triangles), the energy ratio method (green triangles), and the proposed M-Res-LSTM in this study (red triangles). In Data 3, the energy ratio method is significantly affected by noise; in contrast, relying on the time-frequency dual-domain multi-feature constraint and coordinate attention mechanism, M-Res-LSTM still achieves favorable picking performance.

Figure 10 presents the absolute errors of the proposed method relative to manual picking (only valid picked channels are counted). The average absolute error of the energy ratio method reaches 5.9 ms, with the maximum error exceeding 150 ms, which is far beyond the acceptable range for seismic processing. In contrast, the average absolute error of M-Res-LSTM is only 1.34 ms, and the error of more than 87% of the gathers is controlled within 5 ms, which meets the accuracy requirements for near-surface inversion. These results indicate that by virtue of multi-domain feature learning and the attention mechanism, M-Res-LSTM effectively avoids overfitting to the features of the training work areas and can adapt to new work areas with significantly different geological conditions and noise levels.

3.4. Comparison against deep learning-driven first-arrival picking methods

To further verify the effectiveness of the proposed method, this study conducted deep learning-based tests on 1,000,000 seismic traces collected from multiple distinct work areas, in addition to testing traditional methods. The evaluation metrics selected include pixel accuracy, F1-score, first-arrival time deviation (characterized by mean absolute error, MAE), picking rate, and single-trace picking time (unit: milliseconds per trace),[40,45] which are used to comprehensively assess the performance of different networks in the seismic first-arrival picking task. Table 4 presents the performance differences between

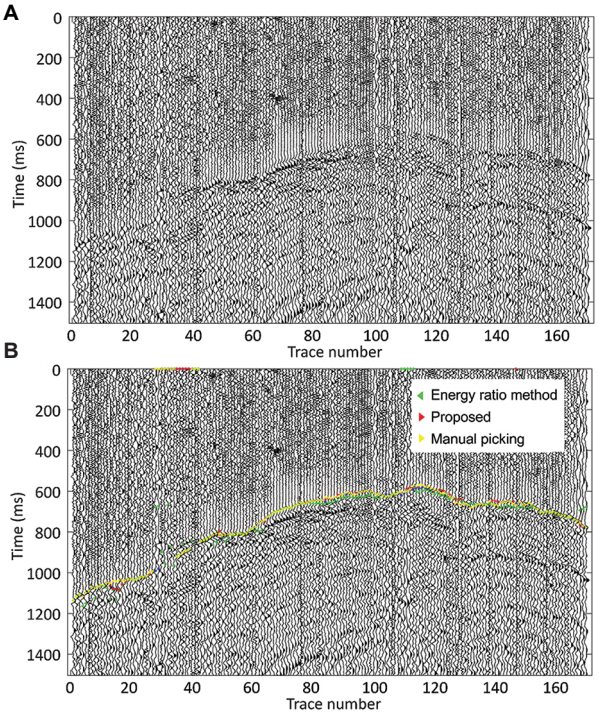


Figure 9. Data 3. (A) Original seismic record. (B) Picking results.

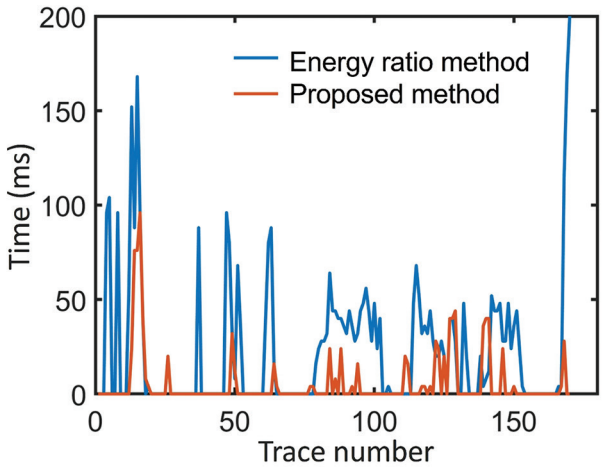


Figure 10. Absolute errors of the picking results between the energy ratio method and the method proposed in this paper (for Data 3)

Table 4. Comparison of picking results of different networks

Network	PA	F1-score	MAE (ms)	Picking rate (%)	Time (ms/trace)
Res-LSTM	0.975	0.942	7.8	92.1	0.65
Swin-Transformer	0.983	0.965	6.5	92.7	0.77
M-Res-LSTM	0.985	0.964	5.9	93.4	0.73

Abbreviation: M-Res-LSTM: Multi-perspective Residual Long Short-Term Memory; MAE: Mean absolute error; PA: Pixel accuracy; Res-LSTM: Residual Long-Short Term Memory.

the proposed M-Res-LSTM network, Res-LSTM-based networks, and Swin-Transformer-based networks[40] in seismic first-arrival picking.

Although Res-LSTM integrates the advantages of ResNet and LSTM networks, its simple cascaded structure prevents it from fully exploiting multi-dimensional information, resulting in limited overall performance and the lowest values across all metrics. Swin-Transformer achieves the highest F1-score by virtue of its self-attention mechanism; however, its single-input design restricts multi-domain feature fusion capability, and the high computational complexity of the self-attention mechanism causes it to underperform M-Res-LSTM in both first-arrival time accuracy and picking comprehensiveness.

The proposed M-Res-LSTM network in this study processes spatiotemporal signals, amplitude, frequency, and phase features in parallel through its multi-branch structure and coordinate attention mechanism, enabling comprehensive capture of spatiotemporal features. As shown in the table data, except for a slightly lower F1-score than Swin-Transformer, M-Res-LSTM outperforms Swin-Transformer in PA, picking rate, and single-trace picking time, with the first-arrival time deviation as low as 5.9 ms. This indicates that while ensuring classification accuracy comparable to Swin-Transformer, M-Res-LSTM achieves better performance in picking efficiency, picking comprehensiveness, and time accuracy through multi-domain feature parallel processing and efficient structural design. It fully verifies the effectiveness of the multi-branch structure and coordinates attention mechanism in the seismic first-arrival picking task, and can better balance accuracy, efficiency, and robustness.

4. Conclusion

The M-Res-LSTM network enables high-precision automatic picking of seismic first arrivals using time-frequency dual-domain features and an attention mechanism. Its multi-branch architecture supports parallel processing of amplitude, frequency, and phase features, thereby fully exploiting the multi-dimensional discriminative information inherent in seismic signals. The combination of residual modules and LSTM not only solves the degradation problem of deep networks but also strengthens the joint capture of spatiotemporal features. The coordinate attention mechanism effectively suppresses noise interference by dynamically adjusting feature weights, reducing the impact of incorrect first arrivals on subsequent velocity modeling.

Compared with traditional methods, manual picking achieves relatively high accuracy but suffers from the drawbacks of being time-consuming and labor-intensive. Furthermore, its results are significantly influenced by the experience of interpreters, making it difficult to meet the requirements of large-scale data processing. In contrast, the traditional energy ratio method enables automated processing yet is highly sensitive to the SNR of data, and tends to produce disorganized picking results or misjudgments in low-SNR scenarios. When compared with existing deep learning methods, the M-Res-LSTM network, leveraging a multi-feature parallel processing mechanism, exhibits superior robustness in practical applications compared to the single-branch Res-LSTM network. Meanwhile, in comparison with the Swin-Transformer network, although the M-Res-LSTM is slightly inferior in terms of pixel accuracy, it demonstrates distinct advantages in the average deviation of first-arrival time, picking rate, and single-trace computation time, thereby effectively balancing the accuracy and efficiency of first-arrival picking. Verified through data processing across different work areas, the proposed method can still obtain relatively ideal picking results even in complex scenarios with low SNR and weak first-arrival energy.

M-Res-LSTM adopts an end-to-end training mode, requiring no manual intervention. Moreover, as the amount of training data increases, its transfer ability to data from new work areas is expected to further improve. However, the complexity of the network structure makes its computation time slightly longer than that of simple models. In the future, efficiency can be optimized through model lightweighting. In addition, this paper verifies the effectiveness of time-frequency dual-domain features. Future research can explore fusion methods of more features or combine transfer learning to solve the training problems in small-sample work areas, promoting the large-scale application of this method under complex surface conditions.

Acknowledgments

None.

Funding

This work was supported in part by Sichuan Science and Technology Program (2024NSFSC0808) and in part by the Talent Project of Chengdu Technological University (2024RC028).

Conflict of interest

The authors declare that there are no (potential) conflicts of interest with any institutes, organizations, or agencies that might influence the integrity of the results or the objective interpretation of this submitted work. All research activities and manuscript preparation have been

conducted in an objective and impartial manner, free from any factors that could compromise the validity and fairness of the study.

Author contributions

Conceptualization: Ziyu Qin, Xianju Zheng
Formal analysis: Ziyu Qin
Investigation: Wenhua Wang
Methodology: Ziyu Qin
Supervision: Xianju Zheng
Visualization: Ziyu Qin, Wenhua Wang
Writing–original draft: Ziyu Qin
Writing–review & editing: Xianju Zheng, Wenhua Wang

Availability of data

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Yuan S, Liu J, Wang S, Wang T, Shi P. Seismic waveform classification and first-break picking using convolution neural networks. *IEEE Geosci Remote Sens Lett*. 2018;15(2):272-276.

   doi: 10.1109/LGRS.2017.2785834

2. Peraldi R, Clement A. Digital processing of refraction data study of first arrivals. *Geophys Prospect*. 1972;20(3):529-548.

   doi: 10.1111/j.1365-2478.1972.tb00653.x

3. Molyneux JB, Schmitt DR. First-break timing: Arrival onset times by direct correlation. *Geophysics*. 1999;64(5): 1492-1501.

   doi: 10.1190/1.1444653

4. Gelchinsky B, Shtivelman V. Automatic picking of first arrivals and parameterization of travel time curves. *Geophys Prospect*. 1983;31:915-928.

   doi: 10.1111/j.1365-2478.1983.tb01097.x

5. Coppens F. First arrival picking on common-offset trace collections for automatic estimation of static corrections. *Geophys Prospect*. 1985;33(8):1212-1231.

   doi: 10.1111/j.1365-2478.1985.tb01360.x

6. Qin Z, Pan S, Hu L, Cui Q, Gou Q. First-arrival automatic picking based on improved energy ratio method and outlier detection theory. *Acta Geophysica*. 2021;69(5):1667-1677.

   doi: 10.1007/s11600-021-00639-w

7. Allen RV. Automatic earthquake recognition and timing from single traces. *Bull Seismol Soc Am*. 1978;68(5): 1521-1532.

   doi: 10.1785/bssa0680051521

8. Wong J, Han L, Bancroft J, Stewart RR. *Automatic Time-Picking of First-Arrivals on Noisy Microseismic Data*. Alberta: CSEG Microseismic Workshop; 2009. p. 1-6.

9. Akram J, Eaton DW. A review and appraisal of arrival-time picking methods for downhole microseismic data. *Geophysics*. 2016;81(2):KS71-KS91.

   doi: 10.1190/geo2014-0500.1

10. Pan S, Qin Z, Lan H, Badal J. Automatic first-arrival picking method based on an image connectivity algorithm and multiple time windows. *Comput Geosci*. 2019;123:95-102.

    doi: 10.1016/j.cageo.2018.12.001

11. Leonard M. Comparison of manual and automatic onset time picking. *Bull Seismol Soc Am*. 2000;90(6):1384-1390.

    doi: 10.1785/0120000026

12. Zhang H, Thurber C, Rowe C. Automatic p-wave arrival detection and picking with multiscale wavelet analysis for single component recordings. *Bull Seismol Soc Am*. 2003;93(5):1904-1912.

    doi: 10.1785/0120020241

13. Jin H, Li L, Cheng S, Li X. Time-arrival pickup method of tunnel water inrush microseismic signals based on kurtosis value and AIC method. *Tunnell Undergr Space Technol*. 2024;154:106135.

    doi: 10.1016/j.tust.2024.106135

14. Boschetti F, Dentith MD, List RD. A fractal-based algorithm for detecting first arrivals on seismic traces. *Geophysics*. 1996;61(4):1095-1102.

    doi: 10.1190/1.1444030

15. Chi-Duráur-Dur/1.14440Dur- M, Silva J. Automatic detection of p- and s-wave arrival times: New strategies based on the modified fractal method and basic matching pursuit. *J Seismol*. 2017;21(5):1171-1184.

    doi: 10.1007/s10950-017-9658-0

16. Jiao LX, Moon WM. Detection of seismic refraction signals using a variance fractal dimension technique. *Geophysics*. 2000;65(1):286-292.

    doi: 10.1190/1.1444719

17. Esmaeili S. A new automatic first break picking method based on the STA/LTA fractal dimension algorithm. *J Seismic Explor*. 2022;31(3):253-265.

18. Lan Z, Gao P, Wang P, Wang Y, Liang J, Hu G. Automatic first arrival time identification using fuzzy C-Means and AIC. *IEEE Trans Geosci Remote Sens*. 2022;60:5907613.

    doi: 10.1109/TGRS.2021.3121032

19. Zhang Y, Lan Z, Xue Y, Wang J, Zhu K, He J. Automatic first arrival picking based on self-similarity and multicenter fuzzy clustering. *IEEE Trans Geosci Remote Sens*. 2025;63:5906013.

    doi: 10.1109/TGRS.2025.3538757

20. Mousa WA, Al-Shuhail AA. Enhancement of first arrivals using the ivaltransform on energy-ratio seismic shot records. *Geophysics*. 2012;77(3):V101-V111.

    doi: 10.1190/geo2010-0331.1

21. Gaci S. The use of wavelet-based denoising techniques to enhance the first-arrival picking on seismic traces. *IEEE Trans Geosci Remote Sens*. 2014;52(8):4558-4563.

    doi: 10.1109/tgrs.2013.2282422

22. Cheng Y, Li Y, Zhang C. First arrival time picking for microseismic data based on shearlet transform. *J Geophys Eng*. 2017;14(2):262-271.

    doi: 10.1088/1742-2140/aa5777

23. Sheng GQ, Tang XG, Xie K, Xiong J. Hydraulic fracturing microseismic first arrival picking method based on non-subsampled shearlet transform and higher-order-statistics. *J Seismic Explor*. 2019;28(6):593-618.

24. Dong XT, Jiang H, Li Y, Yang B. Arrival time picking of micro-seismic data by using SPE algorithm. *J Seismic Explor*. 2019;28(5):475-494.

25. Qin Z, Pan S, Chen J, Cui Q, He J. Method of automatically detecting the abnormal first arrivals using delay time. *IEEE Trans Geosci Remote Sens*. 2022;60:4504908.

    doi: 10.1109/tgrs.2021.3118921

26. Murat ME, Rudman AJ. Automated first arrival picking: A neural network approach. *Geophys Prospect*. 1992;40: 587-604.

    doi: 10.1111/j.1365-2478.1992.tb00543.x

27. McCormack MD, Zaucha DE, Dushek DW. First-break refraction event picking and seismic data trace editing using neural networks. *Geophysics*. 1993;58(1):67-78.

    doi: 10.1190/1.1443352

28. Maity D, Aminzadeh F, Karrenbach M. Novel hybrid artificial neural network based autopicking workflow for passive seismic data. *Geophys Prospect*. 2014;62:834-847.

    doi: 10.1111/1365-2478.12125.

29. Maity D, Salehi I. Neuro-evolutionary event detection technique for downhole microseismic surveys. *Comput Geosci*. 2016;86:23-33.

    doi: 10.1016/j.cageo.2015.09.024

30. Mousavi SM, Horton SP, Langston CA, Samei B. Seismic features and automatic discrimination of deep and shallow induced-microearthquakes using neural network and logistic regression. *Geophys J Int*. 2016;207:29-46.

    doi: 10.1093/gji/ggw258

31. Wu Y, Pan S, Lan H, Badal J, Wei Z, Chen Y. Automatic seismic first-break picking based on multi-view feature fusion network. *Geophys Prospect*. 2024;72(9):3547-3559.

    doi: 10.1111/1365-2478.13592

32. Tsai KC, Hu W, Wu X, Chen J, Han Z. Automatic first arrival picking via deep learning with human interactive learning. *IEEE Trans Geosci Remote Sens*. 2020;58(2):1380-1391.

    doi: 10.1109/tgrs.2019.2946118

33. Wang J, Xiao Z, Liu C, Zhao D, Yao Z. Deep learning for picking seismic arrival times. J *Geophys Res Solid Earth*. 2019;124(7):6612-6624.

    doi: 10.1029/2019jb017536

34. Zhu W, Beroza GC. Phasenet: A deep-neural-network-based seismic arrival time picking method. *Geophys J Int*. 2019;216(3):1831-1841.

    doi: 10.1093/gji/ggy423

35. Wu H, Zhang B, Li F, Liu N. Semiautomatic first-arrival picking of microseismic events by using the pixel-wise convolutional image segmentation method. *Geophysics*. 2019;84(3):V143-V155.

    doi: 10.1190/geo2018-0389.1

36. Mardan A, Blouin M, Fabien-Ouellet G, Giroux B, Vergniault C, Gendreau J. A fine-tuning workflow for automatic first-break picking with deep learning. *Near Surface Geophys*. 2024;22:539-552.

    doi: 10.1002/nsg.12316

37. Zhang Z, Yang J. Seismic first break picking based on multi-task learning. *Front Earth Sci*. 2025;13:1601134.

    doi: 10.3389/feart.2025.160113433

38. Li H, Sun Y, Li J, Li H, Dong H. A meta-learning based approach for automatic first-arrival picking. *IEEE Trans Geosci Remote Sens*. 2024;62:5922615.

    doi: 10.1109/tgrs.2024.3436817

39. Yuan P, Hu W, Wu X, Nguyen H. Adaptive first arrival picking model with meta-learning. In: *SEG International Exposition and 90th Annual Meeting*. United States: IEEE; 2020. p. 1486-1490.

    doi: 10.1190/segam2020-3420195.1

40. Jiang P, Deng F, Wang X, Shuai P, Luo W, Tang Y. Seismic first break picking through swin transformer feature extraction. *IEEE Geosci Remote Sens Lett*. 2023;20:7501505.

    doi: 10.1109/LGRS.2023.3248233

41. Chen K, Li M, Li X, *et al*. Enhancing microseismic event detection with TransUNet: A deep learning approach for simultaneous pickings of P-wave and S-wave first arrivals. *Artif Intell Geosci*. 2025;6:100129.

    doi: 10.1016/j.aiig.2025.100129

42. Jiang P, Deng F, Wang X, Luo W, Ye C. 3-D seismic first break picking based on two-channel mask strategy. *IEEE Trans Geosci Remote Sens*. 2024;62:5918115.

    doi: 10.1109/TGRS.2024.3412673

43. Liao X, Cao J, Hu J, You J, Jiang X, Liu Z. First arrival time identification using transfer learning with continuous wavelet transform feature images. *IEEE Geosci Remote Sens Lett*. 2020;17(11):2002-2006.

doi: 10.1109/lgrs.2019.2955950

44. Zhou Y, Yue H, Kong Q, Zhou S. Hybrid event detection and phase-picking algorithm using convolutional and recurrent neural networks. *Seismol Res Lett*. 2019;90(3):1079-1087.

doi: 10.1785/0220180319

45. Wang H, Zhang J, Wei X, Zhang C, Long L, Guo Z. MSSPN: Automatic first-arrival picking using a multistage segmentation picking network. *Geophysics*. 2024;89(3): U53-U70.

doi: 10.1190/geo2023-0110.1

46. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2016. p. 770-778.

doi: 10.48550/ARXIV.1512.03385

47. Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. United States: IEEE; 2021.

doi: 10.48550/arXiv.2103.02907

48. Taner MT, Koehler F, Sheriff RE. Complex seismic trace analysis. *Geophysics*. 1979;44(6):1041-1063.

doi: 10.1190/1.1440994

49. Robertson JD, Nogami HH. Complex seismic trace analysis of thin beds. *Geophysics*. 1984;49(4):344-352.

doi: 10.1190/1.1441670

50. Kingma D, Ba J. Adam: A Method for Stochastic Optimization. *In: Conference Paper at the 3rd International Conference for Learning Representations*, San Diego, CA, USA; 2015.

doi: 10.48550/ARXIV.1412.6980

51. Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Trans Pattern Anal Mach Intell*. 2017;39(4):640-651.

doi: 10.1109/tpami.2016.2572683

ACCSCIENCE PUBLISHING

Journal of Seismic Exploration

ARTICLE

Fully automatic differentiation with coupling deep neural networks for full-waveform inversion

Pengyuan Sun[1,2], Jun Zheng[3], Jingyi Zhao[3], Ying Yang[3], and Yufeng Wang[3]*

[1]BGPInc., China National Petroleum Corporation, Zhuozhou, Hebei, China

[2]National Engineering Research Center of Oil and Gas Exploration Computer Software, Zhuozhou, Hebei, China

[3]Hubei Subsurface Multi-Scale Imaging Key Laboratory, China University of Geosciences, Wuhan, Hubei, China

(This article belongs to the *Special Issue: Advanced Artificial Intelligence Theories and Methods for Seismic Exploration*)

Abstract

Seismic full-waveform inversion (FWI) is a powerful technique used in geophysical exploration to infer subsurface properties. However, FWI often suffers from challenges such as cycle skipping and sensitivity to uncertainties in seismic observations. This study aims to tackle these challenges by developing a novel fully automatic differentiation (AD) strategy for seismic FWI, coupling U-Net-based reparameterization inspired by the deep image prior concept into a reformulated wave equation simulation framework utilizing recurrent neural networks (RNNs). We demonstrate that the U-Net reparameterization serves as a form of implicit regularization for FWI, mitigating the ill-posed nature of the inversion problem and enhancing the stability of the optimization process. In addition, the RNN reformulation offers a flexible approach for backpropagating the FWI misfit, allowing the gradient with respect to the velocity parameters to be computed using the AD capabilities inherent in deep learning frameworks. Through extensive experiments on synthetic datasets, we showcase the regularization effect of our proposed method, leading to improved inversion results in terms of accuracy and robustness. This study offers a promising avenue for enhancing the reliability and accuracy of FWI through the lens of deep learning methodologies.

Keywords: Full-waveform inversion; U-Net; Deep image prior; RNN-based FWI

***Corresponding author:**
Yufeng Wang
(wangyufeng@cug.edu.cn)

Citation: Sun P, Zheng J, Zhao J, Yang Y, Wang Y. Fully automatic differentiation with coupling deep neural networks for full-waveform inversion. *J Seismic Explor*. 2025;34(6):16-28.
doi: 10.36922/JSE025410085

Received: October 11, 2025

Revised: October 27, 2025

Accepted: October 28, 2025

Published online: November 19, 2025

1. Introduction

Seismic full-waveform inversion (FWI) stands out as a crucial method in geophysical exploration, allowing for high-resolution reconstruction of subsurface properties.[1-3] FWI iteratively refines velocity models by minimizing the difference between synthetic and observed data, employing gradient descent algorithms with gradients computed through the adjoint-state method.[4,5] Despite its significance, traditional FWI methods face challenges such as cycle-skipping and ill-posedness, making it difficult to accurately represent complex subsurface velocity models when initial models are inaccurate and

observations are incomplete or contaminated with noise.[6-8] Moreover, computation of gradients in traditional FWI through the adjoint state method can be cumbersome, requiring formulation for each wave equation, making numerical implementation challenging and prone to errors.[9,10]

Deep learning has emerged as a promising approach in seismic FWI, offering novel solutions to longstanding challenges in conventional seismic inversion.[11,12] By leveraging annotated seismic data pairs consisting of observed seismogram and corresponding subsurface models, supervised learning-based FWI methods train neural networks to learn the complex mapping between seismic data and subsurface properties.[13,14] However, supervised learning FWI heavily relies on the availability of large volumes of labeled training data. Obtaining such datasets can be challenging and resource-intensive. Apart from directly mapping seismic data to inverted models, the integration of deep learning to aid seismic FWI has been intensively explored over the recent years. This includes employing deep learning techniques for tasks such as data augmentation, model initialization, optimization, misfit function design, and learned regularization.[15-20] Supervised learning FWI shows promise but comes with limitations, such as reliance on high-quality labeled data and potential overfitting to specific datasets. In addition, its generalization to diverse geological settings can be problematic, limiting its effectiveness in real-world applications.[21,22]

Recent research has shifted toward physics-based deep learning FWI, where the neural network architecture or loss function encodes underlying physical principles. This approach aims to enhance the interpretability and generalization capabilities of FWI models by explicitly incorporating prior knowledge of the physics governing seismic wave propagation.[22-24] Physics-based deep learning for FWI can be approached in various ways. First, the utilization of deep learning tools like automatic differentiation (AD) and optimization methods has streamlined the FWI process, making it more straightforward and robust.[9,10,23,25-27] These techniques reformulate the time-marching finite-difference discretized wave equation as a recurrent neural network (RNN), which is often referred to as RNN-based FWI. This approach allows for the automatic calculation of gradients and facilitates efficient model updates. Second, integrating the wave equation into neural networks, as demonstrated by physics-informed neural networks (PINNs), enhances the ability of neural networks to grasp the fundamental physics of wave propagation, thereby improving inversion accuracy.[22,28,29] This approach allows the model to leverage known physical principles, reducing the reliance on large datasets and improving generalization across different scenarios. As a result, PINNs-based FWI can offer more robust solutions in complex subsurface environments, addressing some of the limitations faced by traditional methods. Third, the recently developed neural operator learning methods aim to approximate implicit operators defined by partial differential equations (PDEs) between functional spaces.[30,31] These methods can serve as a rapid surrogate for the wave equation, enhancing the efficiency of seismic inversion by reducing the need for multiple wave equation simulations.[24,32] In addition, the concept of deep image prior suggests that the architecture of a neural network itself can act as a potent prior for inversion tasks.[33-35] In the realm of linear inversion, the deep neural network (DNN) parameterization method is referred to as regularization by architecture, where the spatial and temporal features of DNNs are harnessed to adjust inversion results to meet specific expectations.[34] The efficacy of regularization by architecture relies, to some extent, on the meticulous design of network architectures. In the geophysics community, the use of DNNs, particularly convolutional neural networks (CNNs), has gained traction for regularized estimation in FWI. This approach leverages the inherent structure of CNNs to capture spatial dependencies in geophysical data, improving the accuracy and robustness of subsurface model estimations.[36-41]

In this study, we propose a novel seismic FWI framework with coupling DNNs for reparameterization and reformulation, termed fully automatic differentiation-based FWI (FAD-FWI). In this approach, the subsurface velocity models are reparameterized by the weights of DNNs and then fed into an RNN-based FWI module. The seamless integration of these two neural networks enables FAD, allowing the weights of the DNNs to be updated by backpropagating the misfit between synthetic and observed seismograms. The integration of DNNs and RNNs enhances inversion outcomes by eliminating the need for manual tuning of regularization parameters and the reliance on error-prone adjoint state methods. In addition, the FAD-FWI framework offers flexibility in handling complex geological structures and can potentially outperform traditional FWI techniques by exploiting the strengths of deep learning for regularization and optimization. By harnessing the hierarchical feature extraction capabilities of the U-Net, our proposed FAD-FWI method can effectively map a Gaussian random field (GRF) input to the inverted velocity model, aligning it with seismic observations through RNN-based FWI. FAD-FWI outperforms traditional FWI with lower dependency on initial model estimations and better robustness in the face of uncertainties in seismic observations. Through extensive

experiments on synthetic seismic datasets, we demonstrate that FAD-FWI provides more accurate subsurface models and offers greater computational flexibility compared to conventional FWI methods. The main contributions of this work are twofold. We propose the FAD-FWI architecture, which overcomes a key limitation of prior deep learning-based FWI methods. By employing a linear activation in the final layer, our model directly outputs velocity values in a physically realistic range, eliminating the need for a problem-dependent scaling factor. In addition, we develop a fully AD framework that seamlessly integrates a DNN-based model parameterization with the physics of wave propagation. This unified approach automatically computes the gradients of the FWI objective function through both the U-Net and the wave equation solver, removing the dependency on manually derived and implemented adjoint-state equations.

2. Methodology

2.1. FWI with regularization

Seismic FWI seeks to estimate subsurface properties by iteratively updating the velocity model until synthetic seismic data closely match the observed seismic data. FWI minimizes an objective function that measures the discrepancy between recorded and simulated seismic data. This function typically combines a data-misfit term with regularization to ensure stability and to guide the solution. Mathematically, the objective function can be expressed as:

$$\mathcal{J}(m)=\frac{1}{2}\sum_{s,r} d_{obs}(x_r,t;x_s)-d_{cal}(x_r,t;x_s;m)^2+\lambda\mathcal{R}(m) \quad \text{(I)}$$

Where d_{obs} and d_{cal} are observed wavefield and the calculated wavefield recorded at receivers associated with sources , respectively. R denotes the regularization term on velocity model m with weighting coefficient λ. The handcrafted regularization terms, often based on expert knowledge or empirical observations, help guide the inversion process toward solutions that are physically plausible and consistent with prior expectations. However, these priors can sometimes be overly general, as the models generated using their associated probability density functions may encompass a broader range of possibilities than those specifically relevant to geophysics. As a result, there is a risk of introducing biases or inaccuracies into the inversion results, particularly when the priors do not accurately capture the true distribution of subsurface properties in the study area.[42,43] This limitation has prompted researchers to explore alternative approaches to regularization that can adaptively incorporate domain-specific knowledge and better capture the complexities of subsurface structures in geophysical inversion tasks.

2.2. FWI with U-Net reparameterization

As demonstrated in the seminal work of deep image prior, a randomly-initialized neural network can serve as effective prior in inverse problems.[33,34] In this study, we employ U-Net reparameterization for seismic FWI with the following objective function:

$$\mathcal{J}(\theta)=\frac{1}{2}\sum_{s,r} d_{obs}(x_r,t;x_s)-d_{cal}\left(x_r,t;x_s;\mathcal{N}(z;\theta)\right)^2 \quad \text{(II)}$$

Where velocity model is reparameterized by a U-Net $N(z;\theta)$ with weights θ and fixed latent variable z. In this study, we use a latent variable z generated by GRF. In contrast to traditional FWI with the velocity m updated in model space, the proposed FAD-FWI updates the U-Net weights θ iteratively to match the observed data d_{obs} using a gradient-based optimization method with the gradient computed through the chain rule as follows:

$$\frac{\partial\mathcal{J}}{\partial\theta}=\left(\frac{\partial m}{\partial\theta}\right)^T\frac{\partial\mathcal{J}}{\partial m}=\left(\frac{\partial m}{\partial\theta}\right)^T\sum_{t=1}^{N_t}\frac{\partial\mathcal{J}}{\partial u_t}\frac{\partial A(m)}{\partial m}u_{t-1} \quad \text{(III)}$$

Here, we assume the time-marching finite-difference discretization of the wave equation $u_t = A(m)\ u_{t-1} + s_{t-1}$ with $A(m)$ being the finite difference coefficient matrix. In general, regularization by U-Net architecture ensures that the inverted subsurface models maintain consistency with observed seismic data while also achieving desired properties such as spatial coherence and smoothness.[36,38,41] The input to the U-Net-reparameterized FWI consists of GRF realizations of random latent variables, with dimensions matching those of the velocity model. In this study, GRF has a covariance kernel function as follows:

$$C_\alpha(d)=\sigma^2\frac{2^{1-\alpha}}{\Gamma(\alpha)}\left(\sqrt{2\alpha}\frac{d}{\beta}\right)^\alpha K_\alpha\left(\sqrt{2\alpha}\frac{d}{\beta}\right) \quad \text{(IV)}$$

Where σ is the variance of the Gaussian process, α is known as the smoothness of GRF.[44,45] The constants used in the GRF kernel are determined by the complexity of the velocity model. Figure 1 presents the GRF latent variables with different smoothness, (a) α = 1.0, (b) α = 3.0, and (c) α = 5.0, respectively. In this study, we let α = 3.0 for all experiments. The U-Net architecture consists of an encoder-decoder structure: the encoder extracts features through a series of convolutional and downsampling layers, while the decoder upsamples the features to recover spatial resolution. Skip connections between corresponding layers in the encoder and decoder allow for detailed feature preservation.[46,47] The output of the U-Net is the predicted velocity model, which is subsequently fed into the FWI module to ensure that the synthetic data

match the observed data. In this U-Net reparameterization scheme, the U-Net serves as a regularizer, leveraging its inductive biases, such as spatial consistency and hierarchical feature extraction. These biases help preserve important structural patterns in the velocity model, promoting smoother and more geologically plausible solutions. By incorporating this learned regularization, the FWI process becomes more stable and less prone to overfitting, improving the accuracy and robustness of the inversion results.

2.3. FAD-based FWI

U-Net-reparameterized FWI leverages the inductive biases of the U-Net architecture, enhancing the accuracy and robustness of the inversion process. However, challenges arise when coupling the gradients of DNNs ∂m/∂θ with that of the PDEs ∂J/∂m, where the former is typically computed using backpropagation by AD within deep learning framework,[48] whereas the latter is commonly derived through the adjoint-state method.[4] Fortunately, the recently developed RNN-based FWI reformulates FWI into an RNN, enabling the gradient calculation of velocity parameters using AD as well.[23,25] The schematic architecture of the proposed FAD-FWI is shown in Figure 2, which seamlessly integrates two parts, with a U-Net architecture playing a pivotal role in reparameterizing the velocity model and an RNN enabling the gradients with respect to inversion parameters calculated by AD in a modern deep learning framework. This integrated approach holds promise for overcoming traditional FWI limitations and advancing the capabilities of seismic imaging in characterizing subsurface properties. In our proposed framework, two neural networks are combined, allowing the gradient of the cost function with respect to the U-Net weights to be fully computed through AD. This seamless gradient calculation eliminates the need for manual derivation of adjoint equations, enabling efficient updates to the U-Net weights during the inversion process. This is why we refer to the method as FAD-FWI, as it takes advantage of AD to optimize both the velocity model and neural network parameters simultaneously, streamlining the FWI workflow.

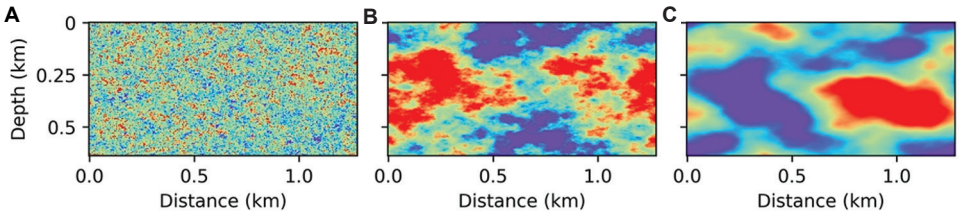


Figure 1. The Gaussian random field latent variables with different smoothness: (A) α = 1.0, (B) α = 3.0, and (C) α = 5.0, respectively

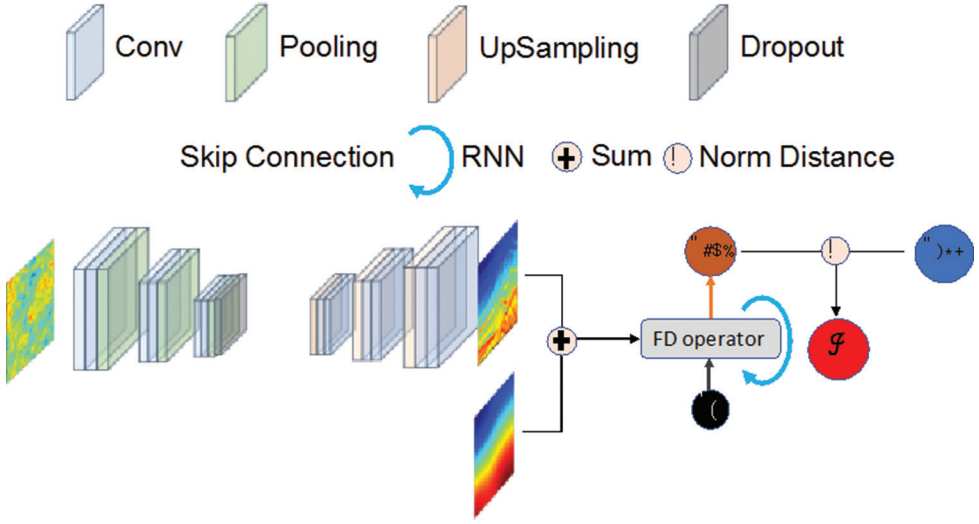


Figure 2. The schematic architecture of the proposed FAD-FWI. The GRF latent variable is fed into an encoder to generate a fused feature map, which is decoded and subsequently directed into an RNN-FWI module. The gradient of the cost function with respect to the U-Net weights is fully computed through AD. The plot of the RNN-based FWI module is adopted from Ref.[25]
Abbreviations: AD: Automatic differentiation; FAD: Fully automatic differentiation; FWI: Full-waveform inversion; GRF: Gaussian random field; RNN: Recurrent neural network.

3. Numerical examples

3.1. FAD-FWI regularized by U-Net architecture

In the first experiment, we aim to demonstrate the advantages of our proposed FAD-FWI framework regularized by the U-Net architecture compared to generative neural networks (GNNs). Specifically, we seek to verify that the inductive biases of U-Net, such as its ability to capture spatial hierarchies and maintain structural consistency, provide more robust regularization, and lead to more accurate velocity model predictions. By comparing the performance of FAD-FWI with U-Net regularization against GNN, we will highlight its effectiveness in producing geologically plausible models while improving the stability and convergence of the inversion process. The optimization configuration is consistent across all experiments. We use the Adam optimizer to minimize the objective function. The learning rate is set to 0.01 and kept constant throughout the inversion process. Each experiment is run for a fixed budget of 2000 iterations, which is empirically determined to be sufficient for the loss to converge in all tested scenarios.

We utilize an angular unconformity geological model shown in Figure 3A to demonstrate the superiority of our proposed FAD-FWI framework, regularized by the U-Net architecture (FAD-FWI-U-Net), in comparison to the version regularized by a GNN (FAD-FWI-GNN) with varying scaling factors as developed by Zhu *et al.*[38] The scaling factors are critical for mapping the bounded output of the neural network to the physically meaningful range of velocity models. As pointed by Zhu *et al.*,[38] applying scaling factors to the output of neural networks depends on the physical parameters and units. The final layer of the neural network used a Tanh activation function, which constrains its output to the range [-1, 1]. To map this bounded output to a meaningful velocity perturbation, a scaling factor is required. In contrast, our primary proposed method, FAD-FWI-U-Net, uses a linear activation (*i.e.*, no activation function) in its final layer. This is a significant advantage, as it allows the network to output velocity values in an unbounded range directly, without the need to assume or tune a predefined scaling factor. This makes FAD-FWI-U-Net more robust and easier to apply to new datasets where the appropriate velocity range may not be known *a priori*.

Figure 3A and B present the velocity model and a homogeneous initial model for FWI, respectively. The seismic acquisition configuration includes a total of 20 shots indicated by red stars and 256 receivers indicated by white dots, as shown in Figure 3B. The inverted velocity model obtained by FAD-FWI-U-Net is displayed in Figure 3C, showcasing good agreement with the true model in Figure 3A. Figure 3D-F depicts the recovered

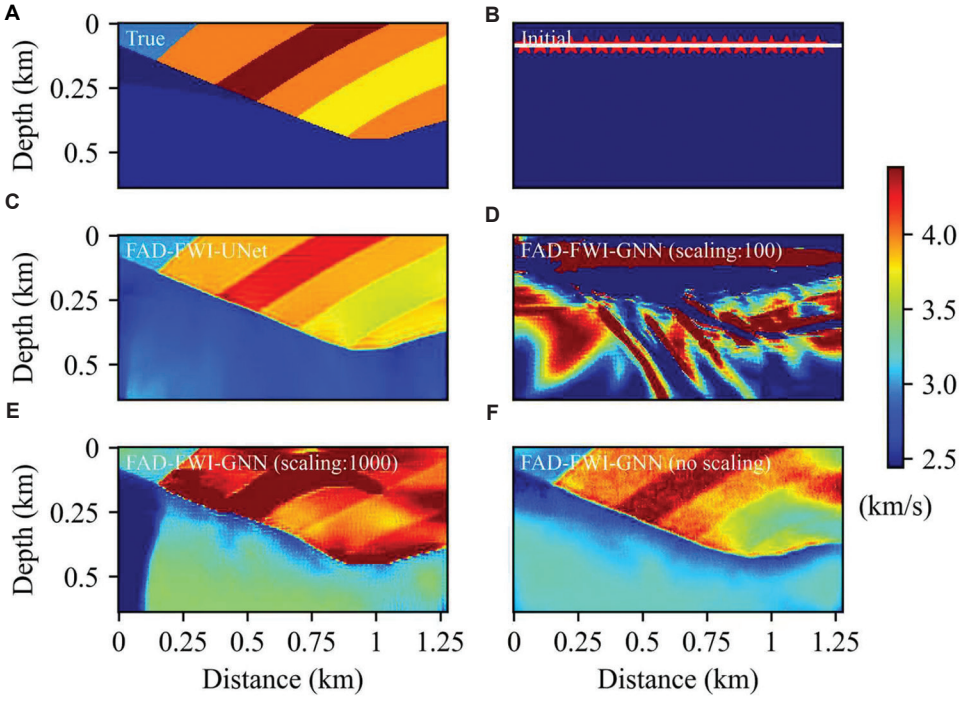


Figure 3. Comparison of inverted velocity models obtained using the proposed FAD-FWI-U-Net and FAD-FWI-GNN with different scaling factors. (A) The angular unconformity geological model; (B) homogeneous initial model; (C) inverted model by FAD-FWI-U-Net, and the inverted models by FAD-FWI-GNN with scaling factor of 100 (D), 1000 (E), and no scaling (F).
Abbreviations: FAD: Fully automatic differentiation; FWI: Full-waveform inversion; GNN: Generative neural network.

velocity models obtained by FAD-FWI-GNN with scaling factors of 100, 1000, and no scaling, respectively. Our results illustrate that FAD-FWI-U-Net can accurately reconstruct the velocity model without the need for predefined scaling, whereas FAD-FWI-GNN requires appropriate scaling for successful inversion; otherwise, it fails to accurately recover the velocity model. In addition, even after removing the Tanh activation in FAD-FWI-GNN, the inversion result shown in Figure 3F remains inferior to that of our proposed FAD-FWI-U-Net shown in Figure 3C. This difference is primarily due to the U-Net architecture, which effectively captures spatial hierarchies, preserves structural consistency, and offers more robust regularization for FWI. Figure 4 shows the loss curves for the proposed FAD-FWI-U-Net and FAD-FWI-GNN with scaling factors of 100, 1000, and no scaling. The results indicate that our proposed FAD-FWI-U-Net achieves a faster convergence rate after 500 iterations and a lower L1 error compared to FAD-FWI-GNN across different scaling factors. The impact of DNN architecture on parameterized FWI is a promising area for exploration in future work.

3.2. FAD-FWI across initial models

The second experiment focuses on evaluating the sensitivity of the proposed FAD-FWI method to the initial model estimation. Traditional FWI methods are often highly dependent on an accurate initial model; they tend to perform poorly when initialized with a less accurate model. In contrast, our FAD-FWI method, utilizing U-Net for parameterization, provides a strong regularizer that helps mitigate this dependency, enhancing inversion results even with suboptimal initial models. This robustness highlights the potential of our framework to improve inversion accuracy in challenging scenarios where initial model quality is compromised. In this experiment, we compare our proposed FAD-FWI method with traditional FWI implemented using Deepwave (https://ausargeo.com/deepwave/) and referred to as DW-FWI for simplicity.

Figure 5 presents the Marmousi velocity model, along with three different initial velocity models: smoothed model, smoothed 1D model, and homogeneous model, respectively. The acquisition configuration consists of 20 shots and 256 receivers, positioned at a depth of 85 m. Figure 6 presents the inverted velocity models obtained by DW-FWI and the proposed FAD-FWI with three different initial velocity models as shown in Figure 5. As the accuracy of the initial models decreases, the inversion results from DW-FWI deteriorate significantly. In contrast, our FAD-FWI approach achieves acceptable inversion results even when starting from a homogeneous initial model. Figure 7 depicts the comparison of the extracted

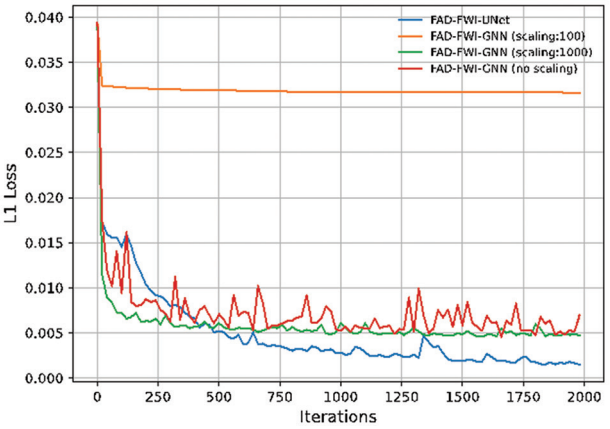


Figure 4. Comparison of the loss curves of the proposed FAD-FWI-U-Net and FAD-FWI-GNN with scaling factors of 100, 1000, and no scaling. We observe that FAD-FWI-GNN with a scaling factor of 100 fails to converge to lower L_1 errors due to improper scaling.
Abbreviations: FAD: Fully automatic differentiation; FWI: Full-waveform inversion; GNN: Generative neural network.

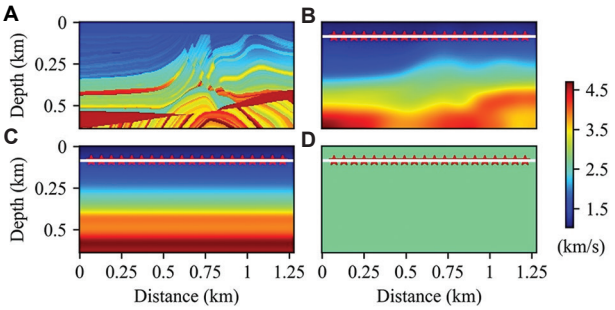


Figure 5. The Marmousi velocity model (A), along with three different initial velocity models: (B) smoothed model, (C) smoothed 1D model, and (D) homogeneous model. The acquisition configuration consists of 20 shots marked by red stars and 256 receivers by white dots, positioned at a depth of 85 m.

traces from Figure 6 at 0.25 km, 0.5 km, 0.75 km, and 1.0 km. It indicates that DW-FWI struggles to recover the velocity model when provided with an inaccurate homogeneous initial model, while the proposed FAD-FWI method still achieves satisfactory results, although with a slightly degraded quality. Figure 8 depicts the loss curves of DW-FWI and the proposed FAD-FWI with different initial velocity models. It is evident that conventional FWI without regularization exhibits a faster convergent rate at the early stages. However, the proposed FAD-FWI, which incorporates regularization by U-Net architecture, demonstrates lower L_1 errors overall. We further conduct sensitivity analysis of DW-FWI and the proposed FAD-FWI given a smoothed 1D initial model with different velocity perturbations from −30% to +30%. The inverted velocity models are displayed in Figure 9. The implication is

that DW-FWI is more susceptible to velocity perturbations compared to FAD-FWI. This experiment demonstrates the robustness of the proposed FAD-FWI to variations in the initial model estimation.

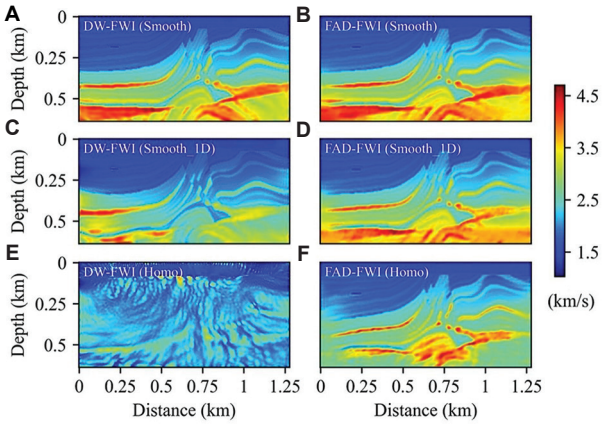


Figure 6. Comparison of inverted velocity models obtained using DW-FWI (A, C, E) and the proposed FAD-FWI (B, D, F) with different initial velocity models shown in Figure 5. From top to bottom, the rows correspond to inverted models with the smoothed model, smoothed 1D model, and the homogeneous model.
Abbreviations: DW: Deepwave; FAD: Fully automatic differentiation; FWI: Full-waveform inversion.

3.3. FAD-FWI across uncertainties in seismic observations

In our third experiment, we apply FAD-FWI to the 2D Overthrust velocity model, as depicted in Figure 10, in the presence of uncertainties in seismic observations such as noise and incomplete frequency components. We use a smoothed 1D initial model shown in Figure 10B for all tests. The acquisition configuration for this experiment is the same as in the previous experiments, with 20 shots and 256 receivers positioned at a depth of 85 m. Figure 11 presents shot gathers under various uncertainties in seismic observations, including clean and noisy data, as well as complete and incomplete data with missing frequencies below 2.5 Hz. The extracted traces at the left-most position, along with their spectra, are displayed alongside the shot gathers.

In this experiment, we perform FWI under conditions of uncertainty in seismic observations, including noise and incomplete frequency components. Figure 12 compares the inverted velocity models obtained by DW-FWI using different seismic observations with clean and noisy data, as well as complete and incomplete data with missing frequencies below 2.5 Hz. In this scenario, DW-FWI demonstrates acceptable performance with clean

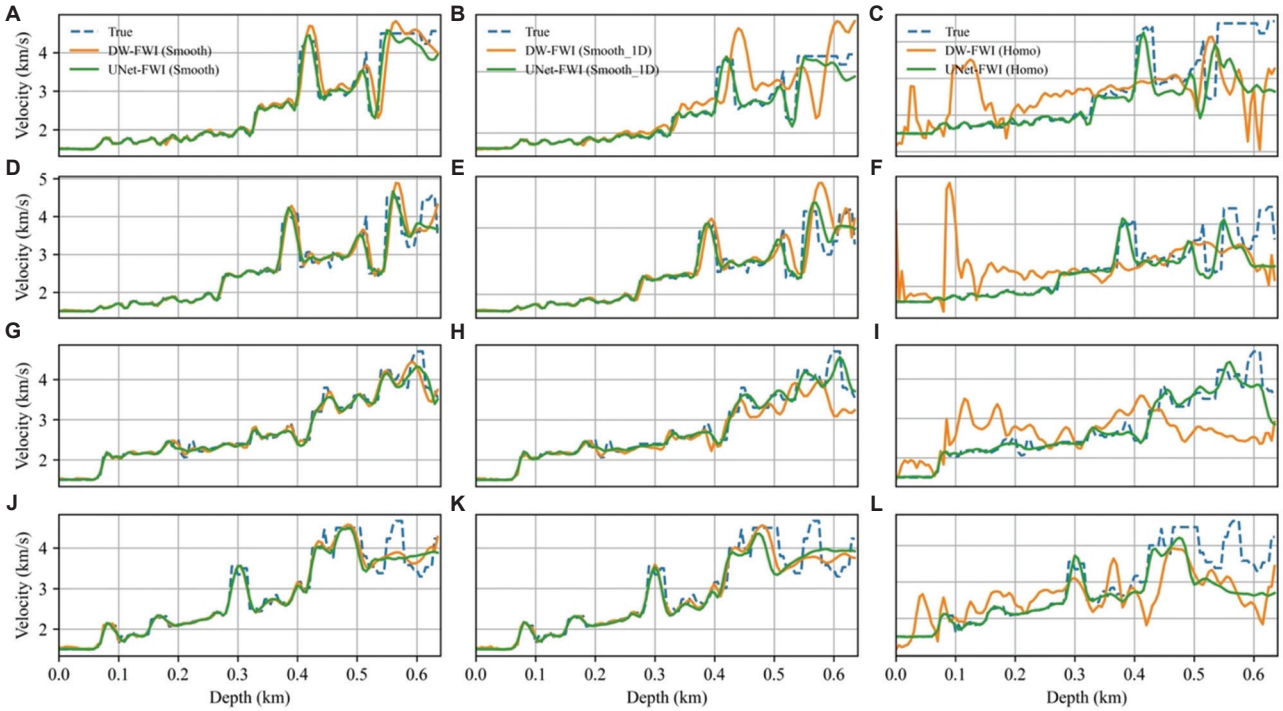


Figure 7. Comparison of the traces extracted from the inverted velocity models shown in Figure 6 at four trace locations. From top to bottom, the rows correspond to trace positions at 0.25 km (A, B, C), 0.5 km (D, E, F), 0.75 km (G, H, I), and 1.0 km (J, K, L). From left to right, the columns correspond to smoothed model (A, D, G, J), smoothed 1D model (B, E, H, K), and homogeneous model (C, F, I, L).
Abbreviations: DW: Deepwave; FAD: Fully automatic differentiation; FWI: Full-waveform inversion.

observations; however, its accuracy declines significantly when the seismic data are contaminated with random noise. It fails to recover the velocity model effectively when faced with both noisy data and missing low-frequency components. Figure 13 compares the inverted velocity models obtained using the proposed FAD-FWI with different seismic observations. The FAD-FWI method demonstrates robust performance, maintaining high-quality inversion results even as data quality degrades due to noise or incomplete frequency information. This resilience highlights the effectiveness of FAD-FWI in handling challenging data conditions. We provide quantitative evaluation metrics in Table 1 to assess the performance of the proposed FAD-FWI. The quantitative metrics mean squared error (MSE), structural similarity index measure (SSIM), and peak signal-to-noise ratio (PSNR) are defined by:

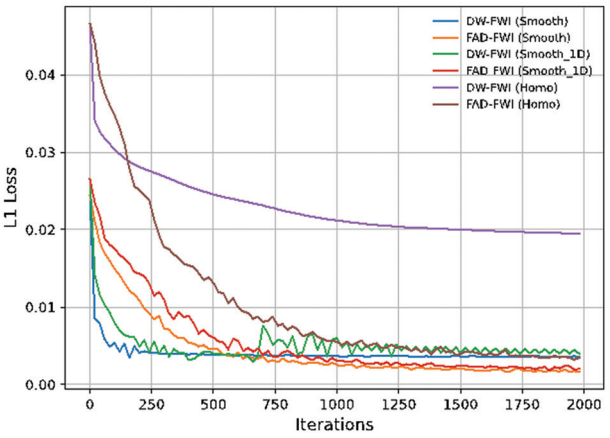


Figure 8. Comparison of the loss curves of DW-FWI and the proposed FAD-FWI with different initial velocity models.
Abbreviations: DW: Deepwave; FAD: Fully automatic differentiation; FWI: Full-waveform inversion.

$$MSE\left(m_{true},m_{inv}\right)=\frac{1}{N}\sum_{i=1}^{N}\left[m_{true}\left(i\right)-m_{inv}\left(i\right)\right]^{2} \quad \text{(V)}$$

$$PSNR\left(m_{true},m_{inv}\right)=10log_{10}\left(\frac{\left[MAX\left(m_{true}\right)-MIN(m_{true})\right]^{2}}{MSE\left(m_{true},m_{inv}\right)}\right) \quad \text{(VI)}$$

$$SSIM\left(x,y\right)=\frac{\left(2\mu_{x}\mu_{y}+C_{1}\right)\left(2\sigma_{xy}+C_{2}\right)}{\left(\mu_{x}^{2}+\mu_{y}^{2}+C_{1}\right)\left(\sigma_{x}^{2}+\sigma_{y}^{2}+C_{2}\right)} \quad \text{(VII)}$$

Where μ_x and μ_y are the mean intensities, and σ_x^2 and σ_y^2 are the variances of true model and the inverted model,

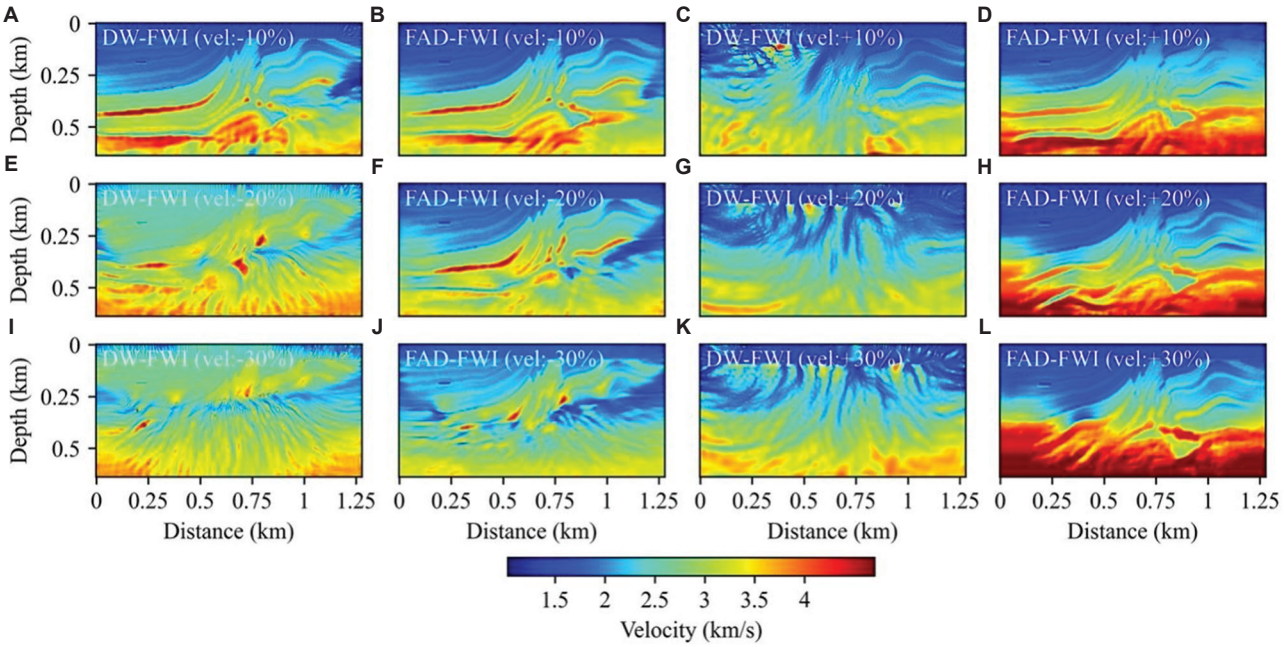


Figure 9. Sensitivity analysis of DW-FWI and the proposed FAD-FWI given smoothed 1D initial model with different velocity perturbations. The first column denotes the inverted models from DW-FWI with −10% (A), −20% (E), and −30% (I) deviated from the smoothed 1D initial model shown in Figure 5C. The second column denotes the inverted models from FAD-FWI with velocity perturbations of −10% (B), −20% (F), and −30%(J). The third and fourth columns correspond to the inverted models from DW-FWI and FAD-FWI with velocity perturbations of +10% (C and D), +20% (G and H), and +30% (K and L), respectively.
Abbreviations: DW: Deepwave; FAD: Fully automatic differentiation; FWI: Full-waveform inversion.

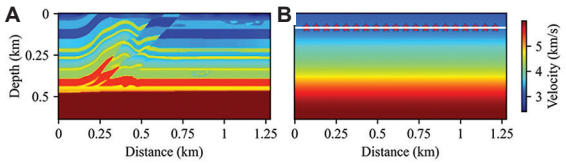


Figure 10. The Overthrust velocity model (A) and the smoothed 1D initial model (B). The acquisition configuration consists of 20 shots marked by red stars and 256 receivers by white dots, positioned at a depth of 85 m.

respectively. C_1 and C_2 are small constants stabilizing the division. The overall SSIM index is the mean of the SSIM values across all windows. The SSIM value ranges from −1 to 1, and a value of 1 indicates perfect structural similarity. This quantitative comparison suggests that the proposed FAD-FWI is more robust than DW-FWI in scenarios where observations lack low-frequency components and are contaminated by noise. In addition, we compare the

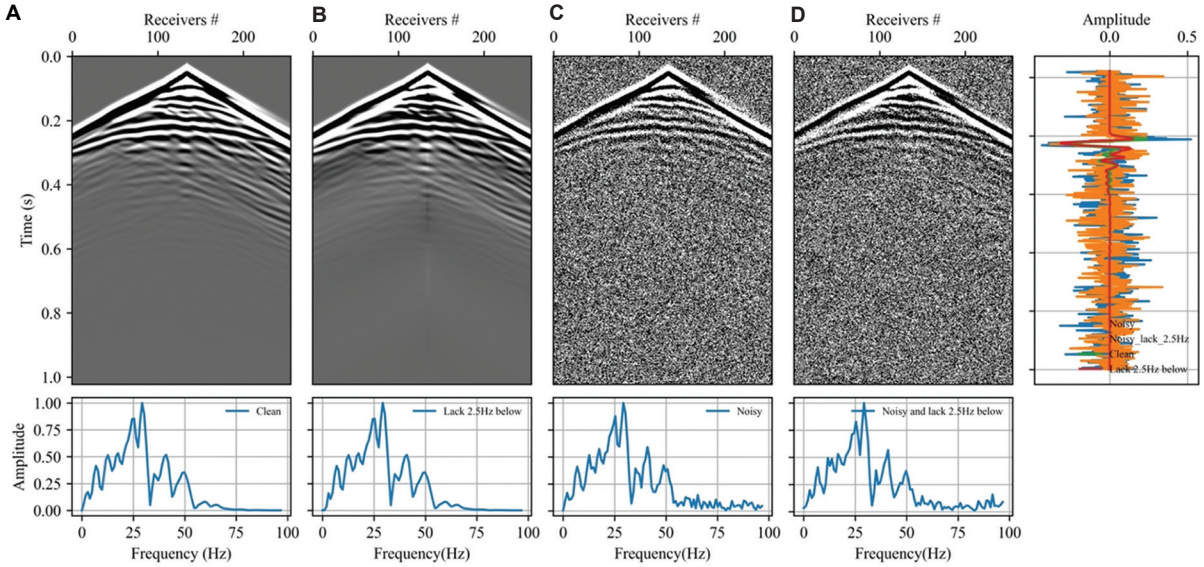


Figure 11. Shot gathers under various uncertainties in seismic observations, including clean data (A), clean data with missing frequencies below 2.5 Hz (B), noisy data with random noises $\sigma = 0.5\sigma_0$ (C), and noisy data with missing frequencies below 2.5 Hz (D). The extracted traces at the left-most position, along with their spectra, are displayed alongside the shot gathers.

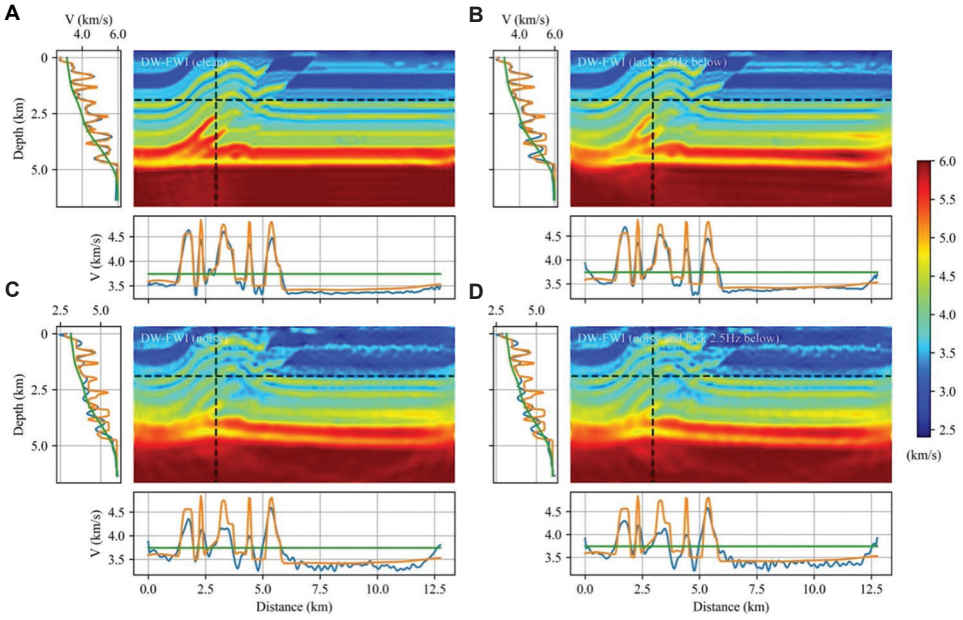


Figure 12. Comparison of inverted velocity models obtained using DW-FWI given different observations with clean data (A), clean data with missing frequencies below 2.5 Hz (B), noisy data with random noises $\sigma = 0.5\sigma 0$ (C), and noisy data with missing frequencies below 2.5 Hz (D).
Abbreviations: DW: Deepwave; FWI: Full-waveform inversion.

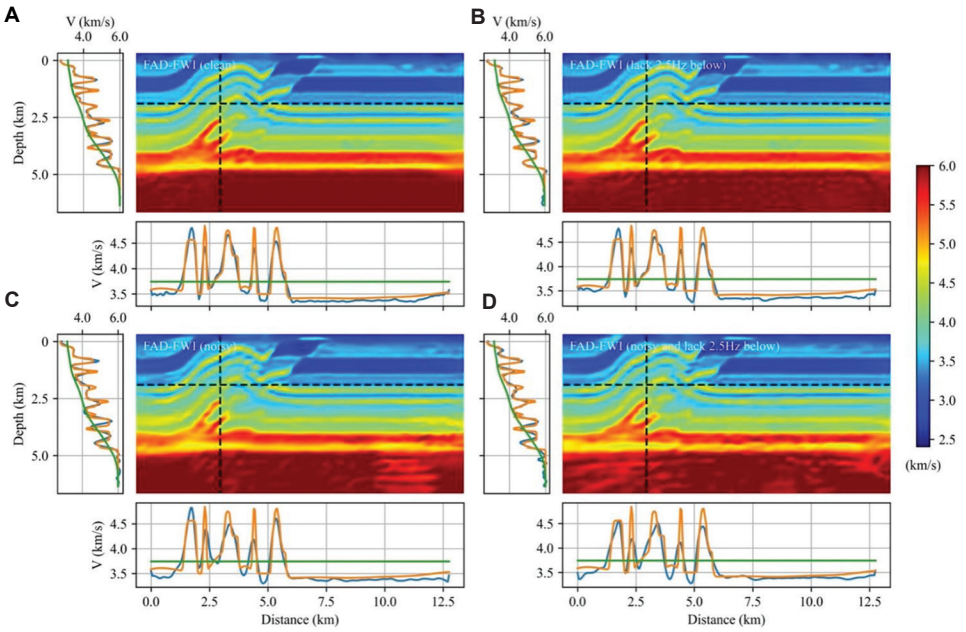


Figure 13. Comparison of inverted velocity models obtained using FAD-FWI given different observations with clean data (A), clean data with missing frequencies below 2.5 Hz (B), noisy data with random noises $\sigma = 0.5\sigma 0$ (C), and noisy data with missing frequencies below 2.5 Hz (D).
Abbreviations: FAD: Fully automatic differentiation; FWI: Full-waveform inversion.

Table 1. Quantitative evaluation metrics of the inverted velocity models obtained using DW-FWI and the proposed FAD-FWI under varying uncertainties in seismic observations

Methods	Metrics	MSE	SSIM	PSNR
DW-FWI	Full data	113.94	0.99	50.55
	Filtered data	200.87	0.98	48.09
	Noisy data	317.46	0.96	46.10
	Filtered noisy data	318.96	0.96	46.08
FAD-FWI	Full data	124.59	0.99	50.16
	Filtered data	132.03	0.99	49.91
	Noisy data	186.22	0.98	48.41
	Filtered noisy data	180.29	0.98	48.55

Abbreviations: DW: Deepwave; FAD: Fully automatic differentiation; FWI: Full-waveform inversion; MSE: Mean squared error; PSNR: Peak signal-to-noise ratio; SSIM: Structural similarity index measure.

runtime and memory usage for both DW-FWI and FAD-FWI. On an NVIDIA GeForce RTX 3080Ti (12 GB) GPU, the traditional DW-FWI method completed in 8 min and 31 s with a peak memory usage of 5.3 GB. In comparison, our proposed FAD-FWI method required 9 min and 6 s and 6.4 GB of memory. This represents a modest increase in runtime and memory usage for FAD-FWI, which is a reasonable trade-off given its significant improvements in accuracy and stability, as demonstrated by the quantitative metrics in Table 1.

4. Discussion

Our study proposes the FAD-FWI framework, an innovative approach to FWI that leverages U-Net reparameterization within an RNN-based paradigm. This approach demonstrates potential in handling challenging scenarios where conventional FWI struggles, such as noisy seismic data with missing low-frequency components and imprecise initial models. While the results affirm the robustness and flexibility of FAD-FWI, the method also presents some limitations and potential areas for enhancement. The primary bottleneck of the proposed FAD-FWI framework lies in the memory requirements associated with the RNN-based FWI. Using reverse-mode AD to compute gradients requires storing intermediate variables at each step, which significantly increases memory demands.[9,25,49] This constraint can be addressed by employing efficient boundary-saving techniques and checkpointing, which reduce memory requirements by selectively saving intermediate steps at the cost of increased computational overhead.[49] Balancing this trade-off between memory and computational demand is crucial for scaling FAD-FWI to larger, more complex models.

Furthermore, our FAD-FWI framework is general and

flexible, providing a foundation for integrating DNN-based parameterization and reformulation within FWI workflows. This versatility suggests promising applications beyond single-physics inversion. The framework can be extended to multi-physics coupled inversion and multi-data joint inversion, allowing for the incorporation of complementary data types (*e.g.*, electromagnetic, gravitational) to improve the resolution and accuracy of subsurface models.[9,50,51] Such extensions could enhance imaging and characterization in diverse geophysical applications, from reservoir monitoring to mineral exploration. In summary, the proposed FAD-FWI framework addresses some key challenges in FWI and shows potential for broad applicability. Future work will focus on optimizing memory efficiency and extending the FAD-FWI framework to multi-physics and joint inversion scenarios, further advancing seismic inversion and subsurface imaging capabilities in geophysics.

5. Conclusion

This study introduces a novel FAD-FWI framework that couples U-Net reparameterization within an RNN-based paradigm. Through a series of experiments, we demonstrated the superiority of our proposed FAD-FWI over conventional DW-FWI approach without DNN reparameterization, highlighting its robustness in scenarios with inaccurate initial models and in the presence of uncertainties in seismic observations, such as noise and missing frequency components. Recovering a velocity model from noisy seismic observations that lack low-frequency components and begin with a rough initial model is typically very challenging for conventional FWI methods. However, our proposed FAD-FWI achieves impressive performance in this demanding scenario. Our findings underscore the potential of deep learning techniques to significantly improve seismic inversion processes, thereby advancing subsurface imaging capabilities and contributing to more accurate geophysical explorations.

Acknowledgments

None.

Funding

This work was financially supported by the National Natural Science Foundation of China Program (42327803, 42304121, 42574154, U2344218), the Open Fund of National Engineering Research Center of Oil and Gas Exploration Computer Software, and the Natural Science Foundation of Wuhan (2025040601020137).

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Yufeng Wang
Formal analysis: Jingyi Zhao, Ying Yang
Investigation: Pengyuan Sun, Jun Zheng, Ying Yang
Methodology: Pengyuan Sun, Yufeng Wang
Visualization: Jun Zheng, Ying Yang
Writing–original draft: Pengyuan Sun, Jingyi Zhao
Writing–review & editing: Jingyi Zhao, Yufeng Wang

Availability of data

Data are available from the corresponding author upon reasonable request.

References

1. Tarantola A. Inversion of seismic reflection data in the acoustic approximation. *Geophysics*. 1984;49(8):1259-1266.

   doi: 10.1190/1.1441754

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

   doi: 10.1190/1.3238367

3. Zhang X, Lomas A, Zhou M, Zheng Y, Curtis A. 3-D Bayesian variational full waveform inversion. *Geophys J Int*. 2023;234(1):546-561.

   doi: 10.1093/gji/ggad057

4. Plessix RE. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophys J Int*. 2006;167(2):495-503.

   doi: 10.1111/j.1365-246X.2006.02978.x

5. Wang T, Cheng J, Geng J. Reflection full waveform inversion with second-order optimization using the adjoint-state method. *J Geophys Res Solid Earth*. 2021;126(8).

   doi: 10.1029/2021jb022135

6. Li YE, Demanet L. Full-waveform inversion with extrapolated low-frequency data. *Geophysics*. 2016;81(6):R339-R348.

   doi: 10.1190/geo2016-0038.1

7. Métivier L, Allain A, Brossier R, Mérigot Q, Oudet E, Virieux J. Optimal transport for mitigating cycle skipping in full-waveform inversion: A graph-space transform approach. *Geophysics*. 2018;83(5):R515-R540.

   doi: 10.1190/geo2017-0807.1

8. Teodor D, Comina C, Khosro Anjom F, Brossier R, Socco LV, Virieux J. Challenges in shallow target reconstruction by

3D elastic full-waveform inversion - which initial model? *Geophysics*. 2021;86(4):R433-R446.

doi: 10.1190/geo2019-0523.1

9. Li D, Xu K, Harris JM, Darve E. Coupled time-lapse full-waveform inversion for subsurface flow problems using intrusive automatic differentiation. *Water Resour Res*. 2020;56(8):e2019WR027032.

doi: 10.1029/2019wr027032

10. Zhu W, Xu K, Darve E, Beroza GC. A general approach to seismic inversion with automatic differentiation. *Comput Geosci*. 2021;151:104751.

doi: 10.1016/j.cageo.2021.104751

11. Adler A, Araya-Polo M, Poggio T. Deep learning for seismic inverse problems: Toward the acceleration of geophysical analysis workflows. *IEEE Signal Process Mag*. 2021;38(2): 89-119.

doi: 10.1109/msp.2020.3037429

12. Yu S, Ma J. Deep learning for geophysics: Current and future trends. *Rev Geophys*. 2021;59(3):e2021RG000742.

doi: 10.1029/2021rg000742

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

doi: 10.1190/geo2018-0249.1

14. Deng C, Feng S, Wang H, *et al*. OpenFWI: Large-scale multi-structural benchmark datasets for full waveform inversion. *Adv Neural Inf Process Syst*. 2022;35:6007-6020.

15. Lewis W, Vigh D. Deep Learning Prior Models from Seismic Images for Full-Waveform Inversion. In: *SEG International Exposition and Annual Meeting*. SEG; 2017.

16. Sun H, Demanet L. Extrapolated full-waveform inversion with deep learning. *Geophysics*. 2020;85(3):R275-R288.

doi: 10.1190/geo2019-0195.1

17. Zhang W, Gao J, Gao Z, Chen H. Adjoint-driven deep-learning seismic full-waveform inversion. *IEEE Trans Geosci Remote Sens*. 2020;59(10):8913-8932.

doi: 10.1109/tgrs.2020.3044065

18. Zhang ZD, Alkhalifah T. High-resolution reservoir characterization using deep learning-aided elastic full-waveform inversion: The North Sea field data example. *Geophysics*. 2020;85(4):WA137-WA146.

doi: 10.1190/geo2019-0340.1

19. Li Y, Alkhalifah T, Zhang Z. Deep-learning assisted regularized elastic full waveform inversion using the velocity distribution information from wells. *Geophys J Int*. 2021;226(2):1322-1335.

doi: 10.1093/gji/ggab162

20. Wang F, Huang X, Alkhalifah TA. A prior regularized full waveform inversion using generative diffusion models. *IEEE Trans Geosci Remote Sens*. 2023;61:1-11.

doi: 10.1109/tgrs.2023.3337014

21. Feng S, Lin Y, Wohlberg B. Multiscale data-driven seismic full-waveform inversion with field data study. *IEEE Trans Geosci Remote Sens*. 2021;60:1-14.

doi: 10.1109/tgrs.2021.3114101

22. Rasht-Behesht M, Huber C, Shukla K, Karniadakis GE. Physics-informed neural networks (PINNs) for wave propagation and full waveform inversions. *J Geophys Res Solid Earth*. 2022;127(5):e25.

doi: 10.1029/2021jb023120

23. Richardson A. *Seismic Full-Waveform Inversion Using deep Learning Tools and Techniques*. [arXiv Preprint]; 2018.

doi: 10.48550/arXiv.1801.07232

24. Yang Y, Gao AF, Azizzadenesheli K, Clayton RW, Ross ZE. Rapid seismic waveform modeling and inversion with neural operators. *IEEE Trans Geosci Remote Sens*. 2023;61:1-12.

doi: 10.1109/tgrs.2023.3264210

25. Sun J, Niu Z, Innanen KA, Li J, Trad DO. A theory-guided deep-learning formulation and optimization of seismic waveform inversion. *Geophysics*. 2020;85(2):R87-R99.

doi: 10.1190/geo2019-0138.1

26. Song C, Wang Y, Richardson A, Liu C. Weighted envelope correlation-based waveform inversion using automatic differentiation. In: *IEEE Transactions on Geoscience and Remote Sensing*. Vol. 61. New York: IEEE; 2023.

doi: 10.1109/tgrs.2023.3300127

27. Fang J, Zhou H, Elita Li Y, Shi Y, Li X, Wang E. Deep-learning optimization using the gradient of a custom objective function: A full-waveform inversion example study on the convolutional objective function. *Geophysics*. 2024;89(5):R479-R492.

doi: 10.1190/geo2023-0538.1

28. Zhang Y, Zhu X, Gao J. Seismic inversion based on acoustic wave equations using physics-informed neural network. *IEEE Trans Geosci Remote Sens*. 2023;61:1-11.

doi: 10.1109/tgrs.2023.3236973

29. Schuster GT, Chen Y, Feng S. Review of physics-informed machine-learning inversion of geophysical data. *Geophysics*. 2024;89(6):1-91.

doi: 10.1190/geo2023-0615.1

30. Li Z, Kovachki N, Azizzadenesheli K, *et al*. *Fourier Neural Operator for Parametric Partial Differential Equations*. [arXiv Preprint]; 2020.

doi: 10.48550/arxiv.2010.08895

31. Lu L, Jin P, Pang G, Zhang Z, Karniadakis GE. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nat Mach Intell*. 2021;3(3):218-229.

32. Guo Z, Chai L, Huang S, Li Y. *Inversion-DeepONet: A Novel DeepONet-Based Network with Encoder-Decoder for Full Waveform Inversion*. [arXiv Preprint]; 2024.

    doi: 10.48550/arXiv.2408.08005

33. Ulyanov D, Vedaldi A, Lempitsky V. Deep Image Prior. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2018. p. 9446-9454.

34. Dittmer S, Kluth T, Maass P, Baguer DO. Regularization by architecture: A deep prior approach for inverse problems. *J Math Imaging Vision*. 2020;62:456-470.

35. Xie Y, Chen W, Ge H, Ng MK. Deep image prior and weighted anisotropic-isotropic total variation regularization for solving linear inverse problems. *Appl Math Comput*. 2024;482:128952.

    doi: 10.1016/j.amc.2024.128952

36. Wu Y, McMechan GA. Parametric convolutional neural network-domain full-waveform inversion. *Geophysics*. 2019;84(6):R881-R896.

    doi: 10.1190/geo2018-0224.1

37. He Q, Wang Y. Reparameterized full-waveform inversion using deep neural networks. *Geophysics*. 2021;86(1): V1-V13.

    doi: 10.1190/geo2019-0382.1

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

    doi: 10.1190/geo2020-0933.1

39. Dhara A, Sen MK. Elastic full-waveform inversion using a physics-guided deep convolutional encoder-decoder. In: *IEEE Transactions on Geoscience and Remote Sensing*. New York: IEEE; 2023.

    doi: 10.1109/tgrs.2023.3294427

40. Sun J, Innanen K, Zhang T, Trad D. Implicit seismic full waveform inversion with deep neural representation. *J Geophys Res Solid Earth*. 2023;128(3):e2022JB025964.

    doi: 10.1029/2022jb025964

41. Herrmann L, Bürchner T, Dietrich F, Kollmannsberger S. On the use of neural networks for full waveform inversion. *Comput Methods Appl Mech Eng*. 2023;415:116278.

    doi: 10.1016/j.cma.2023.116278

42. Arridge S, Maass P, Öktem O, Schönlieb CB. Solving inverse problems using data-driven models. *Acta Numerica*. 2019;28:1-174.

    doi: 10.1017/s0962492919000059

43. Sun P, Yang F, Liang H, Ma J. Full-waveform inversion using a learned regularization. In: *IEEE Transactions on Geoscience and Remote Sensing*. New York: IEEE; 2023.

    doi: 10.1109/tgrs.2023.3322964

44. Lord GJ, Powell CE, Shardlow T. *An Introduction to Computational Stochastic PDEs*. Vol. 50. Cambridge: Cambridge University Press; 2014.

45. Benitez JAL, Furuya T, Faucher F, Tricoche X, Hoop M. *Fine-Tuning Neural-Operator Architectures for Training and Generalization*. [arXiv Preprint]; 2023.

46. Ronneberger O, Fischer P, 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, 2015, Proceedings, Part III 18*. Berlin: Springer; 2015. p. 234-241.

47. Siddique N, Paheding S, Elkin CP, Devabhaktuni V. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access*. 2021;9:82031-82057.

48. Baydin AG, Pearlmutter BA, Radul AA, Siskind JM. Automatic differentiation in machine learning: A survey. *J Mach Learn Res*. 2018;18(153):1-43.

49. Wang S, Jiang Y, Song P, Tan J, Liu Z, He B. Memory optimization in RNN-based full waveform inversion using boundary saving wavefield reconstruction. *IEEE Trans Geosci Remote Sens*. 2023;61:1-12.

    doi: 10.1109/tgrs.2023.3317529

50. Yin Z, Orozco R, Louboutin M, Herrmann FJ. Solving multiphysics-based inverse problems with learned surrogates and constraints. *Adv Model Simul Eng Sci*. 2023;10(1):14.

51. Liu M, Vashisth D, Grana D, Mukerji T. Joint inversion of geophysical data for geologic carbon sequestration monitoring: A differentiable physics-informed neural network model. *J Geophys Res Solid Earth*. 2023;128(3):e2022JB025372.

    doi: 10.1029/2022jb025372

ACCSCIENCE PUBLISHING

Journal of Seismic Exploration

ARTICLE

Porosity prediction with Bi-LSTM network for deep methane reservoirs

Qiang Guo[1], Xinyu Zhao[1], Jing Ba[2], and Cong Luo[2]*

[1]School of Resources and Geosciences, China University of Mining and Technology, Xuzhou, Jiangsu, China

[2]School of Earth Sciences and Engineering, Hohai University, Nanjing, Jiangsu, China

(This article belongs to the *Special Issue: Geophysical Inversion and Intelligent Prediction Technologies for Complex Hydrocarbon Reservoirs*)

Abstract

Quantitative prediction of petrophysical parameters, such as porosity, is crucial for the evaluation and development of coalbed methane (CBM) reservoirs. However, conventional methods based on linear assumptions and empirical formulas often fall short due to the strong heterogeneity of coal seams, complex lithologies and structures, and the highly non-linear relationship between seismic elastic parameters and reservoir properties under deep-buried conditions. While machine learning techniques have shown promise in petrophysical prediction, many existing approaches struggle to effectively capture long-range dependencies within sequential log data. This study proposes a deep learning-based method that integrates comprehensive input feature selection with a bidirectional long short-term memory (Bi-LSTM) network incorporating dropout regularization for enhanced petrophysical parameter prediction. The proposed method is designed to fully exploit the non-linear mapping between seismic elastic parameters (e.g., P-wave velocity, S-wave velocity, density, elastic impedance) and petrophysical parameter (porosity). By combining the bidirectional contextual learning capability of Bi-LSTM, the model effectively captures feature relationships within depth sequences. Comparative analysis against a fully connected neural network and a standard LSTM network demonstrates the superiority of the proposed method. The analysis also reveals the optimal feature combination and network parameter setting (sequential length, sampling interval, etc.). Results indicate that the Bi-LSTM model achieves a significant improvement in prediction accuracy, outperforming other models, and demonstrating better generalization capability in blind well tests. The method provides a reliable and effective tool for quantitative reservoir characterization, offering substantial potential for application in deep CBM exploration.

Keywords: Deep coalbed methane; Porosity prediction; Deep learning; LSTM network

***Corresponding author:**
Cong Luo
(lishang0228@163.com)

Citation: Guo Q, Zhao X, Ba J, Luo C. Porosity prediction with Bi-LSTM network for deep coalbed methane reservoirs. *J Seismic Explor*. 2025;34(6):29-44. doi: 10.36922/JSE025410087

Received: October 12, 2025

Revised: October 28, 2025

Accepted: October 30, 2025

Published online: November 19, 2025

1. Introduction

Coalbed methane (CBM), as an important component of unconventional natural gas, is currently one of the hotspots in natural gas exploration.[1] Its efficient exploration and development have become a critical pathway for increasing reserve and optimizing energy structure.[2] Petrophysical parameters of CBM reservoirs, such as porosity,

are key for characterizing reservoir quality, predicting production potential, and formulating development plans.[3,4] Previous studies calculated and predicted reservoir porosity in unknown intervals by establishing empirical formulas or simplified geological models.[5] However, due to factors such as strong coal heterogeneity, complex reservoir structure, and ambiguity in well log responses,[6,7] conventional seismic prediction methods based on linear assumptions or statistical relation are limited for detailed quantitative reservoir evaluation.[8-11] Particularly in deep CBM exploration, high temperature and pressure conditions further intensify the non-linear characteristics of rock physics relationships, making conventional prediction methods inadequate for refined reservoir characterization.[12,13] In general, the heterogeneity and complexity of CBM reservoirs cause the relationship between porosity and elastic parameters to vary significantly under different geological conditions. Traditional linear methods are unable to adapt to these variations, resulting in reduced prediction accuracy. Therefore, the accurate formulation of non-linear relation between seismic elastic parameters and petrophysical parameters is crucial for the quantitative evaluation of deep CBM reservoirs.

With the rapid development of artificial intelligence, an increasing number of machine learning methods have been applied to porosity prediction. Wu *et al.*[14] used an optimized RBF neural network to predict reservoir porosity models from well data, achieving high prediction accuracy. Ahmadi *et al.*[15] proposed a GA-LSSVM model optimized by a genetic algorithm for reservoir porosity prediction to establish more reliable static reservoir simulation models. Zerrouki *et al.*[16] employed an artificial neural network combined with a fuzzy ranking method to predict fracture porosity. Cao *et al.*[17] investigated the use of an extreme learning machine for estimating porosity and permeability in heterogeneous sandstone reservoirs. Zou *et al.*[18] utilized a random forest-based method to predict pore distribution in subsurface reservoirs.

In recent years, the rapid development of deep learning technology has demonstrated significant potential in geophysical exploration.[19-22] Deep learning techniques possess powerful feature extraction and high-dimensional data processing capabilities, enabling effective mining of deeper features from large datasets.[23-25] Their strong ability to learn complex non-linear relationships allows for more accurate approximation of the highly non-linear relationships between seismic/elastic data and target parameters.[26] Wang *et al.*[27] employed a Gaussian Mixture Model Deep Neural Network for porosity prediction, with experimental results showing its capability to reasonably estimate porosity distribution across the entire target area. Wu *et al.*[28] proposed a joint inversion method based on fluid factor and brittleness index. They developed a new P-P wave reflection coefficient approximation formula specifically for coal-measure gas reservoirs and combined it with a Bayesian inversion framework, effectively enhancing the comprehensive evaluation of gas-bearing potential and fracability.[28] Liu *et al.*[29] incorporated a low-frequency porosity model into a deep learning framework, significantly improving the trend continuity and generalization ability of porosity prediction in carbonate reservoirs. Zhang *et al.*[30] optimized the pore aspect ratio using the deep learning network aided by the Hunger Games Search algorithm to achieve joint inversion of multiple parameters in tight sandstone reservoirs, effectively improving the accuracy and reliability of rock physics modeling and inversion. Sun *et al.*[31] proposed a CNN-Transformer model aimed at improving the accuracy and generalizability of log-based porosity prediction. Tao *et al.*[32] introduced a UNet-based bidirectional neural network method to establish a mapping relationship between seismic data and porosity. While these methods have, to some extent, improved the accuracy and interpretability of porosity prediction under complex reservoir conditions, they cannot effectively handle long-range information in sequence data and fail to capture the relationships of reservoir features in deep sequences.

To address the aforementioned issues, this paper proposes a method combining input feature selection and a bidirectional long short-term memory (Bi-LSTM) network for petrophysical parameter prediction. Comparisons are made with fully connected neural (FCN) networks and unidirectional LSTM networks. The proposed method not only fully exploits the non-linear relationship between seismic elastic parameters and reservoir petrophysical parameters, but is also more sensitive to the contextual correlations within reservoir information sequences. Consequently, it can accurately capture the relationships of reservoir features within depth sequences. Furthermore, the analysis on feature selection and network parameter setting (such as sequence length and sampling interval) also provided practical guidance for deep learning-based seismic prediction of CBM reservoirs.

2. Methodologies

2.1. Fully-connected neural network

FCN network is a basic form of deep learning networks. FCN is composed of multiple layers of neurons, where each neuron in the current layer is connected to every neuron in the subsequent layer. A typical neuron receives multiple input signals, computes their weighted sum, introduces non-linearity through an activation function,

and ultimately produces an output signal. This process can be mathematically expressed as:

$$y = f(\sum_{i=1}^{n} w_i x_i + b) \quad \text{(I)}$$

Where x_i represents the input signal; w_i denotes the weight of the input signal, reflecting its importance to the neuron output; b is the bias term, which adjusts the activation threshold of the neuron; n is the dimensionality of the input features; $f(\cdot)$ is the activation function, which provides non-linear transformations; and y is the output signal of the neuron. Figure 1 shows a schematic diagram of a simple FCN network with an input feature of three dimensions, an output of one dimension, and three hidden layers.

The training of an FCN network involves four key steps.[33,34] First, in forward propagation, input feature passes through the network, undergoing weighted sums and activation functions at each layer to generate a prediction. The loss function then compares this prediction to the true value. Next, backpropagation calculates the gradient of the loss with respect to all network parameters using the chain rule. Finally, these gradients are used by an optimization algorithm to update the weights and biases. This cycle repeats until the loss converges or a maximum iteration is reached.

2.2. Long short-term memory

The long short-term memory (LSTM) network is featured by capturing long-term dependencies in sequential data by introducing a gating mechanism.[35] The core component of an LSTM is a memory cell, which contains three gates: a forget gate, an input gate, and an output gate. These gates regulate the flow of information into, within, and out of the cell, enabling the network to learn and maintain long-range dependencies. The structure of a single LSTM cell is illustrated in Figure 2.

The procedure of an LSTM network can be summarized in the following steps:

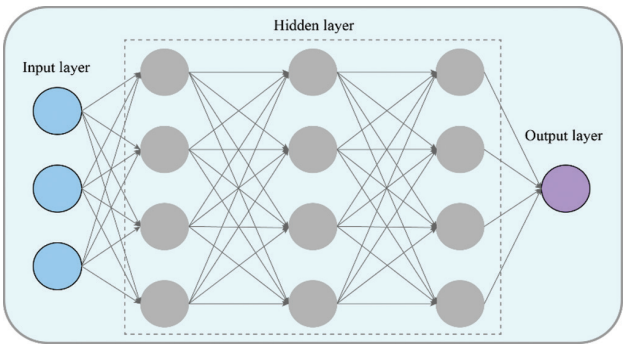


Figure 1. Structure of a simple fully connected neural network

Step 1—Compute the forget gate: This gate determines what information to discard from the cell state, indicating the degree of information retention. It is computed using a sigmoid activation function, which produces an output between 0 and 1 as:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad \text{(II)}$$

Where f_t is the output of the forget gate; σ is the sigmoid activation function; h_{t-1} and x_t represent the hidden state from the previous timestep and the input at the current timestep, respectively; W_f and b_f represent the weight matrix and bias term of the forget gate.

Step 2—Compute the input gate: This gate decides what new information will be stored in the cell state. The calculations take the form as:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad \text{(III)}$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad \text{(IV)}$$

Where i_t is the activation vector of the input gate, determining which values to update; $\tilde{C}_t$ is the candidate value vector, determining the new values to be added; is the hyperbolic tangent activation function; W_i, W_c, b_i and b_c represent the weight matrices and bias terms for the input gate and candidate values, respectively.

Step 3—Update the cell state: The cell state, which embodies the long-term memory of the model, is updated as follows:

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \quad \text{(V)}$$

Where c_t is the current cell state; f_t is the output of the forget gate, representing the information to be discarded; c_{t-1} is the cell state from the previous timestep; i_t is the activation vector of the input gate, representing the information to be updated; and $\tilde{C}_t$ is the candidate value vector.

Step 4—Compute the output gate: This gate determines the value of the next hidden state. The hidden state contains information about the previous timestep and can be used for predicting the output at the next timestep as:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad \text{(VI)}$$

$$h_t = o_t \cdot \tanh(c_t) \quad \text{(VII)}$$

Where o_t is the output of the output gate; h_t is the hidden state at the current timestep; and W_o and b_o represent the weight matrix and bias term of the output gate.

In particular, the Bi-LSTM network is an extension of the standard LSTM. It incorporates two separate LSTM layers: One processing the input sequence in the forward direction

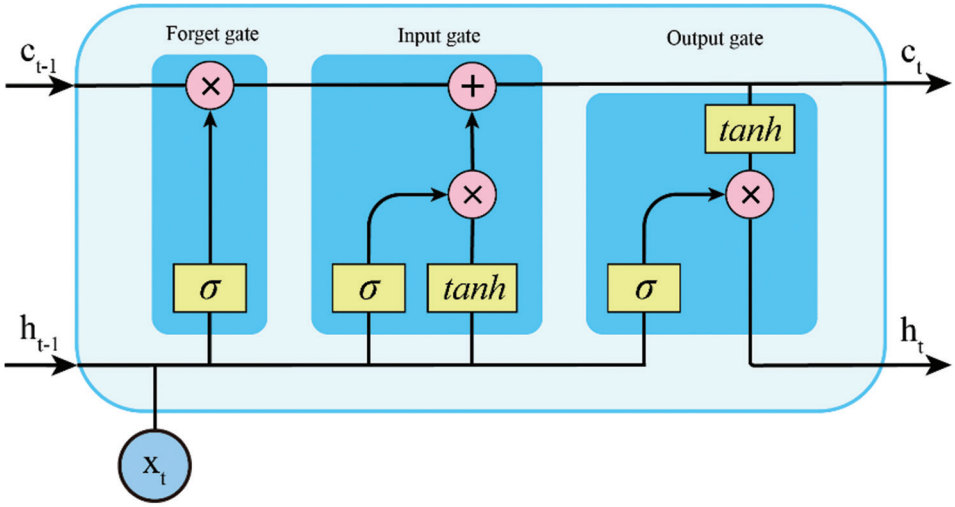


Figure 2. Structure of a single long short-term memory cell

and the other processing it in the reverse direction. The final output is generated by merging (e.g., concatenating or summing) the outputs from both directions (Figure 3). This architecture enables the model to capture dependencies from both past and future contexts simultaneously.

2.3. Activation and loss functions

The activation function is a crucial component in neural networks. Its primary role is to introduce non-linearity, enabling neural networks to learn and represent complex non-linear relationships. Common activation functions include the ReLU function, the Tanh function, and the Sigmoid function. Here, we employ the ReLU function as the activation function, which takes the form as:

$$f(x) = \max\{0, x\} \tag{VIII}$$

The loss function, aiming at training neural networks, quantifies the discrepancy between model predictions and true values, thereby driving the optimization of network parameters. Here, we employ the mean squared error to formulate the loss function, which takes the form as:

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{d}\sum_{j=1}^{d}\left(\hat{y}_j^{(i)} - y_j^{(i)}\right)^2 \tag{IX}$$

Where L_{MSE} represents the average loss over the entire training batch, $\hat{y}_j^{(i)}$ is the predicted value, $y_j^{(i)}$ denotes the ground truth labels, N is the number of samples in the batch, and d indicates the dimensionality of the vectors.

2.4. Workflow

In this study, two different deep learning networks—FCN and LSTM—were employed for predicting petrophysical parameters from well log data. The overall workflow is illustrated in Figure 4. First, after acquiring true log data, dataset preparation was conducted, analyzing the effect of different intervals and sampling rates on prediction accuracy. Subsequently, feature selection was performed using various combinations of elastic parameters from the log data—such as S-wave velocity, P-wave velocity, density, P-to-S-wave velocity ratio, S-wave impedance, and P-wave impedance—as inputs, while using porosity as the network output, to identify the optimal combination of input features for training. Then, the prediction accuracy of the two network models was compared to determine the more suitable model for petrophysical parameter prediction, wherein the effect of sequence length of LSTM on prediction accuracy was also analyzed. Finally, blind well testing was conducted to evaluate the effectiveness of the proposed method. In addition, an attempt was made to introduce Bi-LSTM to enhance prediction accuracy and incorporate dropout to mitigate overfitting during model training.

3. Tests and applications

3.1. Dataset preparation

A total of 45,606 data points from well log measurements acquired across six wells in the study area were compiled to form the dataset. The data underwent min-max normalization, scaling all feature values to the range of 0 to 1. Outliers were removed based on the 3σ rule. These preprocessing steps ensured data quality and provided a reliable foundation for model training. The data were measured from a deep CBM reservoir in north China, with the target coal layer buried at a depth around 2000 m. It includes data from different geological settings, such as

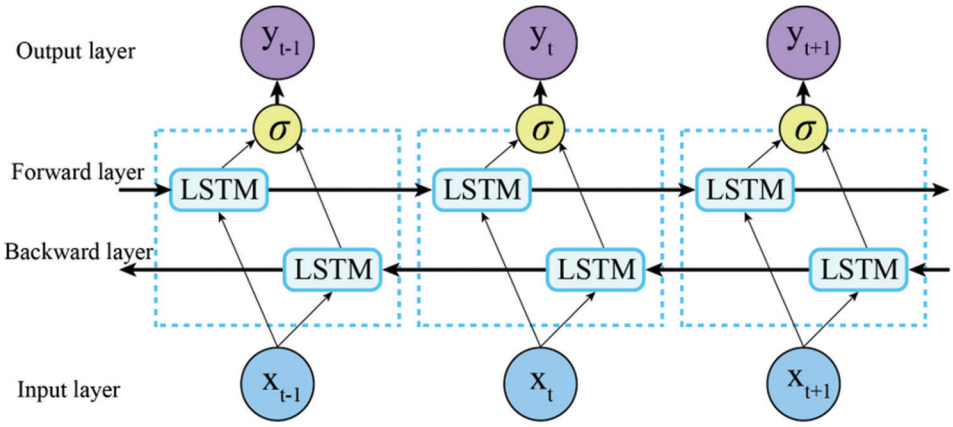


Figure 3. Structure of bidirectional long short-term memory cell

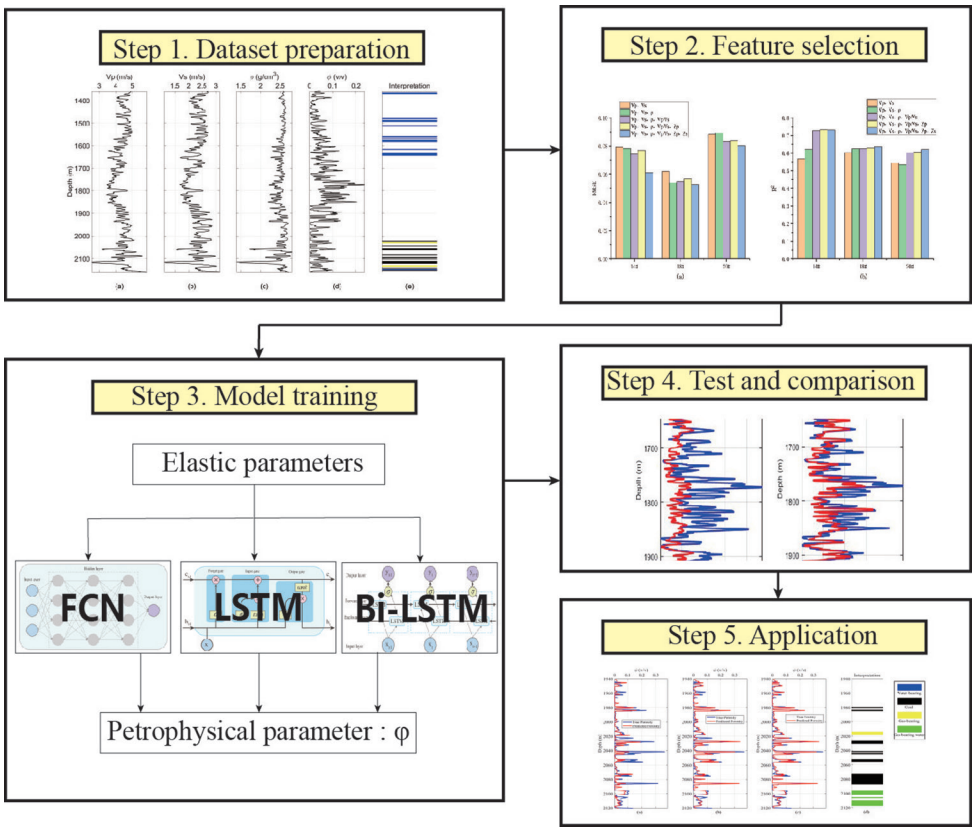


Figure 4. Workflow of the study
Abbreviations: Bi-LSTM: Bidirectional long short-term memory; FCNN: Fully connected neural network; LSTM: Long short-term memory

varying coal thickness, fracture development, and pore structures. This diversity provides a solid foundation for model training and validation, ensuring prediction accuracy and generalization under different geological conditions. The log curves and corresponding lithofacies interpretations for the selected Well B and Well C are shown in Figures 5 and 6, respectively, which exhibit a complex relationship between elastic and petrophysical properties, especially for coal sections. These data points, which include all necessary variables, are suitable for training and testing our models. This study used 80% of the data for training and 20% for validation, with random splitting to ensure consistent distribution between training and validation sets, thereby improving model generalization.

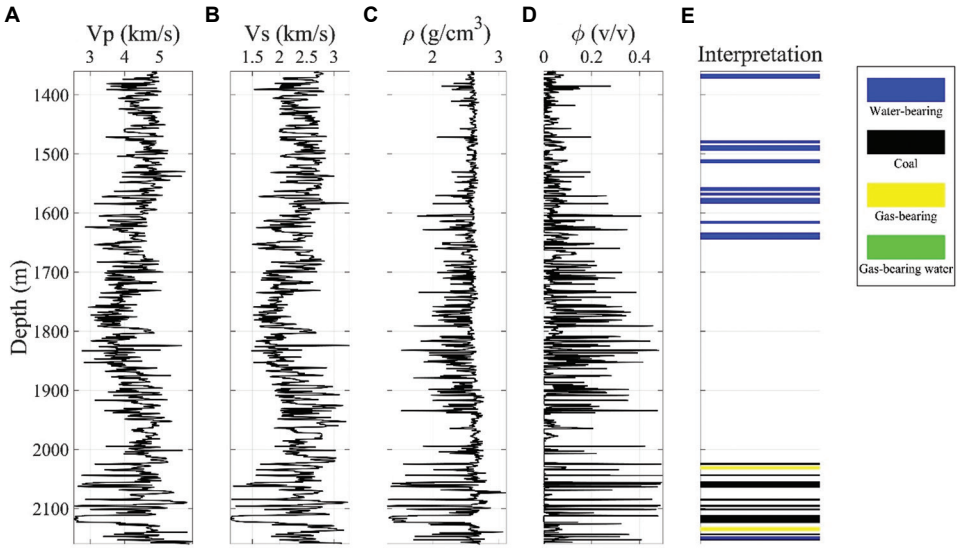


Figure 5. Log curve of P-wave velocity (A), S-wave velocity (B), density (C), porosity (D), and lithofacies interpretation result (E), for Well B

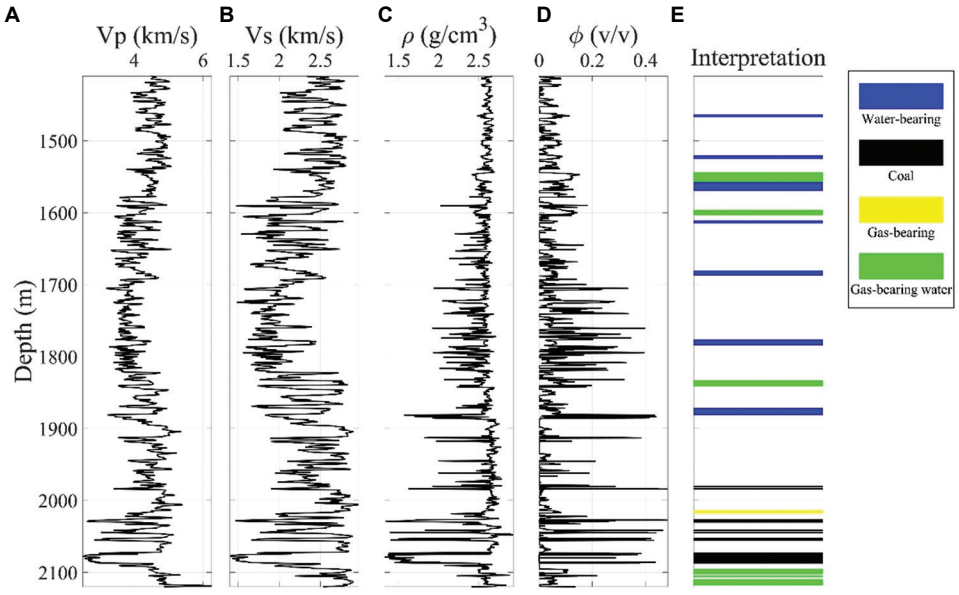


Figure 6. Log curve of P-wave velocity (A), S-wave velocity (B), density (C), porosity (D), and lithofacies interpretation result (E), for Well C

The selection of the dataset for training plays a crucial role in the accuracy and reliability of the network. We conducted a preliminary analysis on the effect of using different stratigraphic sections and sampling intervals on prediction accuracy. The analysis was performed on standard FCN networks with P-, S-wave velocities, and density as input features and porosity as output. The root mean square error (RMSE) and the coefficient of determination (R^2) from various wells were employed as evaluation metrics for prediction performance. As evidenced by the data presented in Figure 7 and Table 1, utilizing the coal section for training can effectively enhance model accuracy. It is due to the highly non-linear relation between elastic properties and porosity primarily exists in coal sections. Moreover, the analysis suggests that appropriately increasing the sampling interval can reduce the prediction error (Figure 8 and Table 2).

3.2. Feature selection

To determine the optimal input features for training, we evaluated five combinations of elastic parameters, i.e., [V_P, V_S], [V_P, V_S, ρ], [V_P, V_S, ρ, V_P/V_S], [V_P, V_S, ρ, V_P/V_S,

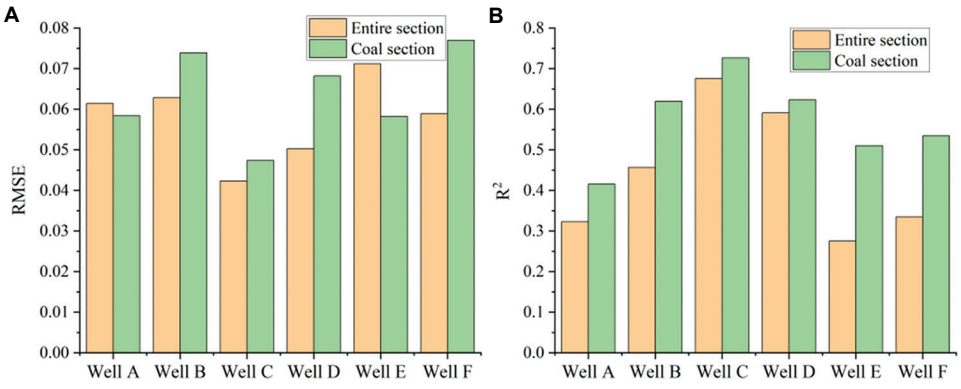


Figure 7. Prediction accuracy in term of root mean square error (A) and R^2 (B) using the training data from different stratigraphic sections

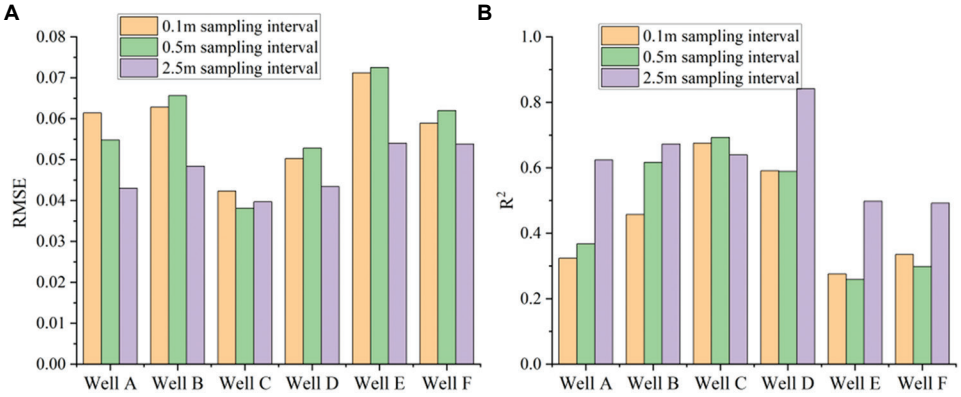


Figure 8. Prediction accuracy in term of root mean square error (A) and R^2 (B) using the training data with different sampling intervals (0.1, 0.5, and 2.5 m)

Table 1. Prediction accuracy with different stratigraphic sections

Stratigraphic section	Mean RMSE	Mean R^2
Entire section	0.0578	0.4429
Coal section	0.0639	0.5714

Abbreviation: RMSE: Root mean square error.

Table 2. Prediction accuracy with different sampling intervals

Sampling interval	Mean RMSE	Mean R^2
0.1 m	0.0578	0.4429
0.5 m	0.0576	0.4706
2.5 m	0.0470	0.6279

Abbreviation: RMSE: Root mean square error.

Z_p], and [V_P, V_S, ρ, V_P/V_S, Z_P, Z_S]. Each combination was used to train the network, and the model performance was validated using test data. When selecting certain wells for testing, the remaining wells are used to train the model. In this experiment, Wells B, D, and F within the study area were selected as the test data, respectively. The true and predicted values were recorded, and the corresponding scatter plots were shown in Figures 9-11. The RMSE and R^2 were calculated to assess the prediction accuracy and identify the optimal input feature combination. To enhance the training outcome, data sampled at an interval of 2.5 m, as suggested by the analysis in Section 3.1, was adopted as the dataset for this experiment.

As observed from the scatter plots in Figures 9-11, the prediction accuracy varies with different combinations of input features, which can be inferred by comparing the predicted against true values with the diagonal reference line. Figure 12 and Table 3 present a comparison of prediction accuracy under these different combinations. The results indicate that using Combination 5—that is, the six parameters V_P, V_S, ρ, V_P/V_S, Z_P, Z_S as input features—yields the best performance, achieving an average RMSE of 0.0647 and an average R^2 of 0.6574, which represents the highest relative accuracy among the five combinations. To better illustrate the prediction performance using different combinations, Figure 13 compares the true and predicted porosity curves for Well B when using the input feature of Combination 1 and Combination 5 with the sampling interval of 2.5 m, respectively, which indicates a significant reducing of prediction error by the optimal feature selection.

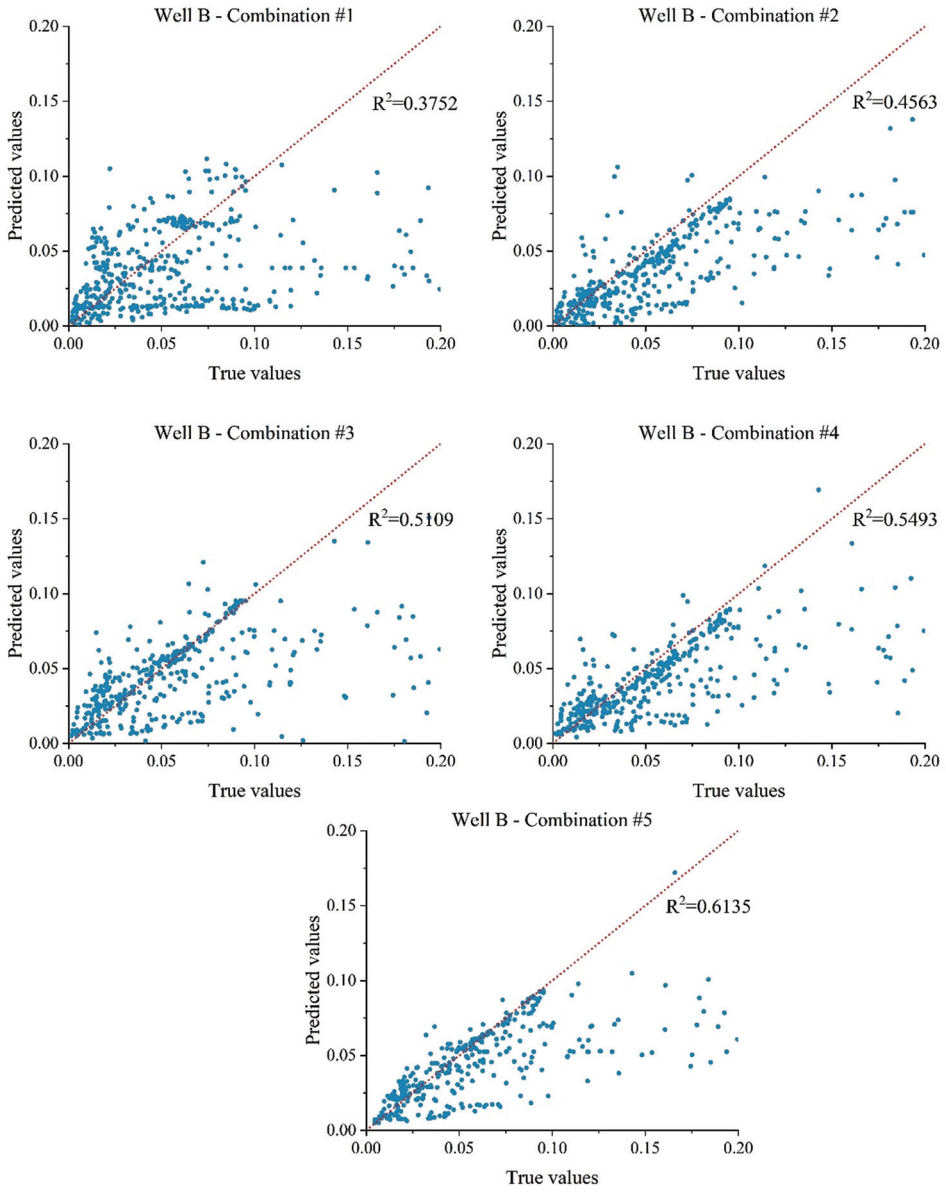


Figure 9. Prediction results with different input feature combinations for Well B

Table 3. Prediction accuracy with different combinations of input feature

Combination	Features	Mean RMSE	Mean R^2
1	[V_P, V_S]	0.0767	0.4078
2	[V_P, V_S, ρ]	0.0737	0.5117
3	[V_P, V_S, ρ, V_P/V_S]	0.0708	0.5710
4	[V_P, V_S, ρ, V_P/V_S, Z_P]	0.0724	0.6283
5	[V_P, V_S, ρ, V_P/V_S, Z_P, Z_S]	0.0647	0.6574

Abbreviation: RMSE: Root mean square error.

3.3. Model tests

To compare the performance of FCN and LSTM networks in petrophysical prediction, we conducted FCN and LSTM models and applied them to the prediction, respectively. The FCN adopts a 5-layer architecture with hidden layer sizes of 32-64-128-64-32 and uses the ReLU activation function. The LSTM model has a hidden size of 64, comprises 2 stacked layers, and is followed by a fully connected layer for output. Both networks have a dropout rate set to 0.2

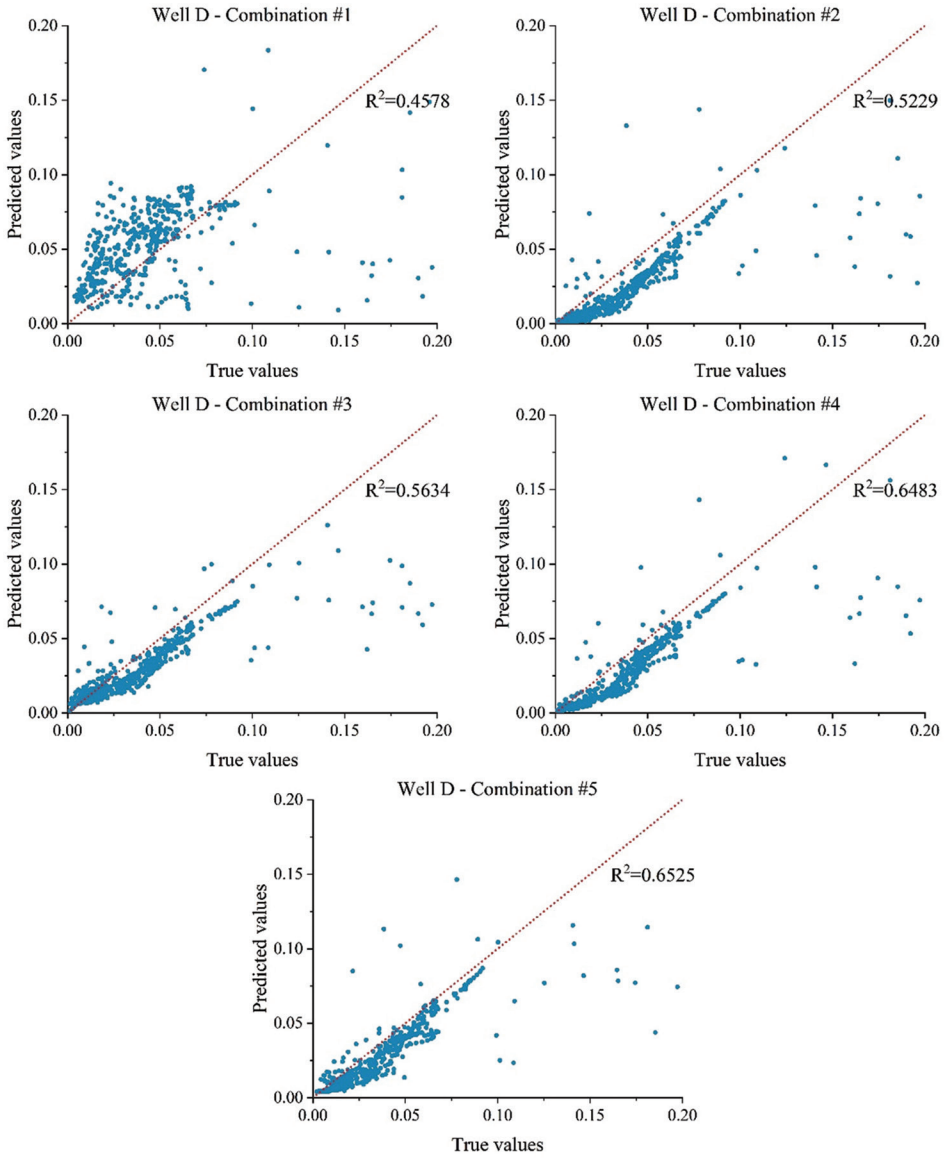


Figure 10. Prediction results with different input feature combinations for Well D

and were trained under identical conditions: Each training session employed a unified loss function, optimizer, learning rate, and batch size. We used the Adam optimizer with a learning rate of 0.001. The optimizer is crucial for efficiently and stably updating network parameters to minimize the loss. The learning rate is adjusted based on model convergence speed and stability to ensure optimal performance within a reasonable time. The models were trained by iteratively updating weights using the same training, validation, and test sets. The RMSE and R^2 for each test well were calculated to assess prediction accuracy.

As indicated by the data in Figure 14 and Table 4, the LSTM network demonstrates superior performance over the FCN network in most wells, with lower RMSE and higher R^2 values, indicating its greater suitability for petrophysical parameter prediction, especially for well-measured sequential data.

Table 4. Prediction accuracy with different network models

Network model	Mean RMSE	Mean R^2
FCN	0.0691	0.5705
LSTM	0.0621	0.6125

Abbreviations: FCN: Fully connected neural; LSTM: Long Short-Term Memory.

When training the LSTM model, the sequence length, which defines the number of sequence length in each input

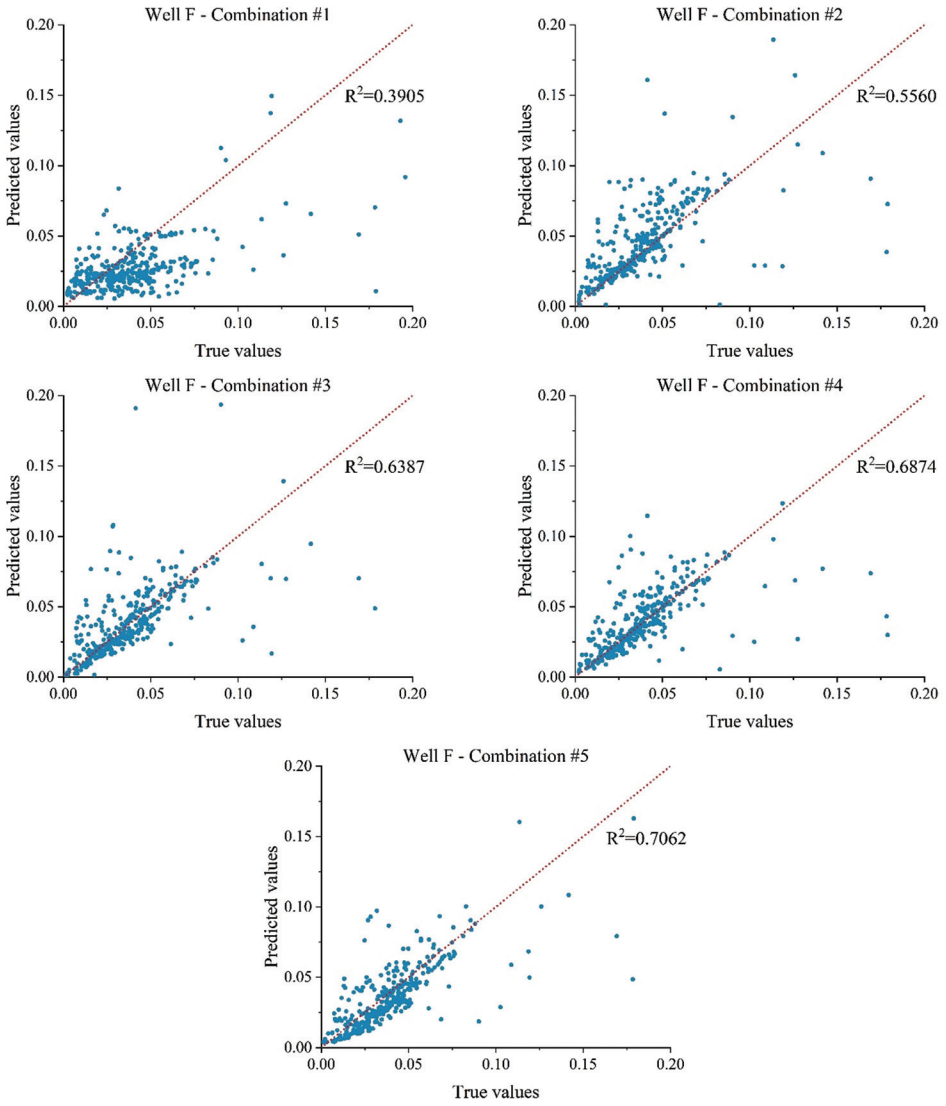


Figure 11. Prediction results with different input feature combinations for Well F

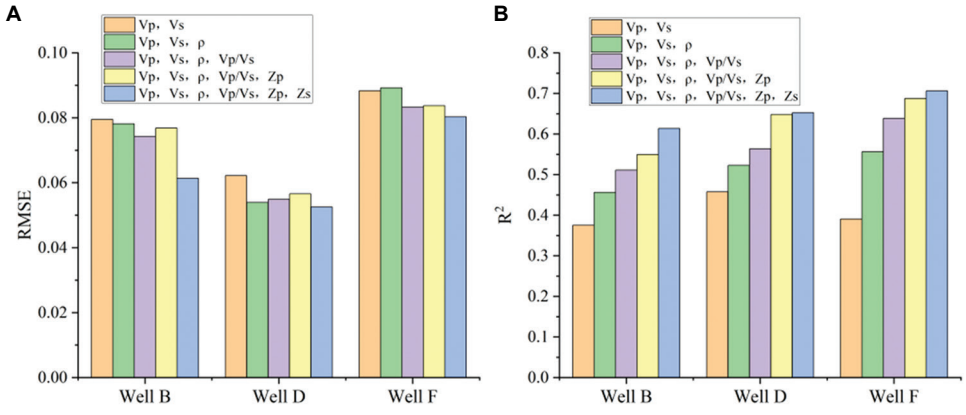


Figure 12. Prediction accuracy in term of root mean square error (A) and R^2 (B) with different combinations of input feature

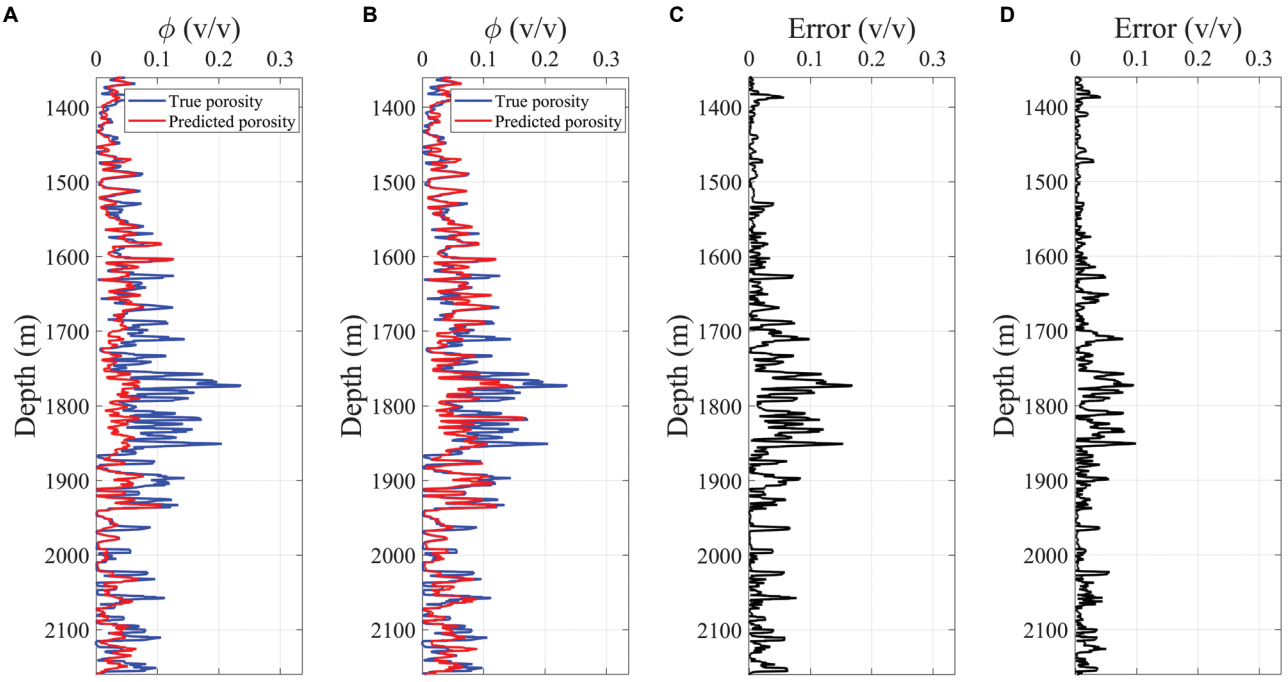


Figure 13. Comparison between the true and predicted porosity using the input feature of Combination 1 (A) and Combination 5 (B) for Well B, and their corresponding absolute residual errors (C and D)

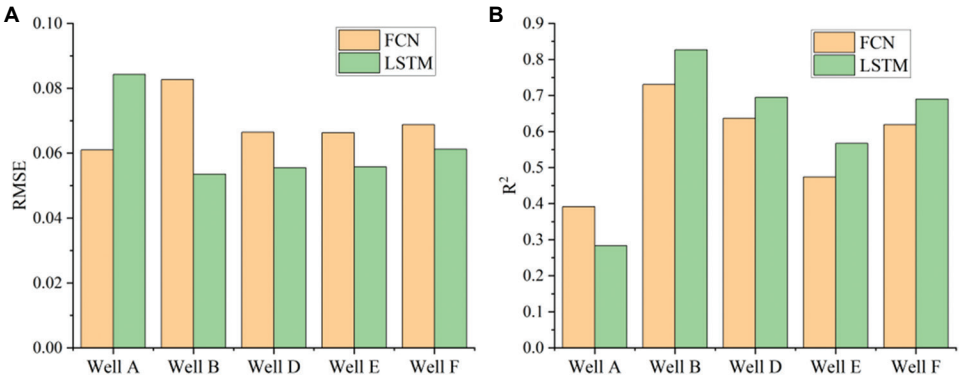


Figure 14. Prediction accuracy in term of root mean square error (A) and R^2 (B) with different network models
Abbreviations: FCN: Fully connected neural; LSTM: Long short-term memory

sequence, is a crucial parameter for data processing. Omitting this step would prevent the LSTM network from learning the influence of historical data on current values. The pore structure and fracture networks of coal seams exhibit similarity within a certain depth range (e.g., coal seams and surrounding rocks), but beyond this range, geological characteristics change significantly. The choice of sequence length is related to the geological variability with depth. A shorter sequence length may overlook the influence of geological layers, while an excessively long sequence length increases training time and may reduce generalization performance due to noise accumulation. Therefore, we analyzed the LSTM network using different sequence length values and evaluated the prediction accuracy for each test well.

As evidenced by the data in Figures 15, 16, and Table 5, a sequence length of 32 yielded the optimal performance in this test, resulting in the lowest average RMSE and the highest average R^2 across all wells. It should also be noted that the choice of sequence length significantly impacts the training duration, requiring a careful balance between sequence length and computational cost (Table 6). If the sequence length is too short, the model may fail to capture sufficient historical information, leading to issues such as underfitting and prediction lag. Conversely, an excessively long sequence

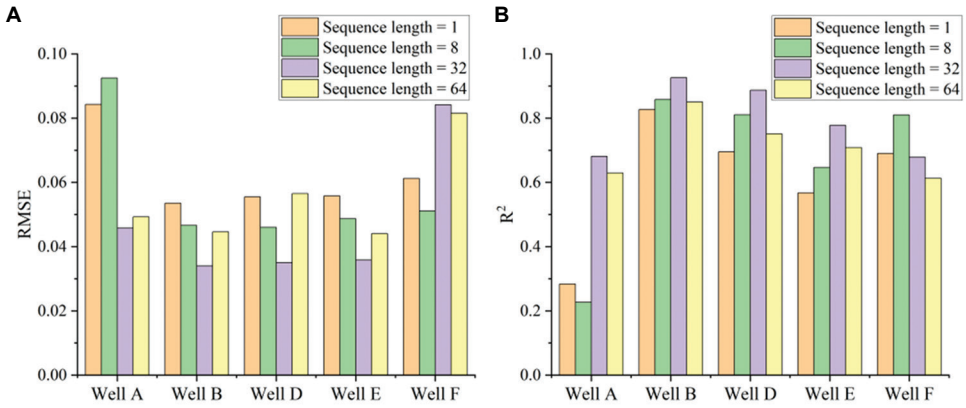


Figure 15. Prediction accuracy in term of root mean square error (A) and R^2 (B) with different sequence length for the long short-term memory network

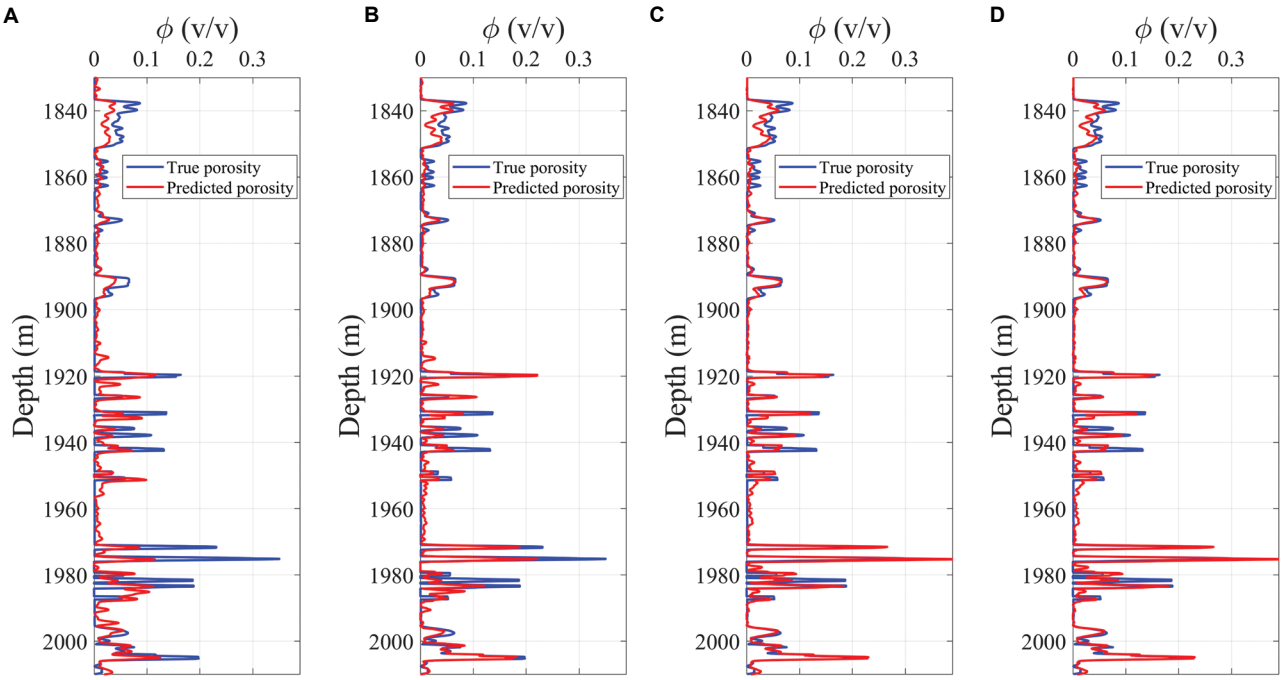


Figure 16. True and predicted porosity comparison using the long short-term memory model with sequence length of 1(A), 8 (B), 32 (C), and 64 (D), for Well D

Table 5. Prediction accuracy with different sequence length for the long short-term memory (LSTM) network

Sequence length	Mean RMSE	Mean R^2
Sequence length=1	0.0621	0.6125
Sequence length=8	0.0570	0.6704
Sequence length=32	0.0470	0.7901
Sequence length=64	0.0552	0.7105

Abbreviation: RMSE: Root mean square error.

length, while theoretically capable of incorporating richer contextual information, can cause an expansion in input dimensions and prolong the gradient backpropagation path through the LSTM hidden states. This not only substantially increases GPU memory usage and training time per iteration but may also degrade generalization performance due to accumulated noise. To reduce time costs and improve engineering feasibility, distributed training with multi-GPU acceleration can be used, or the sequence length and input dimensions can be reduced to shorten training time.

3.4. Application

To validate the effectiveness of the aforementioned method, Well C within the study area was designated as the test well,

while the remaining wells were used for training. Utilizing Combination 5 (i.e., the six parameters V_p, V_s, ρ, V_p/V_s, Z_p, and Z_s as input features), the models were trained for 100 epochs. The remaining hyperparameters were kept at their default and identical values, and the sequence length for the LSTM network was set to 32. Both the FCN and LSTM models were trained under these conditions to generate and compare their prediction results. Figures 17A and B demonstrate the superior prediction performance of the LSTM network. As indicated in Table 7, the LSTM model achieves a reduction in RMSE of approximately 43.41% and an improvement in R^2 to 0.7972 compared to the baseline.

Table 6. Comparison of training time for different sequence lengths

Model	Sequence length	Training time in seconds (100 epochs)
FCN	/	35.8
LSTM	Sequence length=1	44.2
	Sequence length=8	82.3
	Sequence length=32	246.9
	Sequence length=64	506.0

Abbreviations: FCN: Fully connected neural; LSTM: Long short-term memory.

Table 7. Prediction accuracy with different network models for Well C

Network model	Mean RMSE	Mean R^2
FCN	0.0546	0.4104
LSTM	0.0309	0.7972
Bi-LSTM	0.0279	0.8342

Abbreviations: Bi-LSTM: Bidirectional Long Short-Term Memory; FCN: Fully connected neural; LSTM: Long Short-Term Memory; RMSE: Root mean square error.

Furthermore, we employed a Bi-LSTM model to perform the prediction, while keeping all other parameters unchanged. The corresponding results are shown in Figure 17C. The Bi-LSTM model achieved an RMSE of 0.0279, representing a further reduction of 9.71% compared to the standard LSTM, and an R^2 of 0.8342, corresponding to an increase of 0.0370. In Figure 17C, it can be observed that the areas with high porosities accord with the interpreted coals at the depths around 1980 m, 2040 m, and 2080 m, which indicates the prediction could help identify coal sections in good accuracy. These results indicate that the Bi-LSTM model outperforms the standard LSTM both in terms of prediction accuracy and robustness, demonstrating its effectiveness for the task of petrophysical parameter inversion.

4. Discussion

The findings of this study indicate the potential of deep learning, especially sequence models such as LSTM and Bi-LSTM, in addressing the complex challenge of porosity prediction in deep CBM reservoirs. LSTM network inherently captures the contextual dependencies and long-

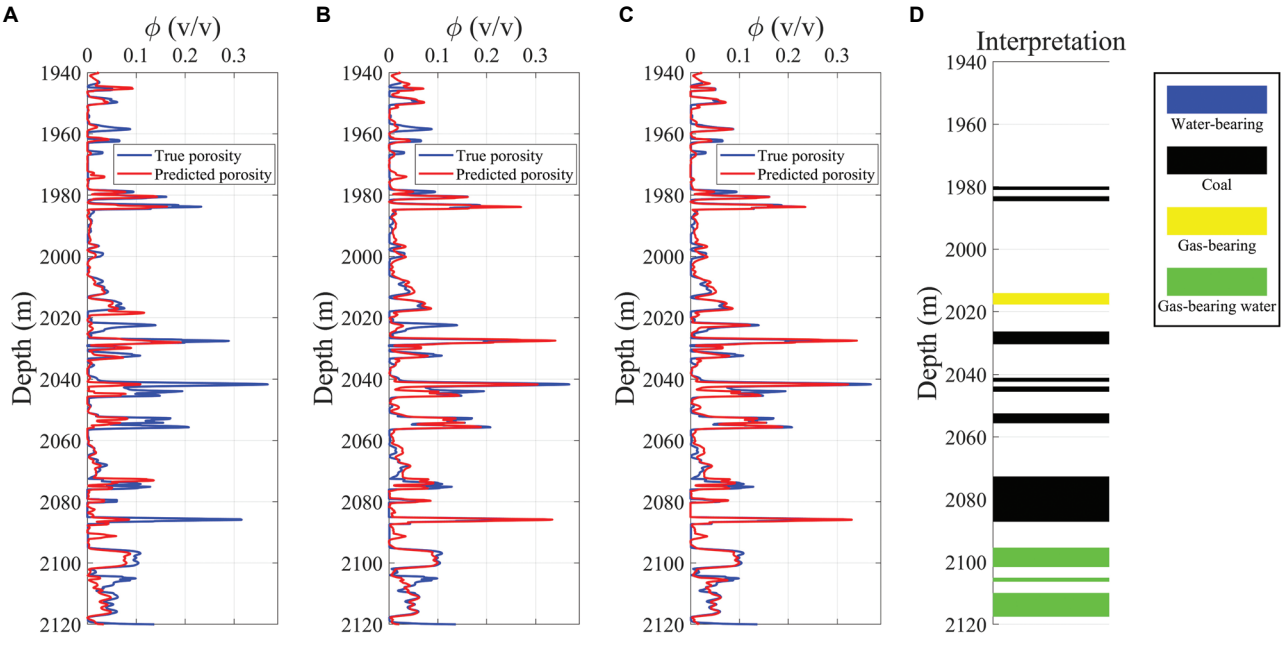


Figure 17. Prediction accuracy with FCN (A), LSTM (B), and Bi-LSTM (C), and the lithofacies interpretation (D) for Well C
Abbreviations: Bi-LSTM: Bidirectional long short-term memory; FCN: Fully connected neural; LSTM: Long short-term memory

range trends within the log curves. This capability is crucial for petrophysical prediction, as reservoir parameters at a given depth are often geologically influenced by the overlying and underlying formations. The further improvement by the Bi-LSTM model indicates that integrating information from both shallower and deeper sections leads to a more accurate prediction. Due to its bidirectional learning capability, Bi-LSTM is theoretically applicable to other sequence data-driven reservoir prediction tasks. For example, shale gas and carbonate reservoirs also have complex pore structures and non-linear relationships. With appropriate feature selection and parameter tuning, the Bi-LSTM model can be applied to porosity prediction in these reservoirs. However, it should be noted that the present study only analyzes the FCN and LSTM models, while a more comprehensive comparison with other advanced networks, such as Temporal Convolutional Networks or Transformer-based models, was not conducted. Future work should include such comparisons to more fairly evaluate the performance of Bi-LSTM model.

Overfitting remains a critical challenge in deep learning. This study employed the dropout method to mitigate overfitting by randomly dropping some neurons during training, thereby reducing the model's reliance on training data. In future research, L2 regularization will be introduced to further constrain model complexity and reduce overfitting by adding the L2 norm of weights to the loss function.

The inherent difficulties in predicting porosity in CBM reservoirs extend beyond the selection of an appropriate algorithm. The complexities of coal seams also present a fundamental task. Coal has a unique dual-porosity system, including the cleat/fracture network and the matrix pores, which governs the storage and transport mechanisms of methane. Porosity measurements and log responses are generally affected by this complex pore structure and the presence of adsorbed gas. Such inherent pore complexities are significant factors influencing the non-linear and challenging nature of the porosity prediction. However, although our data-driven model constructs the relationship between elastic parameters and porosity, it does not explicitly explain or analyze the influence of those pore complexities. A deeper investigation into how these dual-porosity characteristics manifest in the seismic elastic parameters represents a critical area for further research.

Errors may arise from model limitations and the complexity of geological characteristics. For example, the complex and variable pore structure and fracture networks in coal seams result in a highly non-linear relationship between porosity and elastic parameters. Although the Bi-LSTM model performs well in capturing such non-linear relationships, prediction errors may still occur in certain depth intervals. Future research could reduce errors by introducing more complex model structures or increasing the amount of training data.

The proposed method is primarily a data-driven approach. It takes advantage of the powerful non-linear mapping capabilities of deep learning to establish a relationship between input features and the target output, without explicitly considering the governing physical laws. It may limit the model generalizability and physical interpretability in practical applications. To address this issue, our further research will focus on developing a physics-guided deep learning model. In particular, rock physics models can provide the physical relationship between porosity and elastic parameters, offering prior knowledge for deep learning models. By incorporating a coal-specific rock physics model into the network or loss function, we aim to constrain the predictions to be not only data-consistent but also physically plausible for different CBM fields. In addition, rock physics models can supplement labeled data, compensating for the problem of overfitting of deep learning models in small sample scenarios and alleviating the impact of insufficient data.

5. Conclusion

This work proposes a deep learning-based method for predicting porosity in deep CBM reservoirs with well log data. The study investigates the input features of seismic elastic parameters for training, which leads to the optimal combination of P-wave velocity, S-wave velocity, density, and impedance for predicting porosity. The study also focuses on the analysis of network parameters such as sampling interval and sequence length, to achieve an optimal balance between prediction accuracy and computational efficiency. Tests and comparisons indicate that the LSTM network demonstrates a reduction in RMSE of approximately 43.41% and an improvement in R^2 from 0.4104 to 0.7972 compared to the FCN network. Furthermore, the proposed Bi-LSTM model not only enhances bidirectional contextual awareness but also significantly improves generalization capability. Compared to the standard LSTM, it achieved a further RMSE reduction of approximately 9.71% and increased the R^2 to 0.8342. The predictions by the Bi-LSTM model exhibit good capability in identifying potential coal layers. The proposed method provides a reliable approach for porosity prediction with well log data, which could effectively assist in seismic exploration for deep CBM reservoirs.

Acknowledgments

None.

Funding

We appreciate the support provided by the National Key Research and Development Program of China (2024YFC3015802), the National Natural Science Foundation of China (42574178 and 42374128) and the Jiangsu Provincial Science and Technology Plan Project (Natural Science Foundation of Jiangsu Province, BK20252046).

Conflict of interest

Qiang Guo and Jing Ba are Editorial Board Members of this journal but were not in any way involved in the editorial and peer-review process conducted for this paper, directly or indirectly. The authors declare they have no competing interests.

Author contributions

Conceptualization: Qiang Guo
Formal analysis: Xinyu Zhao
Investigation: Cong Luo
Methodology: Jing Ba
Writing–original draft: Qiang Guo, Xinyu Zhao
Writing–review & editing: Qiang Guo, Xinyu Zhao

Availability of data

Data are available from the corresponding author on reasonable request.

References

1. Pan J, Ge T, Liu W, *et al.* Organic matter provenance and accumulation of transitional facies coal and mudstone in Yangquan, China: Insights from petrology and geochemistry. *J Nat Gas Sci Eng*. 2021;94:104076.

   doi: 10.1016/j.jngse.2021.104076

2. Hou X, Liu S, Zhu Y, Yang Y. Evaluation of gas contents for a multi-seam deep coalbed methane reservoir and their geological controls: *In situ* direct method versus indirect method. *Fuel*. 2020;265:116917.

   doi: 10.1016/j.fuel.2019.116917

3. Gong F, Cheng J, Wang G, Peng S, Zhang Z. The effect of lamination on elastic anisotropy of primary coals under confining pressure: Experiment and theoretical modelling. *Geophys Prospect*. 2025;73(4):1228-1242.

   doi: 10.1111/1365-2478.70009

4. Gong F, Huang A, Kang W, *et al.* The influence of lamination and fracture on the velocity anisotropy of tectonic coals. *Geophysics*. 2024;89(6):MR355-MR365.

   doi: 10.1190/GEO2024-0033.1

5. Khaksar A, Griffiths CM. Porosity from sonic log in gas-bearing shale sandstones: Field data versus empirical equations. *Explor Geophys*. 1998;29(4):440-446.

   doi: 10.1071/EG998440

6. Liu X, Shao G, Yuan C, Chen X, Li J, Chen Y. Mixture of relevance vector regression experts for reservoir properties prediction. *J Petrol Sci Eng*. 2022;214:110498.

   doi: 10.1016/j.petrol.2022.110498

7. Wang P, Chen X, Wang B, Li J, Dai H. An improved method for lithology identification based on a hidden Markov model and random forests. *Geophysics*. 2020;85(6):IM27-IM36.

   doi: 10.1190/GEO2020-0108.1

8. Sang K, Yin X, Zhang F. Machine learning seismic reservoir prediction method based on virtual sample generation. *Petrol Sci*. 2021;18(6):1662-1674.

   doi: 10.1016/j.petsci.2021.09.034

9. Guo Q, Ba J, Luo C. Seismic rock-physics linearized inversion for reservoir- property and pore-type parameters with application to carbonate reservoirs. *Geoenergy Sci Eng*. 2023;224:211640.

   doi: 10.1016/j.geoen.2023.211640

10. Luo C, Ba J, Guo Q. Probabilistic seismic petrophysical inversion with statistical double-porosity Biot-Rayleigh model. *Geophysics*. 2023;88(3):M157-M171.

    doi: 10.1190/GEO2022-0288.1

11. Wang P, Cui Y, Zhou L. Multi-task learning for seismic elastic parameter inversion with the lateral constraint of angle-gather difference. *Petrol Sci*. 2024;21(6):4001-4009.

    doi: 10.1016/j.petsci.2024.06.010

12. Zhao L, Nasser M, Han D. Quantitative geophysical pore-type characterization and its geological implication in carbonate reservoirs. *Geophys Prospect*. 2013;61:827-841.

    doi: 10.1111/1365-2478.12043

13. Song L, Yin X, Zong Z, Jiang M. Semi-supervised learning seismic inversion based on Spatio-temporal sequence residual modeling neural network. *J Petrol Sci Eng*. 2022;208:109549.

    doi: 10.1016/j.petrol.2021.109549

14. Wu X, Jiang G, Wang X, *et al.* Prediction of reservoir sensitivity using RBF neural network with trainable radial basis function. *Neural Comput Appl*. 2013;22:947-953.

    doi: 10.1007/s00521-011-0787-z

15. Ahmadi MA, Ebadi M, Shokrollahi A, Majidi SMJ. Evolving artificial neural network and imperialist competitive algorithm for prediction oil flow rate of the reservoir. *Appl Soft Comput*. 2013;13(2):1082-1098.

doi: 10.1016/j.asoc.2012.10.009

16. Zerrouki AA, Aifa T, Baddari K. Prediction of natural fracture porosity from well log data by means of fuzzy ranking and an artificial neural network in Hassi Messaoud oil field, Algeria. *J Petrol Sci Eng*. 2014;115:78-89.

doi: 10.1016/j.petrol.2014.01.011

17. Cao J, Yang J, Wang Y, Wang D, Shi Y. Extreme learning machine for reservoir parameter estimation in heterogeneous sandstone reservoir. *Math Probl Eng*. 2015;2015:287816.

doi: 10.1155/2015/287816

18. Zou C, Zhao L, Xu M, Chen Y, Geng J. Porosity prediction with uncertainty quantification from multiple seismic attributes using Random Forest. *J Geophys Res Solid Earth*. 2021;126:e2021JB021826.

doi: 10.1029/2021JB021826

19. Elkatatny S, Mahmoud M, Tariq Z. New insights into the prediction of heterogeneous carbonate reservoir permeability from well logs using artificial intelligence network. *Neural Comput Appl*. 2018;30(9):2673-2683.

doi: 10.1007/s00521-017-2850-x

20. Wu X, Shi Y, Fomel S. FaultNet3D: Predicting fault probabilities, strikes, and dips with a single convolutional neural network. *IEEE Trans Geosci Remote Sens*. 2019;57(11):9138-9155.

doi: 10.1109/TGRS.2019.2925003

21. Wang P, Xu H, Peng Z, Wang Z, Yang M. Application of data augmentation based on generative adversarial network in impedance inversion. *J Seismic Explor*. 2023;32(2):155-168.

22. Behnia AMO, Reza M, Ali M. A new approach for seismic inversion with GAN algorithm. *J Seismic Explor*. 2024;33(3):1-36.

23. Suraj P, Omer S, Aditya N, *et al.* Model fusion with physics-guided machine learning: Projection-based reduced-order modeling. *Phys Fluids*. 2021;33(6):067123.

doi: 10.1063/5.0053349

24. Xu M, Zhao L, Gao S, Zhu X, Geng J. Joint use of multiseismic information for lithofacies prediction via supervised convolutional neural networks. *Geophysics*. 2022;87(5): 151-162.

doi: 10.1190/GEO2021-0554.1

25. Gao S, Xu M, Zhao L, Chen Y, Geng J. Seismic predictions of fluids via supervised deep learning: Incorporating various class-rebalance strategies. *Geophysics*. 2023;88(4):185-200.

doi: 10.1190/GEO2022-0363.1

26. Yu S, Ma J. Deep learning for geophysics: Current and future trends. *Rev Geophys*. 2021;59:e2021RG000742.

doi: 10.1029/2021RG000742

27. Wang Y, Niu L, Zhao L, *et al.* Gaussian mixture model deep neural network and its application in porosity prediction of deep carbonate reservoir. *Geophysics*. 2022;87(2):59-72.

doi: 10.1190/GEO2020-0740.1

28. Wu H, Wu R, Zhang P, Huang Y, Huang Y, Dong S. Combined fluid factor and brittleness index inversion for coal-measure gas reservoirs. *Geophys Prospect*. 2022;70:751-764.

doi: 10.1111/1365-2478.13172

29. Liu J, Zhao L, Xu M, Zhao X, You Y, Geng J. Porosity prediction from prestack seismic data via deep learning: Incorporating a low-frequency porosity model. *J Geophys Eng*. 2023;20(5):1016-1029.

doi: 10.1093/jge/gxad063

30. Zhang J, Liu Z, Zhou Y, Ai H, Han H. Joint inversion method of rock physics based on hunger games search correction and Bi-LSTM. *IEEE Trans Geosci Remote Sens*. 2024;62:5914310.

31. Sun Y, Pang S, Zhang J, Zhang Y. Porosity prediction through well logging data: A combined approach of convolutional neural network and transformer model (CNN-transformer). *Phys Fluids*. 2024;36(2):026604.

doi: 10.1063/5.0190078

32. Tao B, Zhou H, Chen L, Liu B, Wang R, Liu X. Porosity prediction based on stochastic modeling and facies-controlled dataset constrained by seismic attribute. *IEEE Geosci Remote Sens Lett*. 2025;22:1-5.

doi: 10.1109/LGRS.2025.3580778

33. Ashraf M, Robles WRQ, Kim M, Ko YS, Yi MY. A loss-based patch label denoising method for improving whole-slide image analysis using a convolutional neural network. *Sci Rep.* 2022;12:1392.

doi: 10.1038/s41598-022-05001-8

34. Gao Z, Li C, Yang T, Pan Z, Gao J, Xu Z. OMMDE-Net: A deep learning-based global optimization method for seismic inversion. *IEEE Geosci Remote Sens Lett*. 2021;18:208-212.

doi: 10.1109/LGRS.2020.2973266

35. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9:1735-1780.

doi: 10.1162/neco.1997.9.8.1735

ACCSCIENCE PUBLISHING

Journal of Seismic Exploration

ARTICLE

FaciesGAN: A conditional GAN framework for realistic facies scenario generation as an efficient alternative to multiple-point statistics

Yineth Viviana Camacho-De Angulo[1]*, Tiago Mazzutti[2], Bruno B. Rodrigues[3], and Mauro Roisenberg[1]

[1]Department of Informatics and Statistics, Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, Brazil

[2]Instituto Federal Catarinense, Campus Camboriú, Santa Catarina, Brazil

[3]PETROBRAS, Rio de Janeiro, Rio de Janeiro, Brazil

Abstract

Facies are rock bodies that reflect specific depositional environments and play a central role in reservoir characterization. Accurate facies modeling is a key challenge in generating realistic geological scenarios that honor sparse well data while capturing geological uncertainty. This study introduces FaciesGAN, a novel deep learning framework based on conditional generative adversarial networks (cGANs). The method employs a hierarchical structure of generators and discriminators that progressively refine coarse estimates into high-resolution facies models, ensuring consistency with well data and depositional patterns at each stage. FaciesGAN was validated using the limited Stanford Earth Science Data dataset, demonstrating strong performance even under data scarcity. The quantitative evaluation employed multidimensional scaling and yielded an intersection over union index of 99.96% relative to the conditioning well data. These results confirmed the model's ability to generate diverse scenarios with high fidelity while preserving statistical distributions. Compared with a traditional multiple-point statistics implementation, FaciesGAN produced more realistic and varied geological realizations with significantly greater computational efficiency. These results indicate that cGAN-based approaches, such as FaciesGAN, represent a promising direction for subsurface modeling, offering robust tools for data augmentation, improved uncertainty assessment, and enhanced reservoir characterization.

Keywords: Conditional generative adversarial network; Facies; Hard data; Geostatistical simulations; Seismic inversion

***Corresponding author:**
Yineth Viviana Camacho-De Angulo (y.camacho@posgrad.ufsc.br)

Citation: Angulo YVC, Mazzutti T, Rodrigues BB, Roisenberg M. FaciesGAN: A conditional GAN framework for realistic facies scenario generation as an efficient alternative to multiple-point statistics. *J Seismic Explor*. 2025;34(6):45-59. doi: 10.36922/JSE025370069

Received: September 9, 2025

Revised: October 29, 2025

Accepted: November 3, 2025

Published online: November 28, 2025

1. Introduction

In the context of reservoir characterization, facies are defined as rock units with specific attributes that reflect the depositional environment and directly influence the petrophysical properties and heterogeneity of the reservoir.[1,2] Facies are essential for understanding depositional environments, as they enable geoscientists to correlate these units with seismic and well data, thereby playing a crucial role in the seismic inversion process.[3,4] For example, since sandy facies generally exhibit higher porosity than shale

facies, they help identify and distinguish productive from non-productive zones.[5] From this perspective, the integration of facies into seismic inversion algorithms provides greater consistency between the models and the petrophysical properties obtained, generating more robust and reliable models.[6-8]

The generation of multiple facies scenarios that reproduce complex heterogeneous structures plays a key role in the characterization and modeling of geological reservoirs and the stochastic seismic inversion workflow.[9,10] By enabling multiple plausible realizations of the subsurface, facies scenario modeling provides a rigorous framework to explicitly capture and quantify geological uncertainty.[11] This approach reduces biases arising from single deterministic interpretations and ensures that the resulting reservoir models remain consistent with both geological knowledge and observed field data.[2,12] This ability is particularly critical in seismic inversion workflows, where the relationship between well log measurements, core analysis, and seismic responses must be established in a consistent and geologically meaningful way.[13,14]

Facies scenario generation can be carried out using classical and modern methodologies that combine geology, statistics, and artificial intelligence.[10,15] Classical methodologies include techniques used in geology and geostatistics. For example, sequential indicator simulation (SIS) uses binary indicators for each facies, generating scenarios conditioned on available data.[16] SIS is useful for modeling facies, but has several limitations. It often produces loosely connected patterns and oversimplified geological structures,[17] making it difficult to represent features such as channels or faults. SIS is sensitive to variogram fitting, complicating its use with sparse data. It allows the quantification of uncertainty; however, if not accurately calibrated, it can result in geologically inconsistent models.[18]

Modern methodologies include techniques that have revolutionized geological modeling by allowing the representation of complex patterns and advanced spatial relationships. Multiple-point statistics (MPS) represent a significant advance in this area, enabling the capture of spatial patterns in geological data and modeling of multi-location relationships.[19,20] These techniques are especially useful for simulating facies distributions in regions with limited information. They adhere to spatial distributions observed in training data, such as geological maps and previous simulations.[13,21] MPS may face difficulties in constructing representative training images, as it relies on the analyst's expertise. Furthermore, conditioning to real data may be complex to implement without breaking the continuity of the simulated patterns.[13,22]

In this context, generative adversarial networks (GANs) emerge as an innovative methodology for the generation of facies scenarios. GANs offer significant advantages over traditional geostatistical methods and MPS-based simulation. They can learn directly from real data, preserving first-order statistical features (facies proportions) and second-order statistical features (spatial continuity and body geometry).[23,24] GANs are capable of capturing complex spatial patterns and facies relationships, thereby producing more realistic realizations and reducing the subjectivity inherent in model design.[25-27] Moreover, they open the possibility of training networks as a complement to stochastic facies simulation.[28,29]

There are two competing networks in GANs: A generator network creates synthetic data, and a discriminator network assesses the authenticity of the generated data in relation to the real data.[22,25] In the context of facies, GANs can be used to generate new synthetic records that preserve the statistical characteristics of real data, for example, facies distributions and the geophysical properties of wells. On the other hand, conditional GANs (cGANs) include a conditional layer in the data generation process. This conditional layer allows the generation of synthetic data based on specific previous information, such as the type of facies in a particular depth range, providing even more control over the generation process.[30,31] This characteristic allows the assessment of large-scale geological scenarios and the validation of hypotheses about reservoir connectivity and quality.

In recent years, facies scenario generation has been studied through several case studies, showcasing the effectiveness of advanced technologies. For example, Liu *et al.*[32] proposed an approach for generating 3D subsurface facies map models based on GAN. Miele *et al.*[33] proposed integrating a GAN with spatially-adaptive denormalization (SPADE) to predict realistic facies map patterns while adhering to local probabilities. It combines with geostatistical methods of sequential simulation to model facies-conditioned rock properties. Furthermore, Feng *et al.*[23] proposed a GAN-based method in which the network is trained on facies map images. Research has demonstrated excellent results using facies map data and statistical similarity. However, few studies have incorporated known hard information from the GAN training stage, such as observed facies sequences in wells at specific locations. Specifically, no applications have been published on facies data in 2D vertical sections.

Considering the current progress, this work aims to explore advanced techniques for generating facies scenarios, with a particular focus on cGANs. The objective is to evaluate the effectiveness of this technique on 2D

vertical section facies data conditioned on well data. Using a public dataset with a limited number of samples, this study aims to demonstrate that the proposed method can effectively address one of the main challenges in reservoir characterization: data scarcity. This approach leverages synthetic training models to enhance the integration of well logs, facies distributions, and seismic information, producing scenarios that adhere to geological conditions and maintain statistical and spatial consistency. Accordingly, this study demonstrates that incorporating conditioning information enables cGANs to generate more accurate and robust models for reservoir characterization. cGANs offer an innovative solution to overcome the limitations of traditional techniques, contributing to a more coherent and efficient reservoir modeling.

2. Methodology

This study followed the workflow provided in Figure 1. The methodology comprises several interconnected stages.

2.1. Data collection and preprocessing

The Stanford Earth Science Data dataset was chosen and downloaded from the GitHub repository (https://github.com/SCRFpublic/Stanford-VI-E).[4] The database contains data from oil well logs, with detailed samples of the different facies found in the reservoirs. The facies data are stored in.dat format, facilitating preprocessing and analysis. From the dataset, the available facies classes were floodplain (0), point bar (1), channel (2), boundary (3), and deltaic system (Figure 2A). The dataset primarily represented meandering channel systems, emphasizing facies categories relevant to this study (reservoir and non-reservoir types). The remaining facies were reclassified to simplify the categories into "reservoir" (channel; 1) and "non-reservoir" (floodplain, point bar, and boundary; 0), as shown in Figure 2B.

2.2. Image generation and data labeling

A Python 3.12 environment was configured using image processing and visualization libraries to generate visual representations of the filtered and categorized facies. The tabular data were subsequently converted into images. The 3D Stanford VI reservoir model was employed as training data for the deep-learning workflow. This reference model was defined on a 150×200×200 cell grid, with cell dimensions of 25 m in the horizontal (X and Y) directions and 1 m in the vertical (Z) direction. This resulted in a total physical size of 3,750 m (X-axis) × 5,000 m (Y-axis) × 200 m (Z-axis/depth). The 200 m vertical thickness was composed of three distinct layers (80 m, 40 m, and 80 m). To generate the 2D training images, 200 vertical slices (representing X–Z planes) were extracted, corresponding to one slice for each of the 200 cell positions along the Y-axis. Each slice represented the full horizontal (X-axis) distance of 3,750 m and the top 80 m layer (Layer 1). Subsequently, this physical section of 3,750 m × 80 m was resampled to a 256 × 256-pixel matrix. This process resulted in final images with a resolution of approximately 14.65 m/pixel in the horizontal direction and 0.31 m/pixel in the vertical (depth) direction.

A total of 200 divisions in 2D vertical slices were generated and extracted from the 3D facies model. These were used as training images, with 256 × 256 pixels, and categorized according to the corresponding facies class (Figure 2C). Each image was annotated with the depth condition and used as an external label to guide the process. The annotations delineating vertical polygons indicated the different facies represented, based on the 2D section facies found in the Stanford Earth Science Data dataset, and were used as conditioning data for the cGAN.

2.3. Facies scenario generation with the proposed cGAN

The proposed cGAN, termed FaciesGAN, features a multistage architecture designed to generate geologically realistic facies realizations conditioned on well data.

The FaciesGAN model is an adaptation of SinGAN[28] and WGAN-GP.[34] SinGAN is a generative model that can learn from a single natural image.[28] It consists of a pyramid of fully convolutional GANs, each modeling the distribution of image patches at a distinct spatial scale. This allows for generating new samples of arbitrary size and proportion. Although the generated samples exhibit considerable variability, they retain the overall structure and fine textures of the training image.

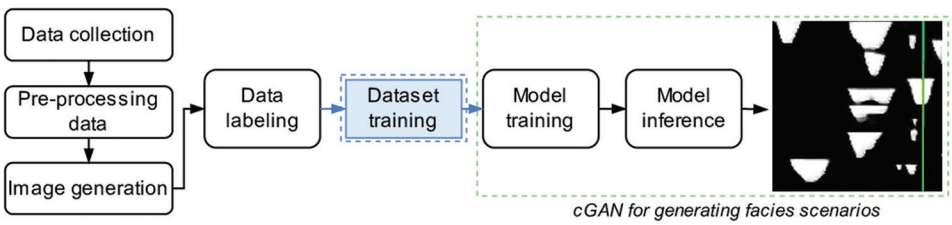


Figure 1. General proposed methodology

The FaciesGAN model is structured as a hierarchy of generators and discriminators operating at progressively higher resolutions, as shown in Figure 3. The process begins with a low-resolution generator that produces

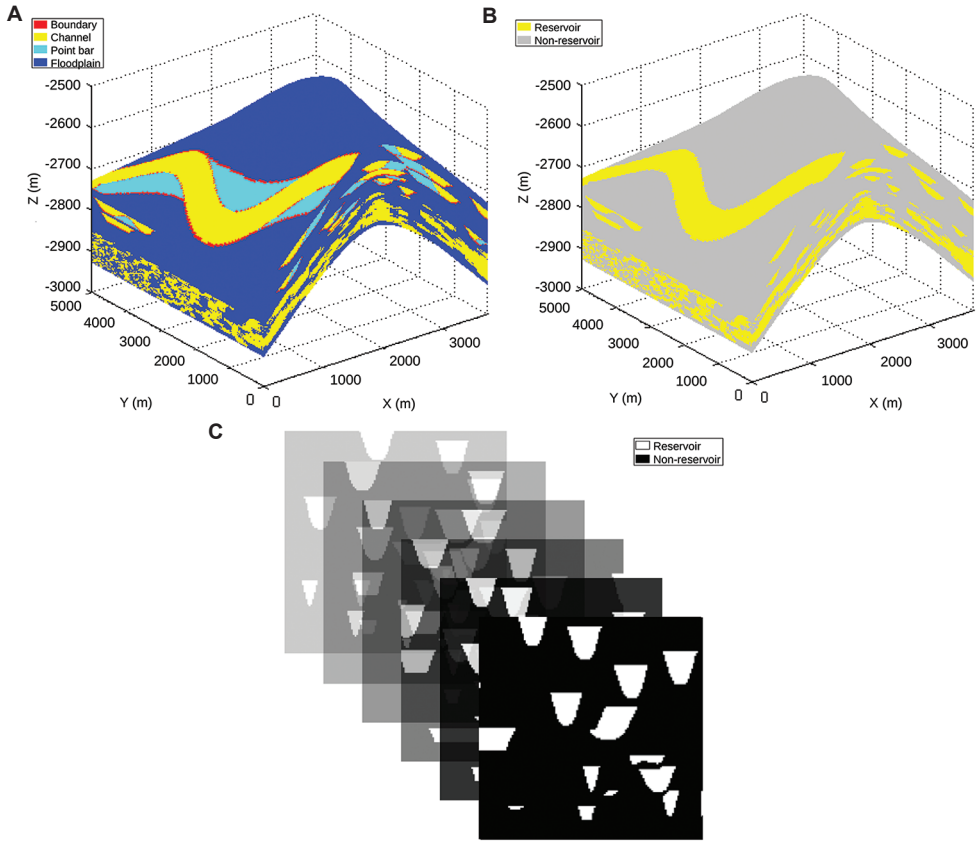


Figure 2. The Stanford Earth Science Data dataset. (A) Multiple sedimentary facies visualization. (B) Facies classification into reservoir (yellow) and non-reservoir (gray). (C) 2D slices examples of projection along the depth, differentiating the reservoir (white) from the non-reservoir (black). Image reproduced and adapted with permission from Lee and Mukerji.[4]

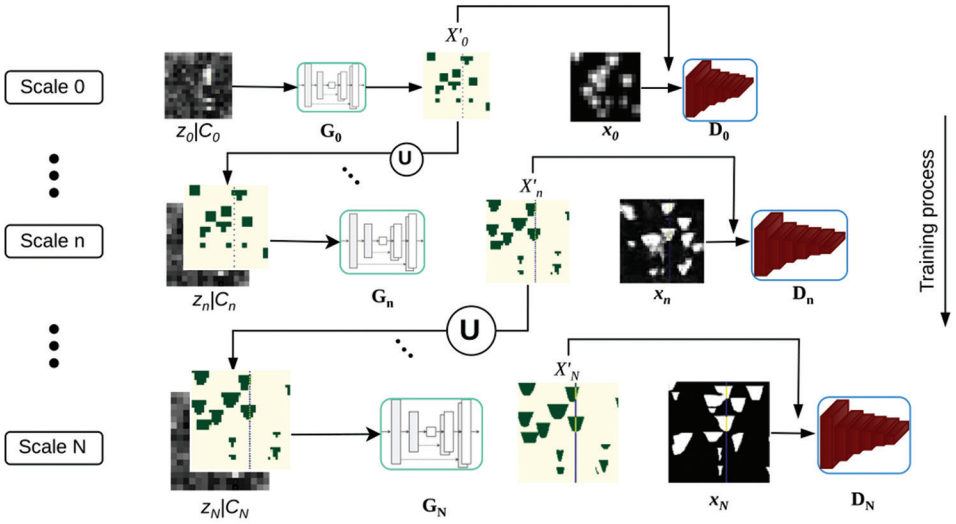


Figure 3. Schematic representation of the proposed cGAN for generating facies scenarios. The generator and discriminator are trained from coarse to fine scale (0 to N).
Abbreviations: cGAN: Conditional generative adversarial network; U: Upsampling.

an initial facies image conditioned on well information (e.g., facies at well locations). Subsequent stages refine this output by adding progressively finer geological details. Each generator stage is paired with a corresponding discriminator that evaluates the realism of the generated facies at its specific resolution while enforcing consistency with the conditioning data. Conditioning is maintained throughout all stages of the generation pipeline, ensuring that the final high-resolution outputs honor well constraints. This progressive refinement strategy allows the model to capture both large-scale geological structures and small-scale heterogeneities, resulting in high-quality, data-consistent facies simulations.

The pyramid of generators $G_0,\ldots,G_n,\ldots,G_N$ is a multi-scale, fully convolutional architecture, as shown in Figure 4A. At each scale, the generator considers a resized version of the previous output x'_{n-1} and a condition z_n, which are concatenated channel-wise. These are then passed through a series of 2D convolutional layers with leaky rectified linear unit (LeakyReLU) to produce a residual output. This is added to the up-sampled input to generate the new 2D section facies map x'_n. Each generator G_n is trained to learn the internal structure of the training images at different scales. G_n finer details from the training images are learned sequentially.[23]

The discriminators, $D_0,\ldots,D_n,\ldots,D_N$, are implemented as a convolutional PatchGAN classifier, which assesses the realism of local image patches rather than making a single global prediction (Figure 4B). The architecture consists of a sequence of convolutional blocks, each comprising a 2D convolutional layer followed by a LeakyReLU activation function.[28] The number of feature channels is progressively reduced across layers (e.g., from 64 to 1), enabling hierarchical feature extraction at multiple spatial resolutions. Notably, normalization layers (e.g., batch normalization) are applied to preserve the raw feature dynamics and stabilize the training process. The final output is a single-channel feature map in which each spatial location corresponds to the discriminator's assessment of whether a specific image patch is real or synthetic.[23,34]

At the n^{th} scale level, an adversarial training process is performed separately: the generator G_n tries to generate fake images x_n to fool the discriminator D_n. The discriminator D_n attempts to distinguish the real images x_n from the fake ones.[23] This multi-scale approach captures the large-scale structures present in the geologic models of interest.[35] The formulation for generating an image sample at the n^{th} level is expressed as follows:

$$x'_n = \begin{cases} G_0(z_0), & n=0 \\ G_n\left(z_n, \bigcup(x'_{n-1})\right), & 0<n\le N \end{cases} \tag{1}$$

where $\cup$ represents the upsampling based on interpolators in the 2D and 3D cases.

The loss function at the n^{th} scale level for G_n and D_n is formulated as:[28]

$$\min_{G_n}\max_{D_n}\mathcal{L}(G_n,D_n)=\mathcal{L}_{adv}(G_n,D_n)+\alpha\mathcal{L}_{rec}(G_n) \tag{2}$$

where $\mathcal{L}_{adv}$ is the adversarial loss for penalizing the distribution distance between the down-sampled images x_n and the generated images x'_n, α is a weighting factor to balance the two loss functions, and $\mathcal{L}_{rec}$ is the reconstruction loss to ensure that x_n can be reproduced given a specific set of random noise maps.

The generator G_n and discriminator D_n at each pyramid scale n are trained with a combined objective inspired by WGAN-GP[34] and SinGAN.[28] The goal is to simultaneously enforce adversarial learning and faithful reconstruction of the image at multiple resolutions.

The discriminator D_n is optimized using the Wasserstein loss with gradient penalty, ensuring Lipschitz continuity. The discriminator loss is formulated as:

$$\mathcal{L}_D^{(n)}=-\mathbb{E}_{x_n\sim p_{data}}\left[D_n(x_n)\right]+\mathbb{E}_{z_n}\left[D_n\left(G_n\left(z_n,\overset{\vee\uparrow}{x}_{n+1}\right)\right)\right]$$
$$+\lambda_{gp}\mathbb{E}_{\hat{x}_n}\left[\left(\nabla_{\hat{x}_n}D_n(\hat{x}_n)_2-1\right)^2\right] \tag{3}$$

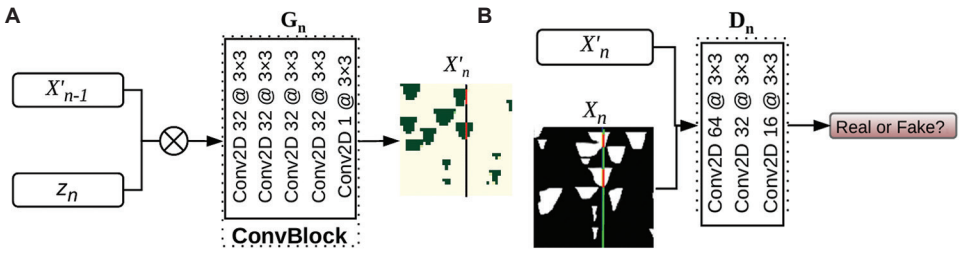


Figure 4. Network architecture at the n^{th} scale level. (A) The generator. (B) The discriminator.

Where x_n denotes a real image at scale n, $G_n\left(z_n, x_{n+1}^{\uparrow}\right)$ is the generated image conditioned on noise z_n and the upsampled output from the next coarser scale, and $\hat{x}_n = \varepsilon x_n + (1-\varepsilon)x'_n$ with $\varepsilon \sim u[0,1]$ is the interpolated sample used for the gradient penalty.

The generator G_n is trained with two complementary objectives: (i) An adversarial loss that encourages generated samples to be indistinguishable from real ones at scale n, and (ii) a reconstruction loss that ensures faithful reproduction of the reference image when a fixed noise map z_n^* is used. The generator loss is:

$$\mathcal{L}_G^{(n)} = -\mathbb{E}_{z_n}\left[D_n\left(G_n\left(z_n, x_{n+1}^{\uparrow}\right)\right)\right] + \alpha_n \mathcal{L}_{rec}^{(n)} \quad (4)$$

With the reconstruction loss defined as:

$$\mathcal{L}_{rec}^{(n)} = x_n - G_n\left(z_n^*, \hat{x}_{n+1}^{\uparrow}\right)_2^2 \quad (5)$$

Where α_n is a scale-dependent weighting factor balancing adversarial and reconstruction objectives, and $\hat{x}_{n+1}^{\uparrow}$ denotes the generated sample from the scale n+1, upsampled to match the resolution of scale n.

This hierarchical optimization scheme allows the generator to progressively capture global structure at coarse scales and fine details at higher resolutions, while the reconstruction term stabilizes training and preserves fidelity.

2.4. Algorithm and implementation

The FaciesGAN model is an architecture designed to generate geologically consistent facies images from a multi-scale noise pyramid. It uses an improved adversarial training framework. FaciesGAN's training loop comprises two alternating main stages, which involve updating the discriminator and generator parameters. The algorithm incorporates additional mechanisms, such as gradient penalty, reconstruction, and masking losses that contribute to improving training stability and fidelity of the generated images. The algorithm and the core procedure for training the FaciesGAN model at a single resolution scale are presented in Algorithm 1. FaciesGAN core training loop (at a single scale).

The training hyperparameters were determined based on the original WGAN-GP and SinGAN models, with empirical adjustments for our specific application. The gradient penalty weight λ_{gp} (referred to as λ in WGAN-GP) was set to 0.115, a value that we found stabilized training effectively for the facies data (in contrast to the λ =10 used in the original WGAN-GP). The reconstruction weight α_{rec} (referred to as α in **Equation [2]** and α_n in **Equation [4]**) was set to 10, a value commonly used in SinGAN-based models that provided an optimal balance between adherence to geological structure and training stability.

Algorithm 1. FaciesGAN core training loop (at a single scale)

Input:
x_{real} ← Real data
M ← Mask
$x_{rec_{in}}$ ← Reconstruction input

Models:
G ← Generator
D ← Discriminator

Hyperparameters:
λ_{gp},α_{rec}, k_d ← Discriminator steps
k_g ← Generator steps

/*Step 1: Train Discriminator */
for j = 1 **to** k_d **do**
  Sample noise pyramid Z←GETNOISE( )
  Generate fake images x_{fake}←G (Z)//Forward pass-through Generator
  $L_{real} \leftarrow -\mathbb{E}\left[D\left(x_{real}\right)\right]$ //Loss for real data
  $L_{fake} \leftarrow \mathbb{E}\left[D\left(x_{fake}\right)\right]$ //Loss for fake data
  $L_{gp} \leftarrow \lambda_{gp}$ CALCULATEGRADIENTPENALTY (D, x_{real}, x_{fake})
  $L_D \leftarrow L_{real} + L_{fake} + L_{gp}$ //Total Discriminator loss
  Update D's parameters θ_d by ascending the gradient of L_D
end for

/* Step 2: Train Generator */
for j = 1 **to** k_d **do**
  Sample noise pyramid Z←GETNOISE( )
  Generate fake images x_{fake}←G (Z)
  $L_{adv} \leftarrow -\mathbb{E}\left[D\left(x_{fake}\right)\right]$ //Adversarial loss
  Sample reconstruction noise Z_{rec} ← GETNOISE (rec=True)
  $x_{rec} \leftarrow G(Z_{rec}, in\_facie = x_{rec\_in})$ //Reconstruction pass
  $L_{rec} \leftarrow \alpha_{rec}$·MSE ($x_{rec}$, x_{real}) //Reconstruction loss
  $L_{mask} \leftarrow 100 \cdot \alpha_{rec} \cdot MSE\,(x_{fake} \odot M, x_{real} \odot M)$ //Masked loss
  $L_G \leftarrow L_{adv} + L_{rec} + L_{mask}$ //Total Generator Loss
  Update G's parameters θ_g by ascending the gradient of L_G
end for

3. Results

This study evaluated the capability of generative models to generate geological facies scenarios. For this purpose, FaciesGAN was trained and validated through visual inspection and multidimensional scaling (MDS) to determine the consistency and representativeness of the generated scenarios with the original facies. For comparison purposes, the same data were modeled using an MPS method, specifically, the single normal equation simulation (SNESIM).[13] The scenarios generated using

the two methodologies were compared through visual inspection of the spatial continuity of the patterns and the facies proportion histogram. This allowed for a qualitative and quantitative analysis of the representativeness and consistency of the simulated models.

Specifically, FaciesGAN was trained to generate facies scenarios using the Stanford Earth Science dataset. The model was developed with a limited training set of 200 samples. During the inference stage, some samples were analyzed using metrics such as visual inspection, average facies proportion, and MDS to determine the consistency and representativeness of the generated scenarios with the original facies.

For FaciesGAN training, appropriate labels were required for each image. The labeling process is shown in Figure 5. These labels are important because they provide information about the characteristics of each image, allowing the model to learn to generate coherent and realistic images based on specific conditions. In this context, the labels corresponded to hard data derived from a simulated well, representing known subsurface information used to condition the facies generation process. The correctly labeled images were integrated into the dataset and associated specifically with each corresponding image. The model used the labels as conditioning input to generate facies scenarios consistent with the characteristics and structures defined by the labels.

The FaciesGAN model was trained for 100 epochs per scale across 10 scales, with a gradient penalty weight λ_{gp} of 0.1, using the Adam optimizer with a learning rate of $5e^{-5}$ and β of 0.5. The kernel size for 2D filters was 3 × 3, with a stride step of 1 × 1. At the coarser scales, image resolution ranged from 16 to 128 pixels. The model was trained on the complete dataset using a workstation with an Intel i7-8700K CPU (6 cores, 3.7 GHz), an NVIDIA GeForce GTX 1080Ti GPU, and 64 GB of RAM.

3.1. Global model evaluation

In the first test, the model generated 1,000 facies scenarios in 20 s; twenty randomly selected conditioned realizations are shown in Figure 6. It was observed that the facies configuration of the conditioning trace, highlighted in green to simulate a real drilled and analyzed well, was closely reproduced in the images generated by FaciesGAN.

The results are promising considering the limited training set, highlighting the applicability of the proposed approach in characterizing oil reservoirs, where well log and facies data are often scarce, costly to obtain, and subject to privacy restrictions. Nevertheless, the model showed remarkable consistency in reproducing the facies spatial distributions. These findings provide insight into the model's capability to produce images that consistently reflect the expected facies proportions. Visual comparisons with real distributions confirmed that the model captured key features of the input data while generating consistent variations. Furthermore, the time required to generate facies scenarios was short, highlighting the computational efficiency of the proposed approach. The short generation time allows for practical integration into workflows that require multiple simulations.

Next, MDS was applied to quantitatively evaluate the trained model and to compare patterns of spatial variability. MDS is a technique commonly used in data analysis and visualization. It represents high-dimensional data in a lower-dimensional space, usually 2D or 3D, while preserving the relative distances (or dissimilarities) between data points and the potential differences between them.[36] The generated facies overlapped closely with the training images in 2D space, demonstrating excellent similarity, as shown in Figure 7. Regions where blue and red overlap indicate highly agreement between generated and real images, suggesting robust model generalization.

The generated facies (red) effectively covered the space of the real facies (blue), indicating the diversity and quality of the generator. The real facies (blue) were closely surrounded by the generated facies, suggesting that the generator interpolates well within the known domain. This indicates a high degree of spatial consistency.

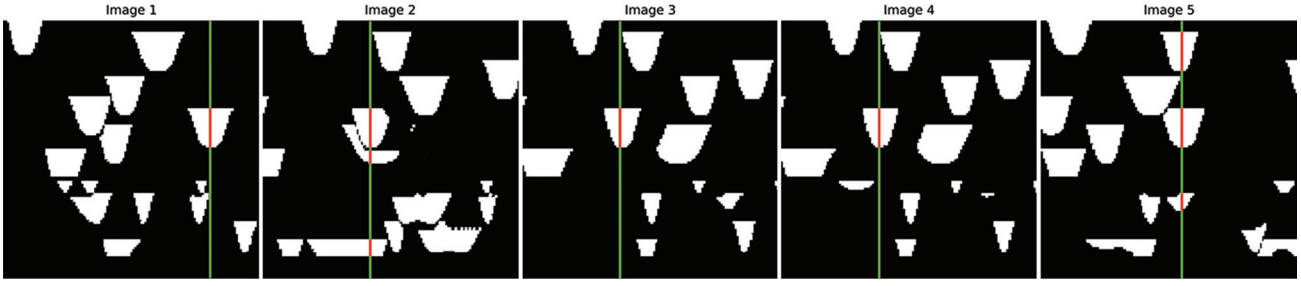


Figure 5. Reservoir (in white) and non-reservoir (in black) with drilled-well conditioning. Conditional traces are highlighted in red (reservoir) and green (non-reservoir). Note: Each image corresponds to a 2D crossline section represented in the pixel domain (256 × 256 pixels) to an 80 m (depth) × 3,750 m (width) vertical section.

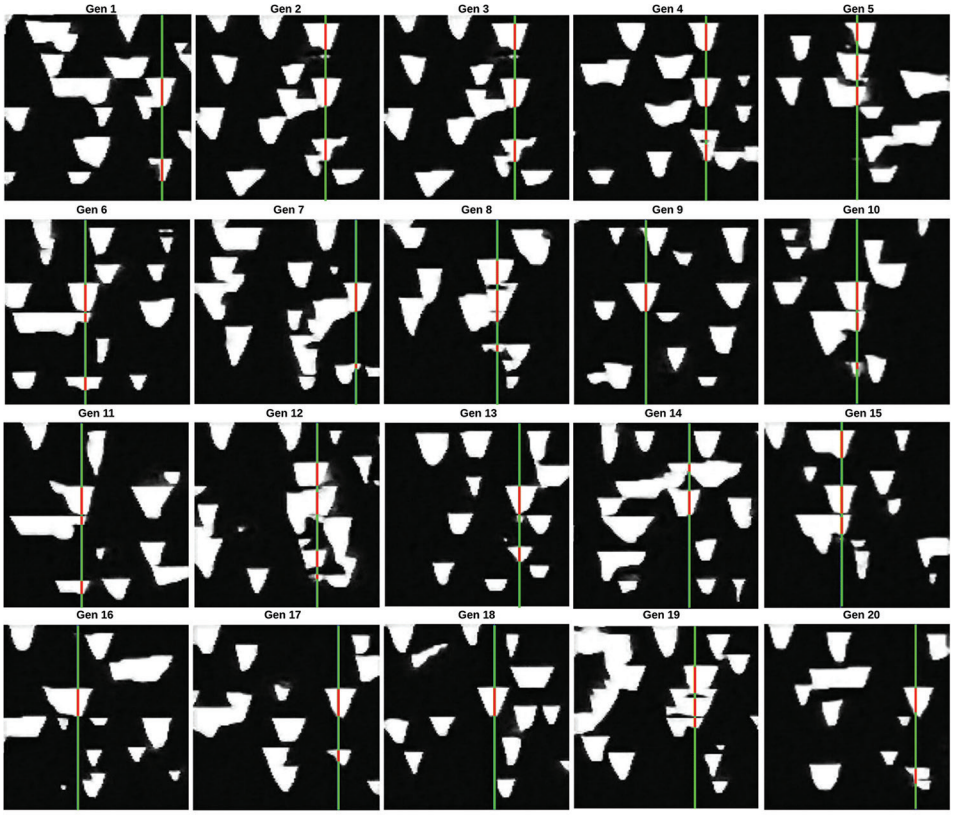


Figure 6. Twenty randomly selected realizations generated by the proposed cGAN. The generated facies are shown in white, while the conditioning is in red (reservoir) and green (non-reservoir). Note: Each image represents a 2D crossline section in the pixel domain (256 × 256 pixels), corresponding to a vertical section 80 m deep and 3,750 m wide.
Abbreviation: cGAN: Conditional generative adversarial network.

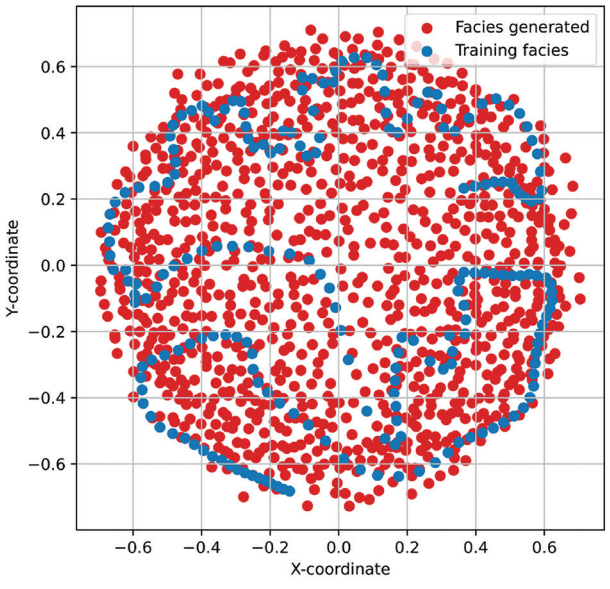


Figure 7. Multidimensional scaling plot of the training images with conditional realizations

In addition, to validate how effectively the conditioning well information was honored by the FaciesGAN model, a procedure was designed that reformulates the problem as a spatial classification task. Specifically, the generated scenarios were compared with their respective original images to evaluate how accurately the location and shape of the facies were reproduced around the actual wells. The intersection over union (IoU) index was used as the evaluation metric. This index is defined as the ratio between the intersection area and the union area of the predicted and reference data:

$$IoU = \frac{GT \cap PD}{GT \cup PD} \tag{6}$$

where PD is the prediction mask and GT is the ground truth. In this evaluation, the prediction mask corresponded to the pixels generated under hard conditioning by FaciesGAN, while the hard-conditioning reference data from the original facies image served as the ground truth. A total of 1,000 images generated from a set of 200 original images were analyzed. The IoU was calculated for each pair of images, yielding an overall mean IoU of 99.96%. This result indicates exceptionally high fidelity in preserving the well-conditioning constraints and demonstrates

that FaciesGAN generates stochastic images that almost perfectly adhere to the geological information observed in the well. These findings validate its effectiveness as a geological conditioning tool.

The effectiveness of the FaciesGAN model was validated by comparing its results with synthetic facies scenarios generated using MPS. Specifically, SNESIM, an improved and scalable extension of the extended normal equation simulation (ENESIM) algorithm for multipoint simulation, was used to generate 1,000 facies realizations based on the 200 training images. The results of 20 representative simulations are presented in Figure 8. These results were generated in approximately 137 min.

The facies scenarios generated by FaciesGAN (Figure 6) exhibited distributions consistent with the expected geology. The scenarios accurately respected the conditions (in green). These realizations reflect the remarkable ability of the model to capture complex spatial patterns with high diversity among simulations. In comparison, the realizations generated by the MPS SNESIM method also preserved the spatial continuity of the facies; however, they exhibited less structural variability than those produced by FaciesGAN. Visual comparison suggests that FaciesGAN accurately reproduced the input conditioning and provided greater structural diversity in its realizations. This demonstrates that the proposed methodology is a robust alternative for generating complex geological scenarios.

Next, we evaluated the overall distribution and class balance within the dataset. The histogram of reservoir facies proportions is shown in Figure 9. In addition to the dataset distribution, the histogram also includes the facies proportion results obtained from the FaciesGAN and MPS SNESIM simulations. This enables a comparative analysis of class balance between the original data and the synthetic realizations produced by the two methods.

The distribution of the dataset (in red) showed a primary peak near 0.16, representing the dominant facies ratio in the real data. The dispersion is moderate, with most realizations concentrated between 0.12 and 0.20. The distribution generated by FaciesGAN (in green) showed a similar behavior, with values concentrated in the same range. However, a slight deviation was observed

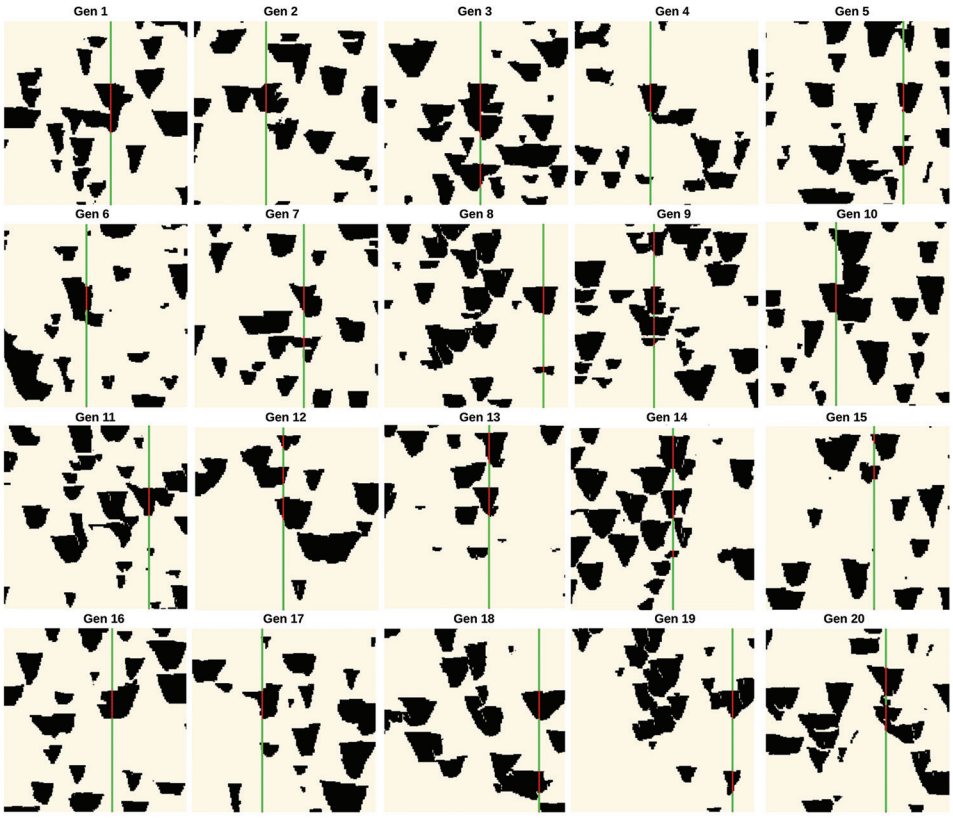


Figure 8. Twenty randomly selected realizations generated by the MPS SNESIM algorithm. The generated facies are shown in black, while the conditioning is in red (reservoir) and green (non-reservoir). Note: Each image represents a 2D crossline section in the pixel domain (256 × 256 pixels), corresponding to a vertical section 80 m deep and 3,750 m wide.
Abbreviations: MPS: Multiple-point statistics; SNESIM: Single normal equation simulation.

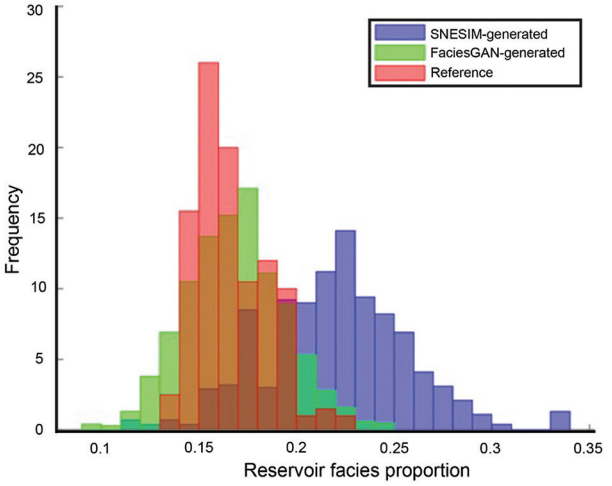


Figure 9. Comparative histogram of reservoir facies proportions generated by FaciesGAN and SNESIM
Abbreviations: GAN: Generative adversarial network; SNESIM: Single normal equation simulation.

toward higher values, indicating a minor overestimation of the proportion in some simulations. In contrast, the distribution generated by SNESIM (in blue) was significantly dispersed, covering a broader range from 0.10 to 0.35. A clear trend toward higher proportions implies lower statistical fidelity compared to the real data. In addition, SNESIM-generated results demonstrated higher variability than those by FaciesGAN and the reference, but were in a controlled interval.

These results indicate that FaciesGAN provides a closer approximation of the observed facies ratios in the reference data compared to SNESIM. The higher variability of SNESIM results in deviations from the true statistical behavior, which can be a major limitation when accurate preservation of facies proportions is required. In addition, the computational efficiency of FaciesGAN is notably superior: while SNESIM took approximately 137 min to generate 1,000 realizations, FaciesGAN produced the same number in only 20 s. This highlights that FaciesGAN has a greater ability to learn and reproduce the distributions observed in real data, enabled by its deep learning-based generative process.

3.2. Well-specific conditioning results

In the second test, five conditioning images distributed in 2D space were selected. For each image, the trained model generated 100 samples at approximately 6 s per image. From each image, five generated facies scenarios are presented in Figure 10. The figure presents the real facies (left column, in green, with the depth condition shown in black) and multiple random model-generated realizations (five columns per well, in black and white) for five different wells, with the depth condition highlighted in green.

The condition incorporated during FaciesGAN training was held constant. The generated facies closely surround or overlap with the real facies, indicating strong spatial consistency between the model realizations and the true data. Although variability was present among generated facies, most realizations maintained structural patterns consistent with the real facies, suggesting that the model adequately learned the underlying spatial patterns.

The real facies were closely surrounded by the generated facies, demonstrating the model's capability to preserve spatial structures across different realizations. This consistency highlights the model's reliability in reproducing subsurface geologic patterns, even under stochastic variability. In such cases, the generation of facies scenarios was fast.

The percentages of pixels corresponding to each facies were compared to evaluate statistical consistency between the real images and those generated by FaciesGAN. The comparison between the percentages observed in the real images and the averages obtained from 100 generated scenarios for the five wells shown in Figure 10 is presented in Table 1. This comparison analyzed the model's ability to reproduce facies distributions realistically, ensuring that the simulations preserve the original geologic characteristics.

A strong correspondence was observed between the facies percentages of the real and generated images. Across all wells, differences between the real values and the generated averages were <4%, indicating that FaciesGAN maintains high fidelity in reproducing facies proportions. For example, in well 181, facies 0 accounted for 72.66% in the real image and 72.71% in the generated average—a practically insignificant difference. Similar cases were observed in the other wells, with the largest deviation occurring in well 63, where facies 1 decreased by approximately 3%. This minor underestimation remains within acceptable ranges for stochastic simulations.

In addition, the IoU metric was calculated for each well to further evaluate the fidelity of facies preservation within the conditioned zones. Table 1 presents the average IoU obtained from each case. The average IoU values ranged from 99.37% to 99.66%, indicating an extremely high agreement between the generated and original facies in the hard-conditioned wells. The results suggest that the FaciesGAN maintains near-identical facies proportions and effectively reproduces the stratigraphic continuity observed in the real data.

The generated scenarios were encouraging, as the model demonstrated a high degree of consistency and realism in reproducing the spatial distribution of facies from a

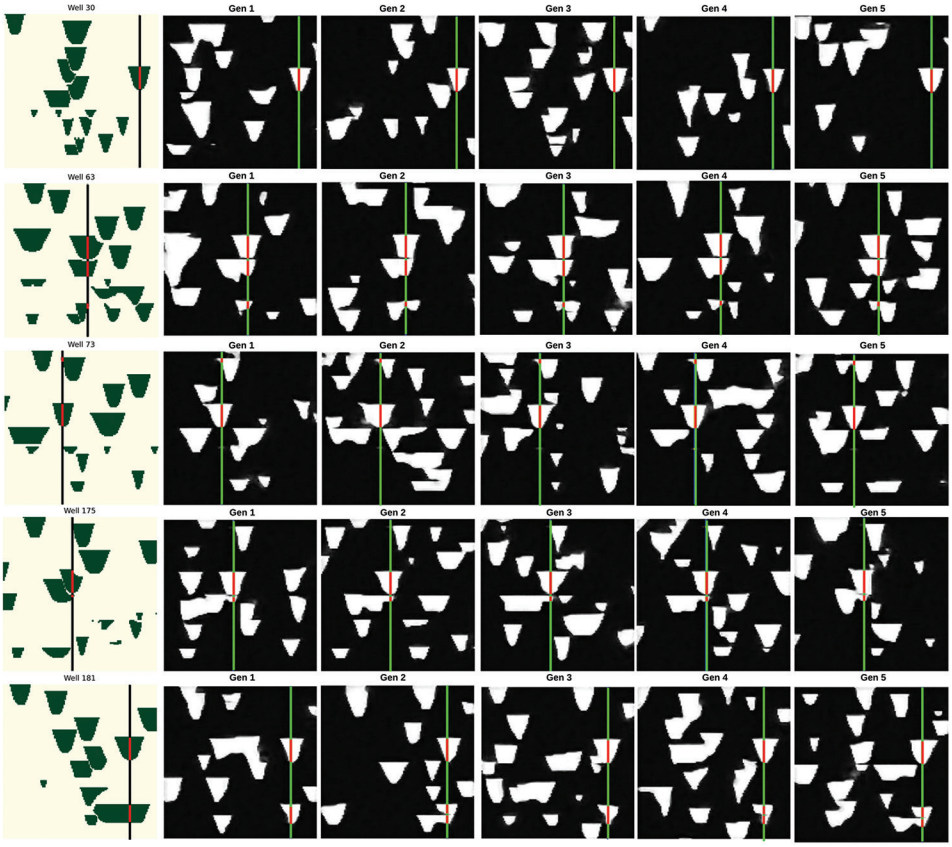


Figure 10. Real and generated facies for five different wells. The first column shows the real 2D facies logs (in green) with the conditioning shown in black. The five columns (Gen 1 to Gen 5) display different random realizations generated by the model, with facies shown in white. The conditioning is highlighted in red (reservoir) and green (non-reservoir).
Note: Each image represents a 2D crossline section in the pixel domain (256 × 256 pixels), corresponding to a vertical section 80 m deep and 3,750 m wide.

Table 1. Averaged facies percentages in real and generated scenarios across five selected wells

Well number	Real images		Generated images		IoU (%)
	Facies 0 (%)	Facies 1 (%)	Facies 0 (%)	Facies 1 (%)	
30	84.77	15.23	85.16	14.84	99.66
63	69.92	30.08	72.35	27.65	99.42
73	79.30	20.70	80.47	19.53	99.38
175	79.30	20.70	81.35	18.65	99.37
181	72.66	27.34	72.71	27.29	99.43

Abbreviation: IoU: Intersection over union.

limited dataset. Furthermore, visual and quantitative comparisons with the actual facies distributions confirmed the model's ability to capture key features of the input data, integrate conditional information, and generate significant variability across realizations. These results suggest that the FaciesGAN model generalizes effectively and serves as a robust tool for generating 2D facies scenarios in petroleum applications, even with limited training datasets. This capability can significantly enhance reservoir characterization and support operations planning.

4. Discussion

The results demonstrate that FaciesGAN produces highly coherent and realistic geological facies scenarios, even when trained on a limited dataset. The generated realizations captured key geological patterns and spatial continuity, showing high fidelity to the conditioning information.

The stochastic simulation methodology employed does not aim to identify a single, "optimal" scenario but rather to quantify geological uncertainty by producing multiple realistic representations of the subsurface that respect the conditioning data. This approach is crucial for risk-based decision-making, as the true facies distribution is unknown. Accordingly, the workflow uses the full ensemble (e.g., as inputs to flow models) to estimate outcome ranges (e.g., P10, P50, and P90) rather than a single "best" result. The statistical consistency of this ensemble with the reference model was evaluated using MDS (Figure 7) and

facies proportion histograms (Figure 9), ensuring that the generated scenarios were statistically representative and effectively explored the geological uncertainty space, as detailed in Section 3.1.

Visual inspection confirmed that FaciesGAN accurately reproduced facies structures and variability, outperforming SNESIM in terms of structural diversity. Quantitative metrics further reinforced this result: the mean IoU (99.96%) indicates near-perfect preservation of well conditioning, and facies-percentage distributions closely aligned with the real data, with deviations generally <4%.

The comparative histogram highlighted that FaciesGAN better approximated the observed facies proportions than SNESIM, which tended to generate realizations with greater variability and statistical deviation. In addition, FaciesGAN demonstrated superior computational efficiency, producing 1,000 scenarios in 20 s compared to SNESIM's 137 min.

These results suggest that deep generative approaches, such as FaciesGAN, offer a robust alternative for simulating geological facies. The model's ability to learn spatial patterns and accurately reproduce conditioning information makes it promising for reservoir modeling workflows that demand both accuracy and efficiency. The results also demonstrate FaciesGAN's potential for generalization, as the model maintained high consistency across different conditioning wells, with stochastic variability remaining within acceptable geostatistical ranges.

The FaciesGAN model was trained and validated exclusively on the Stanford Earth Science Data dataset, which represents a particular meandering-channel system. While the results are promising, especially under conditions of scarce data, further research is necessary to confirm the model's applicability to various geological contexts, including turbidite systems, deltaic environments, and carbonate platforms, which exhibit distinct spatial patterns and heterogeneities.

Furthermore, transfer learning is a promising area for future research. A model pre-trained on a large and diverse set of public geological models could be fine-tuned using smaller, field-specific datasets. This approach could enhance the practicality of FaciesGAN in real-world reservoir characterization projects, where data availability is always limited.

Overall, the tests confirm that FaciesGAN can generate realistic, diverse, and conditionally consistent facies realizations, offering advantages in terms of speed and statistical performance compared with traditional MPS methods. While this study primarily compared FaciesGAN with the conventional MPS SNESIM method, subsequent research should evaluate its performance against additional deep generative models. For example, variational autoencoders could be explored, although cGANs have already demonstrated key advantages in generating scenarios with sharper geological boundaries and closer alignment to the true statistical distribution of reservoir properties[37]—key attributes for realistic facies modeling.

5. Conclusion

The use of artificial intelligence-based techniques to generate facies scenarios is an innovative area aimed at improving the accuracy and robustness of machine learning models in oil exploration and production. cGANs are gaining prominence due to their capacity to generate high-quality synthetic data that preserves known geological characteristics. FaciesGAN was successfully trained with a limited number of 2D facies images, demonstrating strong performance in data-scarce scenarios and offering an effective approach for data augmentation with small datasets. The main advantage of the model lies in its ability to generate synthetic 2D facies scenarios while honoring known conditional information, ensuring consistency with real conditions derived from drilled and analyzed wells. MDS and facies-proportion statistics produced favorable results, highlighting the ability of this network to consistently reproduce the conditioning data. The generated realizations preserve the frequency distributions and spatial correlations characteristic of the original images, ensuring visual consistency and maintaining statistical and geological integrity. Furthermore, a comparative analysis with SNESIM demonstrated that FaciesGAN provides a more accurate representation of facies proportions, with reduced dispersion and skewness relative to the reference data, while maintaining higher spatial and statistical fidelity. Another significant advantage of FaciesGAN is its computational efficiency, which enables the rapid generation of multiple scenarios compared with sequential simulation methods. These synthetic realizations can be integrated into reservoir characterization workflows to support uncertainty estimation and enhance the quality of results. Finally, the proposed methodology can be extended to subsequent workflow steps, such as incorporating facies with acoustic and/or elastic properties, generating synthetic seismic data, and evaluating consistency with actual seismic observations. In particular, future work will focus on applying the approach to real-world field datasets—a logical and most important next step. This extension will enable synthetic seismic generation to be combined with seismic inversion. This integration will

establish a direct and useful link among facies modeling, seismic inversion, and reservoir characterization.

Acknowledgments

None.

Funding

This research was funded by the project "4D Integrated Quantitative Characterization of Reservoirs" (Contract No. FEESC/PETROBRAS/4600671602) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior–Brasil (CAPES)–Finance Code 001.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Yineth Viviana Camacho-De Angulo, Mauro Roisenberg
Data curation: Yineth Viviana Camacho-De Angulo
Formal analysis: Yineth Viviana Camacho-De Angulo, Tiago Mazzutti, Mauro Roisenberg
Funding acquisition: Mauro Roisenberg
Investigation: Yineth Viviana Camacho-De Angulo, Tiago Mazzutti, Mauro Roisenberg
Methodology: Yineth Viviana Camacho-De Angulo, Tiago Mazzutti, Mauro Roisenberg
Project administration: Bruno B. Rodrigues, Mauro Roisenberg
Resources: Bruno B. Rodrigues
Software: Yineth Viviana Camacho-De Angulo, Tiago Mazzutti
Supervision: Mauro Roisenberg
Validation: Yineth Viviana Camacho-De Angulo, Tiago Mazzutti, Mauro Roisenberg
Visualization: Yineth Viviana Camacho-De Angulo
Writing–original draft: Yineth Viviana Camacho-De Angulo
Writing–review & editing: Yineth Viviana Camacho-De Angulo, Tiago Mazzutti, Bruno B. Rodrigues, Mauro Roisenberg

Availability of data

This study employed a publicly available facies dataset from the Stanford VI-E Reservoir: A Synthetic Data Set for Join Seismic-EM Time-lapse Monitoring Algorithms provided by the Stanford Center for Reservoir Forecasting (SCRF), accessible at https://github.com/SCRFpublic/Stanford-VI-E/tree/master.

The code availability for the FaciesGAN model is as follows:

FaciesGAN
Contact: y.camacho@posgrad.ufsc.br
Hardware requirements: No hardware requirement.
Program language: Python
Software required: Python 3.1x
Program size: 62.9 MB

The source codes are available for downloading at the link: https://github.com/mazzutti/FaciesGAN

References

1. Sheriff RE, Geldart LP. *Exploration Seismology*. Cambridge: Cambridge University Press; 1995. Available from: https://books.google.com.br/books?hl=es&lr=&id=wRYgAwAAQBAJ&oi=fnd&pg=PA1901&dq=Sheriff+RE,+Geldart+LP.+Exploration+Seismology.+Cambridge:+Cambridge+University+Press%3B+1995&ots=wCZUKyTUNq&sig=ELLBEg-lVBcYuxFxazP0aboszHY&redir_esc=y#v=onepage&q=Sheriff%20RE%2C%20Geldart%20LP.%20Exploration%20Seismology.%20Cambridge%3A%20Cambridge%20University%20Press%3B%201995&f=false [Last accessed on 2025 Nov 06].
2. Grana D, Mukerji T, Doyen P. *Seismic Reservoir Modeling: Theory, Examples, and Algorithms*. United States: John Wiley and Sons; 2021.
3. Sangree JB, Widmier JM. Interpretation of depositional facies from seismic data. *Geophysics*. 1979;44(2):131-160.
   doi: 10.1190/1.1440957
4. Lee J, Mukerji T. *The Stanford VI-E Reservoir: A Synthetic Data Set for Joint Seismic-EM Time-Lapse Monitoring Algorithms. 25th Annual Report, Stanford Center for Reservoir Forecasting*, Stanford, CA: Stanford University; 2012.
5. Selley RC. *Elements of Petroleum Geology*. United States: Gulf Professional Publishing; 1998.
6. Veeken PCH. *Seismic Stratigraphy, Basin Analysis and Reservoir Characterisation*. Vol. 37. Netherlands: Elsevier; 2006.
7. Grana D, Mukerji T, Dvorkin J, Mavko G. Stochastic inversion of facies from seismic data based on sequential simulations and probability perturbation method. *Geophysics*. 2012;77(4):M53-M72.
   doi: 10.1190/geo2011-0417.1
8. Kemper M, Gunning J. Joint impedance and facies inversion-seismic inversion redefined. *First Break*. 2014;32(9):89-95.
   doi: 10.3997/1365-2397.32.9.77968
9. Pyrcz MJ, Deutsch CV. *Geostatistical Reservoir Modeling*. Oxford: Oxford University Press; 2014.

10. Slatt RM. Geologic controls on reservoir quality. *Dev Petrol Sci*. 2013;61:229-281.

    doi: 10.1016/B978-0-444-56365-1.00006-7

11. Petrovic A, Aigner T, Pontiggia M. Facies heterogeneities in a ramp carbonate reservoir analogue: A new high-resolution approach for 3D facies modelling. *J Petrol Geol*. 2018;41(2):155-174.

    doi: 10.1111/jpg.12698

12. Slatt RM. Basic principles and applications of reservoir characterization. *Dev Petrol Sci*. 2013;61:1-38.

    doi: 10.1016/B978-0-444-56365-1.00001-8

13. Strebelle S. Conditional simulation of complex geological structures using multiple-point statistics. *Math Geol*. 2002;34:1-21.

    doi: 10.1023/A:1014009426274

14. Roksandić MM. Seismic facies analysis concepts. *Geophys Prospect*. 1978;26(2):383-398.

    doi: 10.1111/j.1365-2478.1978.tb01600.x

15. Dembicki H Jr. *Practical Petroleum Geochemistry for Exploration and Production*. Netherlands: Elsevier; 2017.

    doi: 10.1016/C2014-0-03244-3

16. Journel AG. *Fundamentals of Geostatistics in Five Lessons*. Vol. 8. Washington, DC: American Geophysical Union; 1989.

17. Al-Mudhafar WJ. Geostatistical Simulation of Facies and Petrophysical Properties for Hetero-Geneity Modeling in a Tidal Depositional Environment: A Case Study from Upper Shale Member in A Southern Iraqi oil Field. In: *Unconventional Resources Technology Conference, 26-28 July 2021*; 2021. p. 2627-2642.

    doi: 10.15530/urtec-2021-5551

18. Journel AG. Nonparametric estimation of spatial distributions. *J Int Assoc Math Geol*. 1983;15:445-468.

    doi: 10.1007/BF01031292

19. Strebelle SB, Journel AG. Reservoir Modeling using Multiple-Point Statistics. In: *SPE Annual Technical Conference and Exhibition*?; 2001. p. SPE-71324.

    doi: 10.2118/71324-MS

20. Mariethoz G, Caers J. *Multiple-Point Geostatistics: Stochastic Modeling with Training Images. Wiley-Blackwell*; 2014. Available from: https://books.google.com.br/books?hl=es&lr=&id=Fe7sBQAAQBAJ&oi=fnd&pg=PA85&dq=Multiple%E2%80%90Point+Geostatistics:+Stochastic+Modeling+with+Training+Images&ots=P9LR5qUq9B&sig=98MN9Mk5KPtkMgfniK_ECGVnUSQ&redir_esc=y#v=onepage&q=Multiple%E2%80%90Point%20Geostatistics%3A%20Stochastic%20Modeling%20with%20Training%20Images&f=false [Last accessed 2025 Nov 06].

21. Harding A, Strebelle S, Levy M, *et al*. Reservoir facies modelling: New advances in MPS. In: *Geostatistics Banff 2004*. Berlin: Springer; 2004. p. 559-568.

    doi: 10.1007/978-1-4020-3610-1_57

22. Zhang TF, Tilke P, Dupont E, Zhu LC, Liang L, Bailey W. Generating geologically realistic 3D reservoir facies models using deep learning of sedimentary architecture with generative adversarial networks. *Pet Sci*. 2019;16(3):541-549.

    doi: 10.2523/IPTC-19454-MS

23. Feng R, Grana D, Mosegaard K. Geostatistical facies simulation based on training image using generative networks and gradual deformation. *Math Geosci*. 2025;57:1021-44.

    doi: 10.1007/s11004-024-10169-y

24. Tian L, Wang Z, Liu W, Cheng Y, Alsaadi FE, Liu X. A new GAN-based approach to data augmentation and image segmentation for crack detection in thermal imaging tests. *Cognit Comput*. 2021;13(5):1263-1273.

    doi: 10.1007/s12559-021-09922-w

25. Goodfellow IJ, Pouget-Abadie J, Mirza M, *et al*. Generative adversarial nets. *Adversarial Neural Information Processing Systems*; 2014. p. 27.

26. Bowles C, Chen L, Guerrero R, *et al*. *Gan Augmentation: Augmenting Training Data using Generative Adversarial Networks*. [arXiv preprint]; 2018.

    doi: 10.48550/arXiv.1810.10863

27. Patel M, Wang X, Mao S. Data Augmentation with Conditional GAN for Automatic Modulation Classification. In: *Proceedings of the 2nd ACM Workshop on Wireless Security and Machine Learning*; 2020. p. 31-36.

    doi: 10.1145/3395352.3402622

28. Shaham TR, Dekel T, Michaeli T. Singan: Learning a Generative Model from a Single Natural Image. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019. p. 4570-4580.

    doi: 10.48550/arXiv.1905.01164

29. Mosser L, Kimman W, Dramsch J, Purves S, La Fuente Briceño A, Ganssle G. Rapid Seismic Domain Transfer: Seismic Velocity Inversion and Modeling using Deep Generative Neural Networks. In: *80th Eage Conference and Exhibition 2018*. Vol. 2018. p. 1-5.

    doi: 10.3997/2214-4609.201800734

30. Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA. Generative adversarial networks: An overview. *IEEE Signal Process Mag*. 2018;35(1):53-65.

    doi: 10.1109/MSP.2017.2765202

31. Douzas G, Bacao F. Effective data generation for imbalanced learning using conditional generative adversarial networks. *Expert Syst Appl*. 2018;91:464-471.

doi: 10.1016/j.eswa.2017.09.030

32. Liu M, Nivlet P, Smith R, BenHasan N, Grana D. Recurrent neural network for seismic reservoir characterization. In: Bhattacharya S, Di H, editors. *Advances in Subsurface Data Analytics*. Ch. 4. Netherlands: Elsevier; 2022. p. 95-116.

doi: 10.1016/B978-0-12-822295-9.00010-8

33. Miele R, Azevedo L, Grana D. Iterative Stochastic Seismic Inversion Method using GAN with Spatiallyadaptive Normalization. In: *Third EAGE Conference on Seismic Inversion*; 2024. p. 1-5.

doi: 10.3997/2214-4609.202438056

34. Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville AC. Improved training of wasserstein gans. *Advance Neural Information Processing Systems*; 2017. p. 30.

35. Tran TT. Improving variogram reproduction on dense simulation grids. *Comput Geosci*. 1994;20(7-8):1161-1168.

doi: 10.1016/0098-3004(94)90069-8

36. Cox Michael AA, Cox TF. Multidimensional scaling. In: *Handbook of Data Visualization*. Berlin: Springer Berlin Heidelberg; 2008. p. 315-347.

doi: 10.1007/978-3-540-33037-0_14

37. Miele R, Levy S, Linde N, Soares A, Azevedo L. Deep generative networks for multivariate fullstack seismic data inversion using inverse autoregressive flows. *Comput Geosci*. 2024;188:105622.

doi: 10.1016/j.cageo.2024.105622

ACCSCIENCE PUBLISHING

Journal of Seismic Exploration

ARTICLE

Microseismic event picking and classification for hot dry rock hydraulic fracturing monitoring using SeisFormer

Mingjun Ouyang[1], Zenan Leng[2], Haotian Hu[3], Zubin Chen[1], Fa Zhao[1], and Feng Sun[1]*

[1]Key Laboratory of Geo-Exploration Instrumentation of the Ministry of Education, College of Instrumentation and Electrical Engineering, Jilin University, Changchun, Jilin, China

[2]Research and Development Center, Changchun UP Optotech Co., Ltd., Changchun, Jilin, China

[3]Department of Research and Development, Jingwei Hirain Co., Ltd., Changchun, Jilin, China

(This article belongs to the *Special Issue: Advanced Artificial Intelligence Theories and Methods for Seismic Exploration*)

Abstract

Accurate seismic monitoring is vital for the safe operation of enhanced geothermal systems in hot dry rock (HDR) reservoirs; however, robust P- and S-wave classification and precise first-arrival picking remain difficult under low signal-to-noise ratios and complex noise conditions. Hence, in this study, we present SeisFormer, a Transformer-based framework that couples adaptive multi-scale windowing with joint time–frequency analysis. It allocates time–frequency resolution on the fly to overcome the limitations of a fixed-window short-time Fourier transform and slowly extracts varying trends and dominant periodicities from waveform sequences. To stabilize the modeling of long-range dependencies, we introduce regularized pseudoinverse attention, which retains the speedups of low-rank approximations while damping amplification in directions associated with small singular values. We evaluated SeisFormer on a unified, multi-site dataset with data from HDR operations in the Qinghai Gonghe Basin and from an unconventional hydraulic-fracturing field in North China. Compared with baselines (EQTransformer, PhaseNet), it exhibited better performance across real-world data, noise-augmented data with non-stationary composite noise, and overlapping multi-event scenarios. On real-world data, it attained 98.30% classification accuracy, with mean arrival-time errors of 1.42 ms (P) and 2.29 ms (S). Ablations show that each component contributes substantially, indicating robustness for near-real-time monitoring and deployment.

Keywords: Microseismic monitoring; Hot dry rock hydraulic fracturing; Picking and classification; Transformer; Adaptive multi-scale windowing; Time–frequency domain

***Corresponding author:**
Feng Sun
(sunfeng@jlu.edu.cn)

Citation: Ouyang M, Leng Z, Hu H, Chen Z, Zhao F, Sun F. Microseismic event picking and classification for hot dry rock hydraulic fracturing monitoring using SeisFormer. *J Seismic Explor*. 2025;34(6):60-77.
doi: 10.36922/JSE025290036

Received: July 15, 2025

Revised: September 25, 2025

Accepted: October 24, 2025

Published online: December 2, 2025

1. Introduction

As a clean and renewable energy source, geothermal energy offers low carbon emissions, environmental friendliness, operational stability, high efficiency, and abundant resources. Among geothermal resources, hot dry rock (HDR) has attracted significant attention due to its large heat-storage capacity and development potential, and has become a

substantial focus of global exploration.[1] To efficiently extract heat from HDR, hydraulic fracturing is commonly employed to increase reservoir permeability, promote heat flow, and improve energy recovery. Figure 1 illustrates the basic process of heat extraction from HDR:[2] High-pressure fluid is injected to induce rock failure; fractures propagate and release energy, accompanied by microseismic events.

Compared with conventional oil and gas reservoirs, hydraulic fracturing in HDR formations is more complex.[3,4] HDR rocks typically exhibit very low permeability and high strength, and their microseismic signals are broadband with a high-frequency bias. Under high-temperature, high-pressure conditions and elevated injection rates, rock strength is further reduced and microseismicity becomes more complex and heterogeneous in time and space; numerous closely spaced events often occur within short time windows, yielding intricate source distributions and substantially increasing the difficulty of signal processing and interpretation.[5-7] Reliable microseismic monitoring is therefore crucial for assessing fracture-propagation dynamics during HDR stimulation and for providing timely feedback for engineered fracture-extension analysis and field decision-making.[8,9]

In microseismic signal processing, phase identification and first-arrival picking are two core tasks. Traditional methods (e.g., short-time average/long-time average [STA/LTA],[10] and Akaike information criterion[11]) perform well under ideal conditions but are prone to false positives and less effective at low signal-to-noise ratios (SNR) and in complex noise environments, limiting their suitability for HDR field applications.[12-16] Recent advances in deep learning have substantially improved detection and phase picking for seismic and microseismic signals.[17-19] Convolutional and recurrent architectures, such as PhaseNet,[20] PickCapsNet,[21] and a convolutional neural network (CNN) + long short-term memory (LSTM)[22] learn discriminative features but still struggle to model long-range dependencies and cross-scale coupling. Transformer-based models, via self-attention, provide global dependency modeling and have become an important framework for seismic time-series analysis. Representative works include EQTransformer, which jointly models detection and phase picking for regional and teleseismic catalogs; EQCCT, which couples compact CNNs with transformers for efficiency and improves cross-domain robustness via basin-scale transfer learning; SeisT, which uses multitask learning to unify detection, phase classification, and arrival-time estimation; and ICAT-Net, which leverages lightweight attention to balance accuracy and efficiency.[23-28] In mining and engineering scenarios, prior work has also explored handcrafted feature representations and hybrid CNN–transformer classifiers. However, many of these methods target conventional seismic catalogs or relatively stationary noise. In particular, transformer pipelines trained on regional or teleseismic data—characterized by lower event density and more stationary backgrounds—generalize poorly to HDR wavefields featuring overlapping onsets, narrowband harmonics, and low-frequency drift. Moreover, the quadratic cost of full attention and fixed analysis windows can introduce latency and unstable pick times on long streams sampled at 1 kHz with rapid cross-scale variability, motivating a time-frequency-aware architecture with adaptive windowing and a stabilized Nyström attention mechanism.

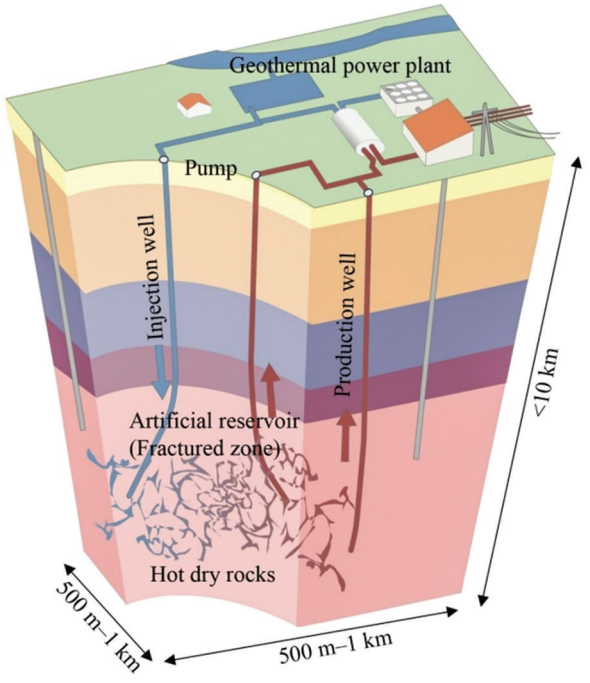


Figure 1. Schematic representation of enhanced geothermal systems using hydraulic fracturing in hot dry rock

In this context, we propose the SeisFormer, a time–frequency transformer framework tailored to HDR hydraulic-fracturing microseismic signals. Our main contributions are as follows:

(i) Time–frequency co-modeling with a transformer. We fuse short-time Fourier transform (STFT)- and two-dimensional (2D)-convolution-derived time–frequency features with self-attention to jointly represent low-frequency trends and high-frequency transients, matching the broadband, high-frequency-biased characteristics of HDR microseismic data.

(ii) Adaptive multi-scale window selection. We adapt window scales based on trend and periodicity cues and employ dynamic grouping and routing for efficient cross-scale modeling in dense-event, strongly non-stationary scenarios.

(iii) Robust and efficient attention. We incorporate a Tikhonov-regularized pseudoinverse into Nyström

attention, delivering low-rank speedups while improving numerical stability, thereby supporting million-sample sequences for engineering deployment.

1.1. Time–frequency characteristics

Microseismic data from HDR hydraulic fracturing in enhanced geothermal system reservoirs are complex, non-stationary time series with distinct time–frequency structure.[29] Waveforms comprise background noise, P waves, S waves, and coda, with phase durations and amplitudes varying across operating conditions. In the frequency domain, P waves have higher-frequency and lower-amplitude, whereas S waves have lower-frequency and higher-amplitude.[30,31] Consequently, models must capture both low- and high-frequency content and adapt across multiple time scales to distribution shifts and transient changes.

1.2. Challenges in manual labeling

P- and S-wave arrival times are commonly picked manually, which is labor-intensive and susceptible to inter- and intra-operator variability. There is a clear need for automated classification and picking methods that are efficient, robust, accurate, and less labor-intensive.

1.3. Multi-event scenarios

Fracture propagation and injection fluctuations often trigger closely spaced, overlapping events. Fixed-window approaches are limited in this regime: Short windows truncate long events, whereas long windows include excessive noise. When signal lengths vary widely and event density is high, processing performance degrades and errors propagate downstream to subsequent modeling stages.

2. Proposed method

2.1. Model architecture

We propose the SeisFormer, a time–frequency modeling framework for P/S classification and first-arrival picking. As illustrated in Figure 2, the model (i) performs per-sample adaptive window selection to choose the processing scale, (ii) derives interpretable time–frequency representations via STFT coupled with 2D convolutions, and (iii) models long-range dependencies with a transformer whose self-attention is stabilized by a Tikhonov-regularized pseudoinverse to enhance numerical robustness and computational efficiency.

Section 2.2 introduces the trend- and dominant-frequency-guided complex routing for window selection. Section 2.3 explains how the selected window jointly determines the STFT/2D-convolution hyperparameters and the construction of the time–frequency tensor. Section 2.4 presents Nyström attention with a Tikhonov-regularized pseudoinverse.

2.2. Adaptive multi-scale time windows

To capture the multi-scale, time-varying characteristics of microseismic signals, we proposed a dynamic

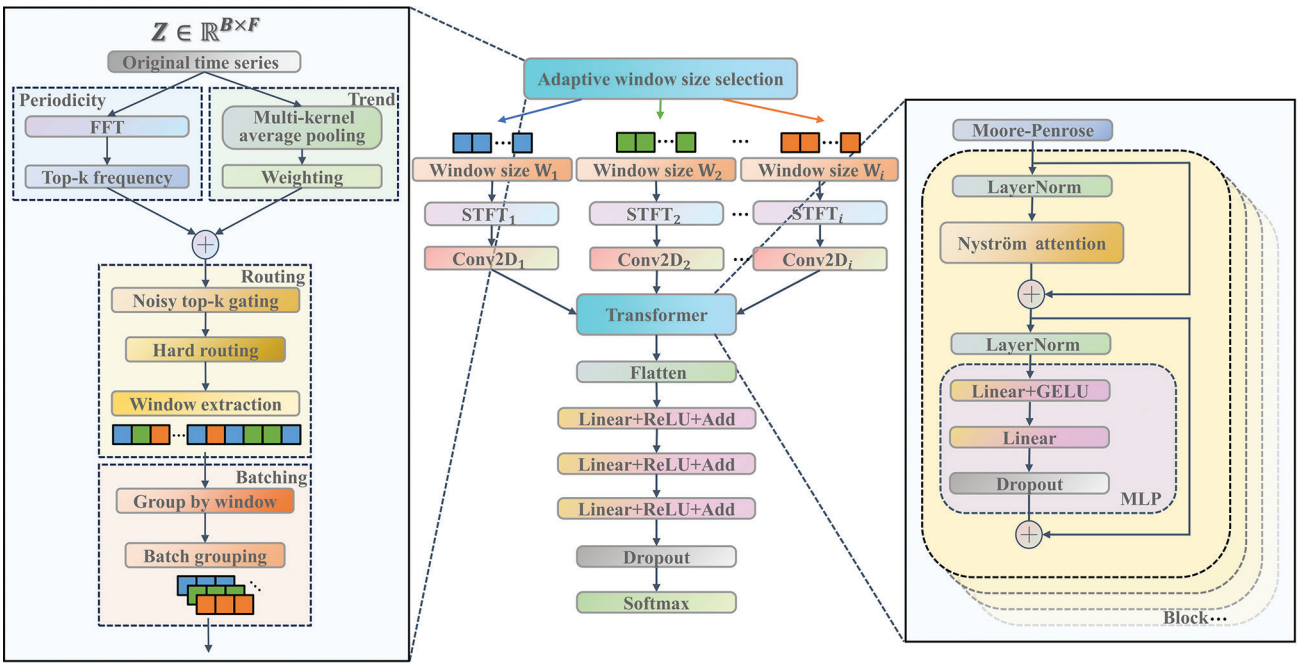


Figure 2. Overview of the SeisFormer model architecture
Abbreviations: Conv2D: Convolutional two-dimensional layer; FFT: Fast Fourier transform; MLP: Multilayer perceptron; STFT: Short-time Fourier transform.

window-selection method that fuses trend and periodicity cues. The technique first extracts multi-scale trend and periodicity representations from the input sequence and compresses them along the temporal axis to obtain a compact indicator vector *f*. We then scored a pre-defined candidate window set {*W*1,…*Wn*} with *f*; during training, light Gaussian noise was injected into the scores to improve generalization. During inference, a hard-routing strategy (k = 1) selected the optimal window length W_{i^*} to drive the subsequent time–frequency modeling.

Let the input be $z \in R^{B\times W\times D}$, where *B* is the batch size, *W* is the maximum observation window considered by the selector, and *D* = 1 is the channel dimension. Given a scale set $K=\{k_1|,\ldots,k_m|\}$, we computed multi-scale moving averages using **Equation (I)**:

$$z_{\text{trend}}^{(k)} = \mathrm{MA}(z;k) \in \mathbb{R}^{B\times W\times D},\ \forall k \in \mathcal{K} \tag{I}$$

The per-scale trend components were fused with learnable weights, as shown in **Equation (II)**:

$$z_{\text{trend}} = \sum_{k=1}^{m} \alpha_k z_{\text{trend}}^{(k)},\ \alpha_k = \frac{\exp(\omega_k)}{\sum_{j=1}^{n}\exp(\omega_j)},$$
$$\omega_k = W_{\text{trend}} \cdot \mathrm{AvgPool}(z_{\text{trend}}^{(k)}) + b_{\text{trend}} \tag{II}$$

Where AvgPool(·) denotes temporal average pooling and $\{a_k\}$ are softmax-normalized scale weights.

In terms of periodicity features, for each sample, we applied a fast Fourier transform to obtain a complex spectrum Z_{FFT}. The magnitude was calculated using **Equation (III)**:

$$|Z_{FFT}| = \sqrt{\mathrm{Re}(Z_{FFT})^2 + \mathrm{Im}(Z_{FFT})^2} \tag{III}$$

To emphasize dominant periodic components, we selected the top-κ frequency indices by magnitude and retained the corresponding real-valued magnitude features, yielding $Z_{freq}^{(\kappa)}$. This reduces dimensionality while preserving the principal periodic structure.

For fusion and selection, we performed temporal average pooling on the trend and periodicity features separately and used additive fusion to obtain the indicator vector, as shown in **Equation (IV)**:

$$f = \mathrm{AvgPool}(z_{\text{trend}}) + \mathrm{AvgPool}(Z_{\text{freq}}^{(\kappa)}) \tag{IV}$$

Candidate windows were scored by logits = $W_g f + b_g$. During training, we added zero-mean Gaussian noise $N(0, \sigma^2)$ to the logits (with σ selected on the validation set) to mitigate overfitting near decision boundaries and improve out-of-distribution robustness. During inference, we adopted hard routing by choosing a single window via $i^* = \arg\max_i \text{logits}_i$, and used W_{i^*} for subsequent time–frequency modeling. The fusion and routing procedure is summarized in Algorithm 1.

For dynamic bucketing and end alignment, since samples within the same batch can select different window lengths, we dynamically grouped (bucketed) samples by their chosen window and formed sub-batches per window. For each group, a sample was fed to the segment obtained

Algorithm 1. Trend-period fusion and hard routing

Inputs: $z \in R^{B\times W\times 1}$, kernel set $K = \{k_1|,\ldots, k_m|\}$, top−κ

Outputs: selected index i^* and window length W_{i^*}

1: //Multi-scale trend extraction

2 For each $k \in K$ do

3: $z_{\text{trend}}^{(k)|} \leftarrow \mathrm{MovingAverage}(\mathbf{z};\ \text{window} = k)$

4: end for

5: //Learnable weighting across scales (softmax on pooled cues)

6: for each $k \in K$ do

7: $u_k \leftarrow \mathrm{AvgPool}(z_{\text{trend}}^{(k)})$ // pool along time

8: $\alpha_k \leftarrow W_{\text{trend}} \cdot u_k + b_{\text{trend}}$

9: end for

10: $\alpha \leftarrow \mathrm{softmax}([\omega_k]_k)$ // $B\times m$, along k

11: $\bar{Z}_{\text{trend}} \leftarrow \sum_k \alpha_k z_{\text{trend}}^{(k)}$ // $B\times W\times 1$

12: //Periodicity via FFT (keep Top-*k* complex components)

13: $Z_{\text{fft}} \leftarrow \mathrm{FFT}(z)$

14: mag $\leftarrow \sqrt{\Re(Z_{\text{fft}})^2 + \Im(Z_{\text{fft}})^2}$

15: idx ←TopK (mag,κ)//per−sample

16: Z_{freq} ←Gather (mag, idx)

17: //Additive fusion and hard routing

18: f $\leftarrow \mathrm{AvgPool}(\bar{Z}_{\text{trend}}) + \mathrm{AvgPool}(Z_{\text{freq}})$

19: logits ←$W_g f + b_g$

20: if training then

21: logits ←logits+ε//$\varepsilon \sim N(0,\sigma^2)$

22: *P* ←softmax (logits)//optional: for logging/analysis

23: i^* ←arg $\max_i p_i$

24: else

25: V//Hard routing at inference: k=1

26: i^* ←arg $\max_i$ (logits)

27: end if

28: return i^*, W_{i^*}

by slicing from the sequence end leftward with length $W_{i\backslash^*}$; the supervision signal (class label and/or arrival-time label) was aligned to the window end. This ensured a one-to-one correspondence among "context length–feature extraction–supervision" while preserving batch efficiency. The selector consisted only of linear transforms and pooling, and dynamic grouping was a tensor reindexing/slicing operation; the overall computational overhead was negligible.

2.3. Time–frequency feature extraction based on STFT and 2D convolution

The STFT preserves both temporal and spectral information and is therefore well suited to low-SNR, compositionally similar, time-varying microseismic signals.[32] In this work, we mapped the input microseismic sequence from the time domain to the time–frequency domain to extract more discriminative spectral features and characterize energy evolution across time. The STFT window length is a key hyperparameter: Increasing the window improves frequency resolution (smaller $\Delta f = f_{s|}/n_{fft}$) but reduces time resolution (larger $\Delta t = H/f_s$)and increases temporal smoothing $\tau_{win} = n_{win}/f_s$; the converse holds for shorter windows.

Let $f_{s|}$ be the sampling rate, n_{fft} the DFT size, n_{win} the window length (we set $n_{fft} = n_{win}$), and $H$the hop size. For a discrete signal $z[n]$, the STFT was calculated using **Equation (V)**:

$$Z(t,f)=\sum_{n=0}^{N-1} z[n]\omega[n-tH]e^{-j2\pi fn/\mathrm{N}} \quad \text{(V)}$$

Where ω[·] is the analysis window (we used a Hann window), t indexes time frames, and f indexes frequency bins. For a batch $z\in R^{B\times T}$, the STFT produced **Equation (VI)**:

$$Z_{STFT}\in\mathbb{C}^{B\times F\times T'},\ F=\left[\frac{n_{fft}}{2}\right]+1,\ T'\approx\left[\frac{t-n_{win}}{H}\right]+1 \quad \text{(VI)}$$

with the Nyquist limit $f_{s|}/2$. To retain both magnitude and phase while matching a 2D CNN, we stacked the real and imaginary parts along the channel dimension, forming stack $[Z_{real}, Z_{imag}]\in R^{B\times 2\times F\times T'}$, which was then passed to a 2D convolution followed by ReLU, as shown in **Equation (VII)**:

$$Z_{conv} = \mathrm{ReLU}(W\circledast \mathrm{stack}[Z_{real}, Z_{imag}]+b) \quad \text{(VII)}$$

where ⊛ denotes 2D convolution on the frequency–time plane. The 2D CNN captures local structures within a single band and cross-band/time dependencies, enabling short-term spectral trend modeling and inter-band coordination.

Following the hard-routing selection in Section 2.2, once the sample-level optimal window $W_i^*\in\{128,256,512\}$ was chosen, we adapted n_{fft} and the hop size H accordingly (50% overlap, $H \approx n_{fft}/2$), and proportionally adjusted the number of 2D convolution channels to achieve comparable time–frequency resolution and controlled computation across scales. The mapping used in this paper is shown in **Equation (VIII)**:

$$\left(n_{fft},H,channels\right)=\begin{cases}(62,31,12), & W_i^*=128\\ (126,63,6), & W_i^*=256\\ (256,127,3), & W_i^*=512\end{cases} \quad \text{(VIII)}$$

motivated by setting $n_{fft}\approx W_i^*/2$ to align $\Delta f = f_{s|}/n_{fft}$ across scales. Larger windows increase frequency resolution but also the number of frequency bins FFF; hence, we reduced the convolution channels inversely (12→6→3) to offset the growth in feature-map size and stabilize throughput. Modern fast Fourier transform implementations handle non-power-of-two lengths efficiently, so the above choices were numerically and computationally sound. This parameterization was empirically validated as the best-performing configuration, yielding the strongest trade-off among arrival-time accuracy, classification metrics, and efficiency on the validation set, and demonstrating stable behavior in ablation studies.

2.4. Nyström attention with Tikhonov-regularized pseudoinverse

After the 2D convolution, the frequency–time maps were reshaped to form a sequence for the transformer. Let $Z_{conv}\in\mathbb{R}^{B\times C\times F\times T''}$ denote the convolutional output (batch B, channels C, frequency bins F, frames T''). We flatten the (C, F) axes to obtain **Equation (IX)**:

$$X\in R^{B\times n\times d},\ n \coloneqq T'',\ d \coloneqq C\times F \quad \text{(IX)}$$

and fed X to the transformer (SeisFormer) for further sequence modeling. Multi-head self-attention captured long-range temporal–spectral dependencies, improving microseismic event discrimination.

For notation unification, we set $n \coloneqq T''$ (sequence length after 2D CNN) and $d \coloneqq C\times F$ (embedding width before head-splitting). With h heads, the per-head width was $d_h = d/h$. Given $X\in R^{B\times n\times d}$, after linear projections and head splitting, we have $Q,K,V\in\mathbb{R}^{B\times h\times n\times d_h}$. In **Equation (X)**, the scaling $\sqrt{d}$ refers to the per-head width, that is, $d\equiv d_h$.

SeisFormer alternates between self-attention and feed-forward neural network (FFN) blocks and, unlike a standard transformer, employs a Nyström approximation to self-attention for efficiency on long sequences.[33-37] In conventional attention, the row-wise scaled dot-product was calculated using **Equation (X)**:

$$S(Q,K)=softmax\left(\frac{QK^{\top}}{\sqrt{d}}\right)\in\mathbb{R}^{n\times n} \quad (X)$$

We chose $m \ll T$ landmarks with index set M, and defined it as **Equation (XI)**:

$$A \triangleq S(Q_M, K_M)\in\mathrm{R}^{m\times m},\ B \triangleq S(Q, K_M)\in\mathrm{R}^{n\times m},\ C \triangleq S(Q_M, K)\in\mathrm{R}^{m\times n} \quad (XI)$$

The classical Nyström approximation is shown in **Equation (XII)**:

$$\hat{S}=BA^{+}C \quad (XII)$$

where A^{+} is the Moore–Penrose pseudoinverse. If the true attention S has rank at most m and the landmark submatrices are full rank, we can write $S = UV^{\top}$ with U, $V\in\mathrm{R}^{n\times m}$, which yields $B=UV_M^{\top}$, $C=U_MV^{\top}$ and $A=U_MV_M^{\top}$. Using $(XY)^{+} = Y^{+}X^{+}$ under the usual full-rank side conditions, we obtain **Equation (XIII)**:

$$BA^{+}C=(UV_M^{\top})((V_M^{+})^{\top}U_M^{+})(U_MV^{\top})=U\underbrace{[V_M^{\top}(V_M^{+})^{\top}]}_{=I_m}\underbrace{[U_M^{+}U_M]}_{=I_m}V^{\top}$$
$$=UV^{\top}=S \quad (XIII)$$

Hence, the Nyström reconstruction is exact in this ideal case. In general, $\hat{S}$ still preserves the landmark rows/columns ($\hat{S}_{:,M}=B, \hat{S}_{M,:}=C$) and gives the minimum-norm solution consistent with them.

For numerical stability, we replaced A^{+} with a Tikhonov-regularized pseudoinverse A_{λ}^{+}. Let the SVD of A be the formula shown in **Equation (XIV)**:

$$A=U_A\Sigma V_A^{\top},\ A^{+}=V_A\Sigma^{+}U_A^{\top},\ \Sigma^{+}=diag\left(\sigma_i^{-1}1_{\{\sigma i>0\}}\right) \quad (XIV)$$

When the landmarks are highly correlated or the subset is skewed, small singular values make the plain inverse amplify noise along those directions. In practice, we used the Tikhonov-regularized pseudoinverse, as shown in **Equation (XV)**:

$$A_{\lambda}^{+}=\left(A^{\top}A+\lambda I\right)^{-1}A^{\top}=V_Adiag\left(\frac{\sigma_i}{\sigma_i^{2}+\lambda}\right)U_A^{\top} \quad (XV)$$

Whose operator norm satisfies **Equation (XVI)**:

$$A_{\lambda}^{+}{}_{2}=\max_i\frac{\sigma_i}{\sigma_i^{2}+\lambda}\le\frac{1}{2\sqrt{\lambda}} \quad (XVI)$$

thereby avoiding the $1/\sigma_i$ blow-up as $\sigma_i\to 0$. The resulting stable reconstruction is shown in **Equation (XVII)**:

$$S_{\lambda}=BA_{\lambda}^{+}C \quad (XVII)$$

Spectrally, the regularization acts as a smooth shrinkage on each singular direction, as seen in **Equation (XVIII)**:

$$AA_{\lambda}^{+}A=U_Adiag\left(\frac{\sigma_i^{3}}{\sigma_i^{2}+\lambda}\right)V_A^{\top},\ A-AA_{\lambda}^{+}A$$
$$=U_Adiag\left(\frac{\lambda\sigma_i}{\sigma_i^{2}+\lambda}\right)V_A^{\top} \quad (XVIII)$$

Hence, the approximation-bias trade-off is monotone in λ. This shrinkage bounds the amplification of perturbations and yields smoother gradients during back-propagation, as the Lipschitz constant along the landmark path is controlled by $A_{\lambda}^{+}{}_{2}$.

In terms of complexity, exact attention incurs $O(n^2d)$ time and $O(n^2)$memory, whereas the Nyström scheme requires $O(nmd) + O(m^3)$ to construct B and $C$and to solve a single $m \times m$ system. Under the common regime, $m << n$, the $O(m^3)$term is negligible, and the overall complexity is effectively $O(nmd)$. Replacing A^{+} with A_{λ}^{+} preserved this low-rank acceleration while improving numerical stability for long-range dependency modeling in microseismic signals; this matches our implementation, which computes the (regularized) pseudoinverse on the landmark attention block.

The processed sequence features were flattened and passed to each transformer layer. In each layer, the features were further optimized through the FFN, which consists of linear layers and GELU activations to extract non-linear relationships and enhance feature representation. Residual connections and layer normalization were applied to both self-attention and FFN blocks to accelerate training and prevent gradient vanishing, ensuring stable signal propagation through the network and better adaptation to complex time–frequency structure. The output features then pass through three linear transformations with ReLU and dropout, followed by a final linear layer that maps to the task space; finally, scores were normalized to predict probabilities for P-waves, S-waves, and noise, completing the microseismic event classification.

3. Experimentation

3.1. Parameter configuration

To enhance data representativeness and rigorously evaluate cross-site generalization, we merged data from two independent sites into a joint dataset under a unified organization and labeling protocol: An HDR project in the Gonghe Basin, Qinghai, China, and an unconventional hydraulic-fracturing site in North China. The Qinghai data were acquired in 2020 using an in-house system. Monitoring at well GR-1 (approximately 2 km from GH-02/3) used a 12-level, three-component downhole array (1,100–1,400 m depth; 20 m inter-level spacing) together with a "surface–shallow-well–deep-well" joint layout: 12 surface lines

within an 8 × 8 km area centered on GH-02/3 (25 m station spacing; at least 1,512 channels) and 60 three-component shallow-well instruments (10–25 m installation depth), providing coverage to a target depth of ~4,000 m. At the North China site, production wells were arranged in belts, targeting formations at depths of ~3,700–4,300 m. The two sites shared a consistent data organization and labeling protocol: Both used a 1 ms sampling interval, and phases were labeled as noise = 0, P = 1, and S = 2 (Figure 3). Across all annotated frames, class proportions were 70.01% noise, 13.70% P, and 16.29% S.

Modeling was conducted on the joint dataset, comprising a total of 4,000 single-channel time series (Δt = 1 ms)as inputs. Each channel was demeaned and standardized via z-score using statistics computed from the training split of the joint dataset to ensure comparability across sites and channels. Unless otherwise stated, the data were split in a ratio of 8:1:1 into training/validation/test sets. We adopted Adam (initial learning rate 1×10^{-4}) with ReduceLROnPlateau (factor = 0.1, patience = 5) based on validation metrics to promote stable convergence. The training objective combined cross-entropy with L2 regularization (weight decay = 0.003). After each epoch, the model was evaluated on the validation set, and early stopping was applied to curb overfitting and improve generalization.

Given the high noise fraction (70.01%) in the joint dataset, we employed class-weighted cross-entropy during training and assigned a weight of 1.2 to P- and S-phase frames to strengthen discrimination around arrivals, thereby improving picking sensitivity and robustness. All training and evaluation settings were applied uniformly across both sites to ensure fair comparison and reproducibility.

3.2. Comprehensive experimental evaluation

To comprehensively evaluate model performance under different conditions, we designed a series of experiments on a strictly held-out test set from the joint dataset. This test set consisted of 200 data segments from each site (400 in total) and was entirely non-overlapping with the training/validation data. The evaluation scenarios included the real environment, the noise environment, and multi-event cases. To ensure fairness and reproducibility, all methods followed a unified pre-processing pipeline before entering their respective models/algorithms. The evaluation protocol then proceeded in four stages: (i) Multiple methods were compared under the real scenario; (ii) better-performing methods were included in the noise tests; (iii) complex-signal handling was assessed via the multi-event scenario; and (iv) ablation studies were conducted to quantify the contributions of key components.

The class set be {0:Noise, 1:P, 2:S}. Define the confusion matrix be $C \in N^{3\times3}$ with entries C_{ii} = #{samples with true class i predicted as j}, i, j∈{0,1,2}.Then, the overall accuracy is calculated using **Equation (XIX)**:

$$\text{Accuracy} = \frac{\sum_{i=0}^{2} C_{ii}}{\sum_{i=0}^{2}\sum_{j=0}^{2} C_{ij}} \tag{XIX}$$

For class *i*, **Equation (XX)** was used:

$$TP_i = C_{ii},\ FP_i = \sum_{k\neq i} C_{ki},\ FN_i = \sum_{k\neq i} C_{ik} \tag{XX}$$

Class-wise precision, recall, and F1 are shown in **Equation (XXI):**

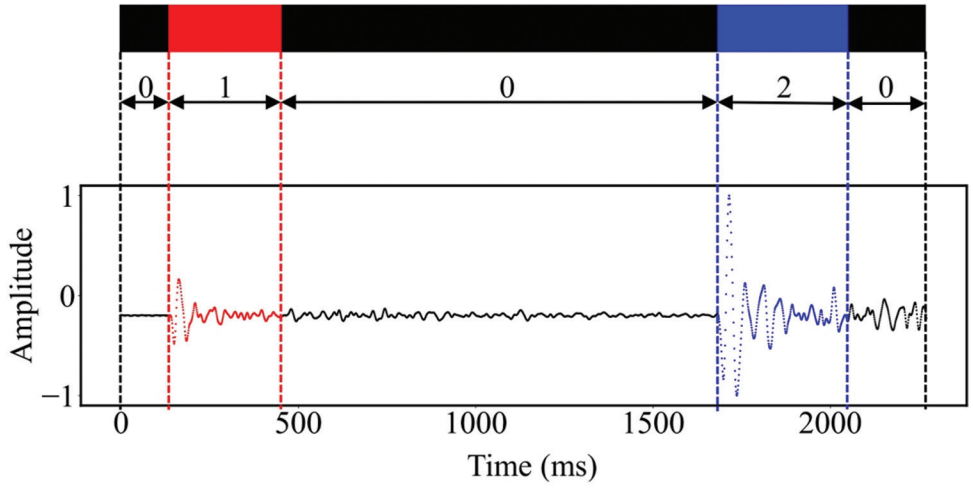


Figure 3. Manual annotation process

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i}, \quad \text{Recall}_i = \frac{TP_i}{TP_i + FN_i},$$
$$\text{F1}_i = \frac{2\text{Precision}_i\text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \tag{XXI}$$

Weighted aggregates used for class support $n_i = TP_i + FN_i$ (i.e., the number of true samples in class i) with $N = \Sigma_i n_i$, as shown in **Equation (XXII)**:

$$\text{Precision} = \sum_{i=0}^{2} \frac{n_i}{N}\text{Precision}_i, \quad \text{Recall} = \sum_{i=0}^{2} \frac{n_i}{N}\text{Recall}_i,$$
$$\text{F1} = \sum_{i=0}^{2} \frac{n_i}{N}\text{F1}_i \tag{XXII}$$

We also reported mean absolute error (MAE)-P and MAE-S for P/S arrival times, defined as the sample-wise mean absolute difference between the predicted and manually annotated arrivals. Together, these metrics and visualizations quantified both classification and arrival-time picking performance and enabled a consistent comparison of methods across the three scenarios.

3.2.1. Real environment experiment

We evaluated classification performance on real microseismic signals using SeisFormer, EQTransformer,[23] PhaseNet,[20] generalized phase detection (GPD),[38] LSTM, CNN, and STA/LTA.[10] Unless otherwise stated, all models were trained from scratch under a unified pre-processing pipeline, with a sampling rate of 1 kHz, identical train/validation/test splits, and identical label definitions. Model configurations were as follows: SeisFormer—an 8-layer Transformer with eight attention heads, model dimension 64, FFN/multilayer perceptron hidden size 256, dropout 0.5 on attention and FFN, and a multilayer perceptron head with 128 hidden units. EQTransformer was implemented following the public release and original architecture (convolutional encoder, residual CNN stack, 3×BiLSTM, detection decoder branch with multi-stage upsampling). PhaseNet is a one-dimensional U-Net with four down- and upsampled stages (downsampling kernel length 7, stride 4). GPD used four Conv1D layers plus two fully connected layers. The LSTM baseline used a two-layer bidirectional LSTM with 100 hidden units per direction. The CNN baseline is a lightweight one-dimensional CNN with three convolutional blocks and a fully connected head (kernel length 7; channels 32/64/128). STA/LTA is a short/long-window energy-ratio trigger under the same pre-processing/segmentation as the deep models (short/long windows 0.2s/2.0s; threshold tuned on the validation set). For fairness, we used matched optimization, regularization, learning-rate scheduling, early stopping, and random seeds across methods, without modifying baseline architectures.

As shown in Figure 4, SeisFormer, EQTransformer, and PhaseNet clearly outperformed the other baselines. Representative numbers are reported in Table 1: SeisFormer (Accuracy: 98.30%, precision: 97.40%, recall: 97.92%, F1: 97.66%; MAE-P: 1.42 ms, and MAE-S: 2.29 ms), EQTransformer (Accuracy: 96.90%; MAE-P: 1.90 ms, and MAE-S: 3.18 ms), PhaseNet (Accuracy: 94.80%; MAE-P: 4.76 ms, and MAE-S: 6.95 ms), while the remaining baselines lagged substantially behind. Overall, these three models constituted the first tier, with SeisFormer leading in both classification and arrival-time accuracy.

We also benchmarked forward-pass latency on an RTX 4060 (8 GB) + Intel i9-13900HX using single-channel 1 kHz/3 s input (the three-second window was used solely to standardize the latency benchmark), FP32, batch = 1. Results were the median of 100 runs after 20 warm-up iterations, measuring wall-clock time for the model forward only—including in-graph STFT and hard routing, and excluding data loading and disk I/O: SeisFormer ≈ 4.2 ms (GPU)/43 ms (CPU), PhaseNet ≈ 5.1 ms (GPU)/55 ms (CPU), EQTransformer ≈ 8.5 ms (GPU/94 ms (CPU). Under this accuracy–latency trade-off, SeisFormer is the most suitable for near–real-time deployment on the target hardware.

Confusion matrices for each method in Figure 5 further illustrate their strengths and weaknesses. SeisFormer attained an overall true-positive rate of 98.1%, with class-wise rates of 96.1% (P-wave) and 94.3% (S-wave), outperforming all other methods. Notably, most SeisFormer errors arise from small discrepancies between predicted and manually labeled endpoints of P- and S-wave arrivals; such endpoint disagreements have limited impact on microseismic monitoring and are therefore of low significance to the overall evaluation. To further substantiate SeisFormer's advantages, Figure 6A-D shows predictions on representative waveforms from the test sets of both datasets, visually demonstrating efficient classification and accurate arrival-time calibration.

3.2.2. Noise environment experiment

To more faithfully emulate field disturbances and align the evaluation with picking/classification objectives, we calibrated noise intensity using an event-referenced SNR (ER-SNR) and conducted stress tests with non-stationary composite noise that included low-frequency drift, power-line fundamentals and harmonics, impulsive interference, and colored background noise. This design better reflected real HDR noise characteristics than conventional whole-trace SNR and enabled an objective assessment of model robustness under realistic conditions. Concretely, for each record, we constructed an event window E (labels >0) and

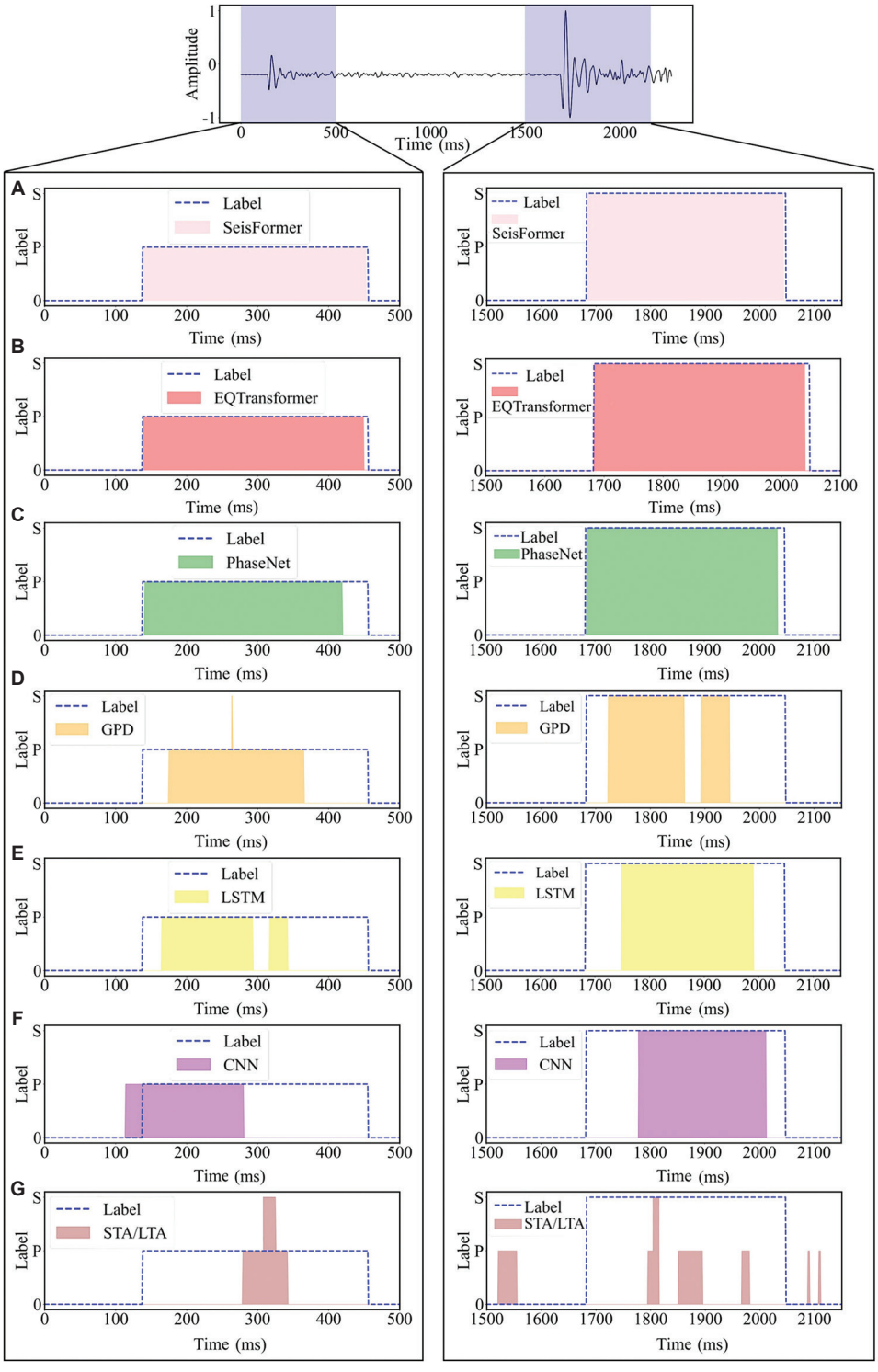


Figure 4. Comparison of classification and picking results from different methods. The same waveform was used to test each model, with the classification results for the 0–500 ms segment extracted to showcase performance in P-wave classification and arrival-time picking, and the 1,500–2,100 ms segment extracted to highlight performance in S-wave classification and arrival-time picking. SeisFormer, EQTransformer, and PhaseNet demonstrated strong performance; further comparisons and evaluations will be conducted in subsequent noise experiments. Classification results of the (A) SeisFormer model, (B) the EQTransformer model, (C) the PhaseNet model, (D) the GPD model, (E) the long short-term memory (LSTM) model, (F) the convolutional neural network (CNN) model, and (G) the short-term average/long-term average (STA/LTA) method.

Table 1. Comparison of classification performance and arrival time calibration errors for different models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	Mean P-wave arrival error	Mean S-wave arrival error
SeisFormer	98.30	97.40	97.92	97.66	1.42 ms	2.29 ms
EQTransformer	96.90	96.15	96.48	96.31	1.90 ms	3.18 ms
PhaseNet	95.80	95.02	95.71	95.36	4.76 ms	6.95 ms
GPD	83.70	81.80	82.45	82.12	24.9 ms	30.6 ms
LSTM	85.90	85.10	85.35	86.10	15.1 ms	45.3 ms
CNN	82.10	80.10	81.95	81.30	21.4 ms	54.9 ms
STA/LTA	68.79	61.60	66.42	64.63	152 ms	224 ms

Abbreviations: CNN: Convolutional neural network; GPD: Generalized phase detection; LSTM: long short-term memory; STA/LTA: Short-term average/long-term average.

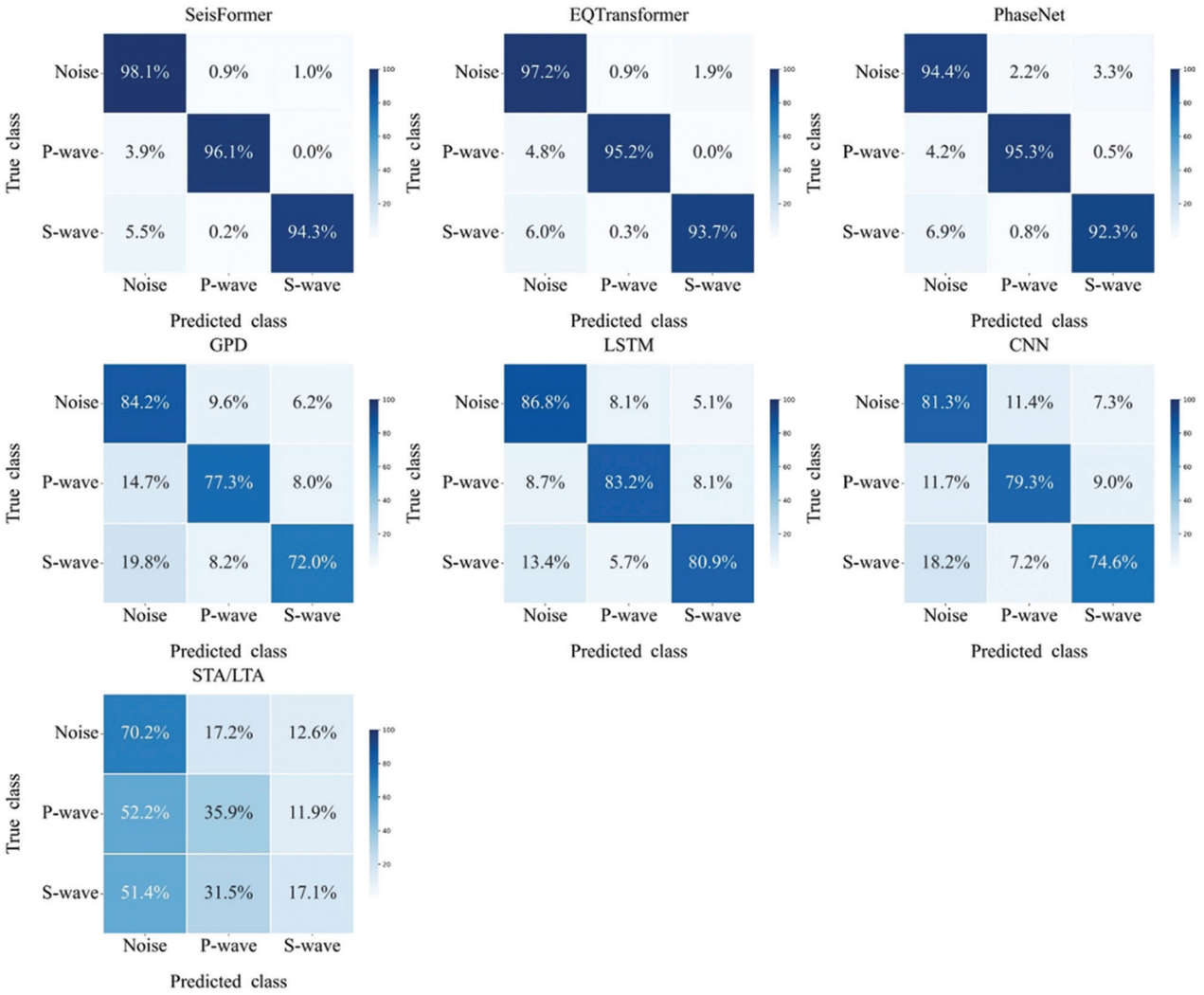


Figure 5. Comparison of confusion matrices for SeisFormer, EQTransformer, PhaseNet, GPD, LSTM, CNN, and STA/LTA on microseismic signal classification Abbreviations: CNN: Convolutional neural network; GPD: Generalized phase detection; LSTM: long short-term memory; STA/LTA: Short-term average/long-term average.

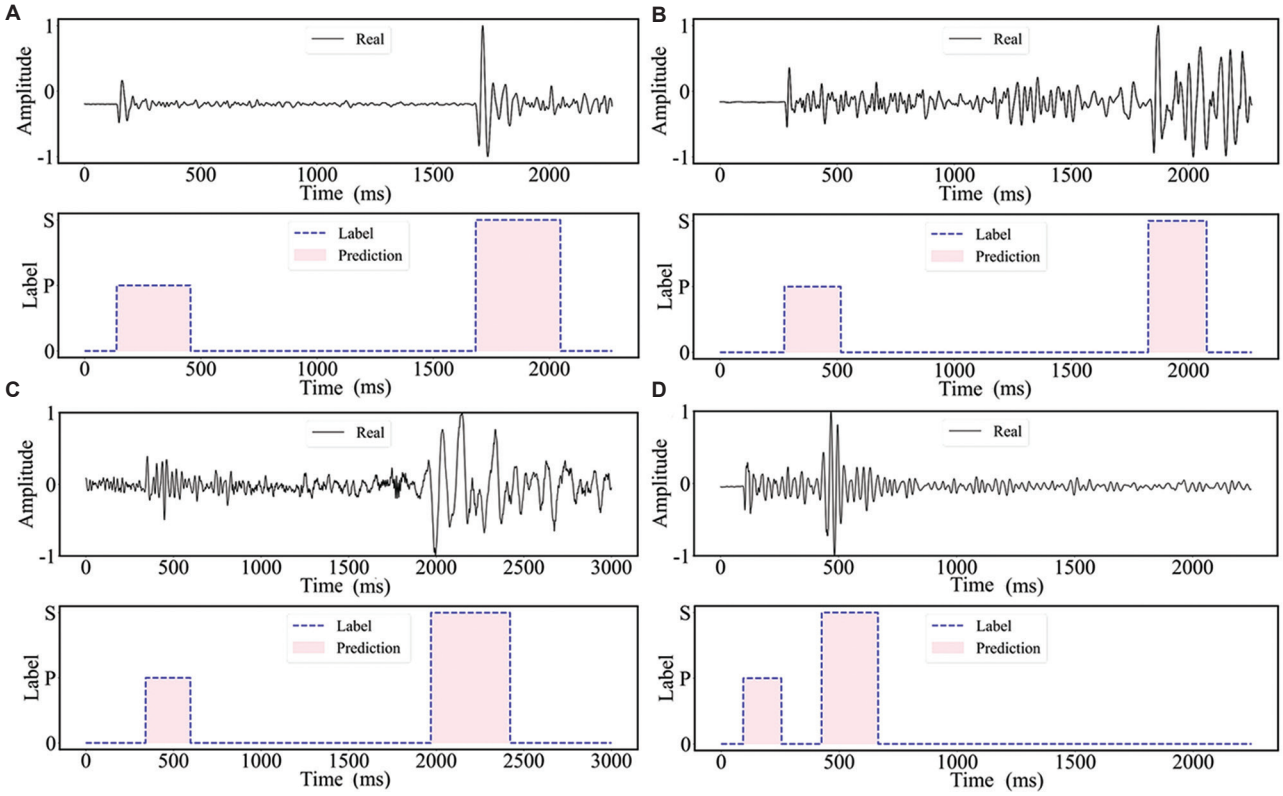


Figure 6. Classification results of the SeisFormer model on representative waveforms from the training set, verifying performance across different data types. (A and B) Qinghai site. (C and D) North China site.

a pre-event baseline window B from the second-column labels. The event mask was dilated by approximately ±100 samples to cover onsets and coda. The waveform was then baseline-centered using $\mu_B = \frac{1}{|B|}\sum_{i\in B} s[i]$, yielding $\tilde{s}[i] = s[i] - \mu_B$, which suppresses low-frequency drift in power estimation. In the presence of impulses and non-stationarity, we obtained stable energy estimates by combining trimmed mean of squares with baseline bootstrap length-matching: For any segment we averaged the squared amplitudes after two-sided 10% trimming to reduce the leverage of outliers; as $|B| \gg |E|$ for most records, we repeatedly sampled from B subsegments of length $|E|$, computed trimmed power for each replicate, and averaged across K = 30 replicates to remove biases due to unequal window lengths. This yielded **Equation (XXIII)**:

$$P_E E = Trim_{0.1}\left(\tilde{s}\left[E\right]^2\right),\ P_B = \mathbb{E}_{boot}\left[Trim_{0.1}\left(\tilde{s}\left[B_{sub}\right]^2\right)\right] \tag{XXIII}$$

and the ER-SNR (in dB) was defined as **Equation (XXIV)**:

$$ER - SNR_{dB} = 10\log_{10}\left(\frac{P_E}{P_B}\right) \tag{XXIV}$$

To match the field noise spectra, we did not add stationary Gaussian white noise. Instead, we synthesized a non-stationary composite of four components—baseline drift (random walk or 1/f -like), power-line harmonics (50/60 Hz and overtones with slow AM/PM), sparse impulsive spikes, and colored AR (1) background—and linearly mixed them with fixed relative weights, as shown in **Equation (XXV)**:

$$n_{raw} = w_d n_{drift} + w_h n_{harm} + w_i n_{imp} + w_c n_{col} \tag{XXV}$$

using w_d = 1.0, w_h = 1.0, w_i = 0.6, and w_c = 0.8. The noise powers within E and B, $P_{nE0}|$ and P_{nB0}, were estimated with the same robust procedure. Given a target ER-SNR level (let $\gamma = 10^{ER-SNRdB/10}$, the composite noise was scaled and added as $x = s + \alpha n_{raw}$ so that the post-augmentation event/baseline power ratio met the target, as shown in **Equation (XXVI)**:

$$\alpha^2 = \frac{P_E - \gamma P_B}{\gamma P_{nB0} - P_{nE0}} \tag{XXVI}$$

If $P_{E|} - \gamma P_B \le 0$ (the trace is already cleaner than the target) or $P_{nB0} - P_{nE0} \le 0$ (the noise recipe concentrates relatively more energy in E than in B, contradicting the

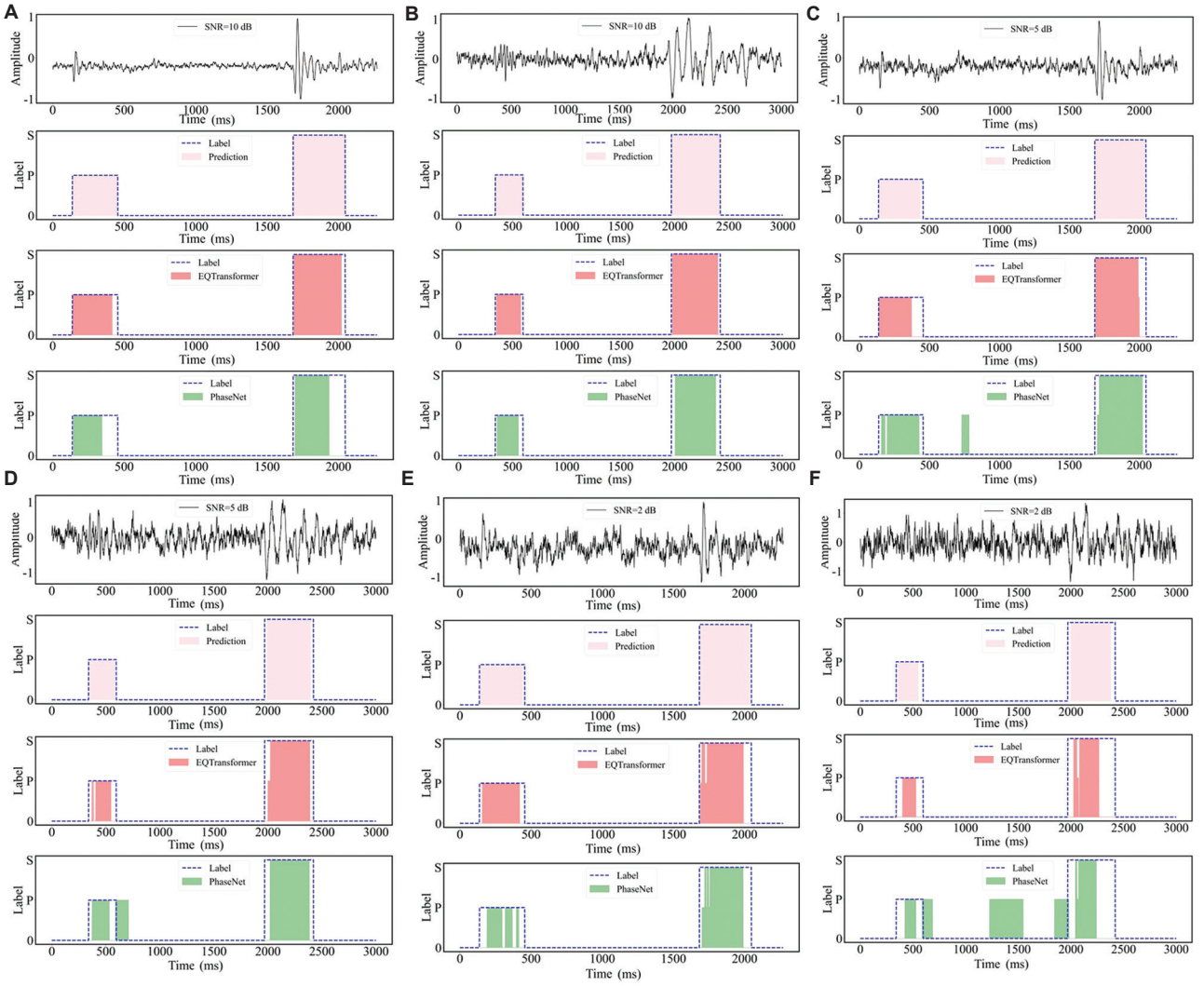


Figure 7. Phase classification and arrival-time picking across event-referenced signal-to-noise ratio (ER-SNR) levels. (A, C, and E) Qinghai. (B, D, and F) North China. Each column shows the same representative record under three noise settings (10/5/2 dB); rows are, from top to bottom, raw/noisy waveform (with ER-SNR), SeisFormer, EQTransformer, and PhaseNet. Performance degrades as ER-SNR decreases; SeisFormer consistently exhibits more precise and temporally coherent P/S predictions, smaller picking bias, and slower growth in false/missed detections across both datasets.

target balance), we kept the original waveform to avoid unrealistic distortion. All augmentations used fixed random seeds for reproducibility and fairness, and for each record and level, we computed and logged the achieved ER-SNR to verify calibration error against the target.

We adopted three ER-SNR levels (10 dB, 5 dB, and 2 dB), corresponding to moderate, strong, and extreme degradation, with a no-noise condition as the baseline. For each level, waveforms, spectrograms, and augmented samples were generated under fixed random seeds for inference and visualization. To provide representative comparisons, we selected two records from Figure 6 (panel A: Qaidam; panel D: North China) and evaluated SeisFormer, EQTransformer, and PhaseNet across the four noise conditions (no noise, 10 dB, 5 dB, 2 dB). All results were reported as raw values (accuracy, precision, recall, F1, and P/S arrival MAE), accompanied by corresponding waveforms and STFT spectrograms to visually demonstrate the degradation trend with increasing noise. Spectrograms used frame-wise adaptive STFT: At each time position, the window length was chosen by the selector in Section 2.2, and the STFT for that frame was computed as in Section 2.3; frames from different windows are interpolated onto a unified time–frequency grid and concatenated to form a continuous spectrogram. We also overlaid a window-identifier ribbon aligned to the time axis to indicate the time–frequency resolution used per segment. Related visualizations and noise-robustness curves are shown in Figures 7 and 8.

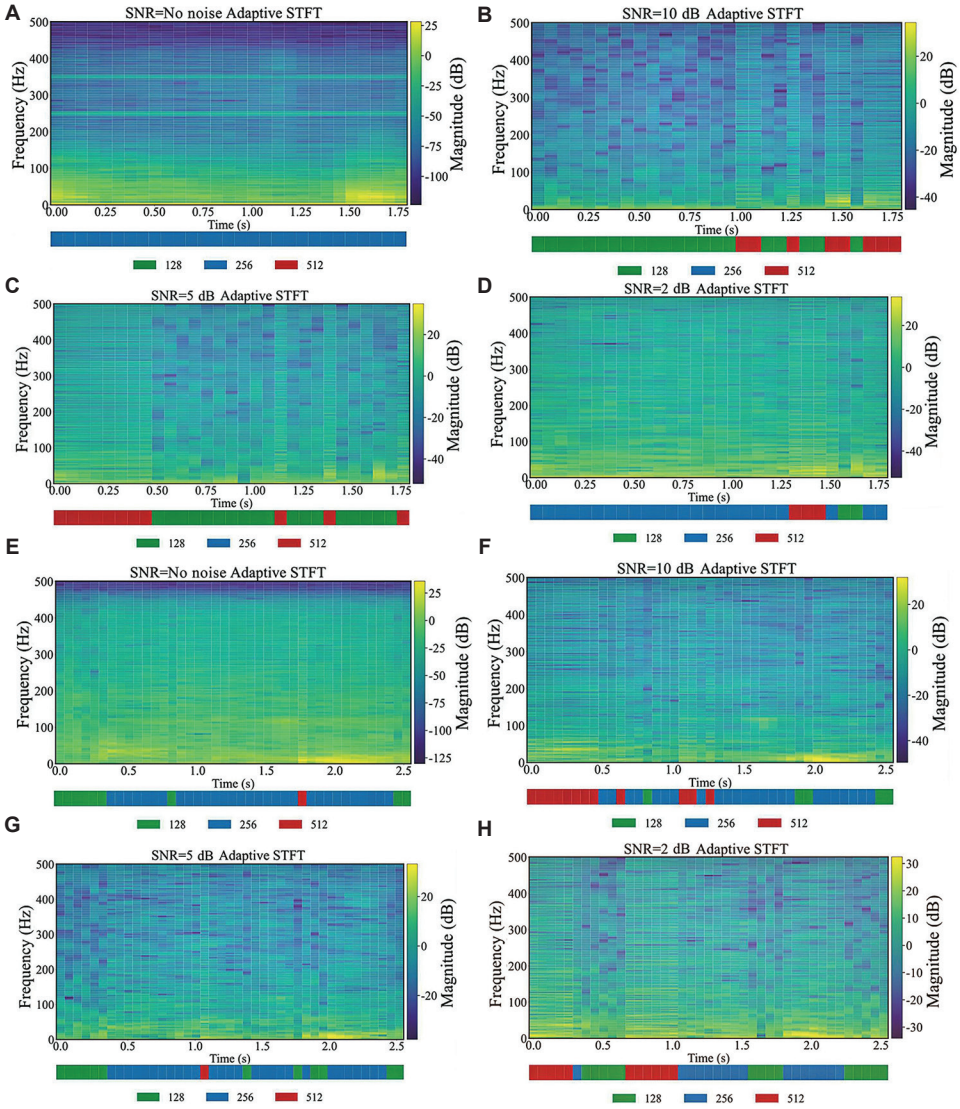


Figure 8. Adaptive short-time Fourier transform (STFT) time–frequency spectrograms with window-selection indicator bars. (A-D) Qinghai. (E-H) North China. Shown are the adaptive STFT magnitude spectra of the same two records as in Figure 7 under different noise/scenario settings (vertical axis: Frequency/Hz; horizontal axis: time/s; color scale: magnitude/dB). Below each spectrogram, the colored bar indicates the frame-wise window selection (green = 128, blue = 256, red = 512) with pixel-wise alignment to the spectrogram's time axis.
Abbreviation: SNR: Signal-to-noise ratio.

As shown in Figure 7A-F, when ER-SNR decreased from 10 dB to 5 dB and 2 dB, the classification and arrival-time accuracy of all three methods degraded: Baseline elevation and impulsive interference raised false alarms on non-event segments and introduced systematic delays in the picks. In contrast, SeisFormer consistently produced more temporally coherent and better-aligned P/S predictions on both datasets while maintaining a tighter temporal window—at 10 dB it nearly coincided with the annotations; at 5 dB it still stably covered the main energy of the events with markedly fewer false positives than EQTransformer and PhaseNet; and under the extreme 2 dB condition, although slightly contracted, its onset/offset remained broadly consistent with the labels, whereas the baselines exhibited fragmented or drifting predictions, leakage of energy into the baseline, or P/S confusion. Detailed metrics are presented in Table 2.

The STFT spectrograms in Figure 8 make the non-stationarity and narrowband harmonics, as well as their evolution with SNR, visually explicit, and empirically demonstrate that frame-wise adaptive windowing dynamically allocated time–frequency resolution: Under high noise it favored longer windows to enhance frequency resolution

and suppressed harmonics and low-frequency drift, whereas at higher SNR it adopted shorter windows to preserve onset transients—thereby highlighting the band-limited event energy even at low SNR. These visualizations provide direct and interpretable evidence for the robustness of the proposed time–frequency strategy. Overall, as noise intensifies, SeisFormer exhibited slower growth in false/missed detections and smaller arrival-time drift, indicating stronger noise resilience.

3.2.3. Multi-event scenario experiment

To assess the model's adaptability in complex signal conditions, we evaluated it on a dense-event window containing multiple consecutive P/S arrivals. As shown in Figures 9 and 10, the proposed model preserved clear P/S boundaries between adjacent events, with onsets and offsets closely matching the annotations. When inter-event intervals shortened and energy overlaps arose, it still robustly localized phase breakpoints and effectively suppressed cross-segment leakage. The spectrogram reveals that the model adaptively switches to shorter windows at rapid energy transitions to retain transient details, while favoring longer windows in regions with background undulations or strong narrowband interference to stabilize spectral structure. Consequently, in dense multi-event scenarios, the model achieved a favorable balance between arrival-time accuracy and noise robustness.

3.2.4. Ablation experiment

To quantify the contribution of each component to overall performance, we conducted ablation studies under realistic settings with a unified training/evaluation protocol (Table 3). Replacing Nyström attention with exact dot-product attention (without Nyström) reduced accuracy/F1 to 91.72%/91.93% and increased P/S arrival MAE to 6.36/7.71 ms, indicating that the Tikhonov-regularized pseudoinverse within the Nyström block

Table 2. Classification performance and arrival time calibration errors of the SeisFormer under different signal-to-noise ratios

Event-referenced signal-to-noise ratio	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	Mean P-wave arrival error	Mean S-wave arrival error
None	98.30	97.40	97.92	97.66	1.42 ms	2.29 ms
10dB	95.73	94.75	94.69	95.38	3.02 ms	5.37 ms
5dB	92.88	93.13	92.90	93.96	6.33 ms	8.41 ms
2dB	87.02	86.76	87.06	87.17	12.48 ms	18.49 ms

Table 3. Comparison of classification performance and arrival time calibration errors under different model configurations

Method	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	Mean P-wave arrival error (%)	Mean S-wave arrival error (%)
SeisFormer	98.30	97.40	97.92	97.66	1.42 ms	2.29 ms
Without Nyström	91.72	92.05	91.68	91.93	6.36 ms	7.71 ms
Frequency domain only	90.21	92.11	92.41	92.16	4.79 ms	5.34 ms
Time domain only	86.02	87.77	88.32	89.18	8.81 ms	9.05 ms
Without an adaptive window	93.82	92.19	90.71	90.30	10.21 ms	14.16 ms

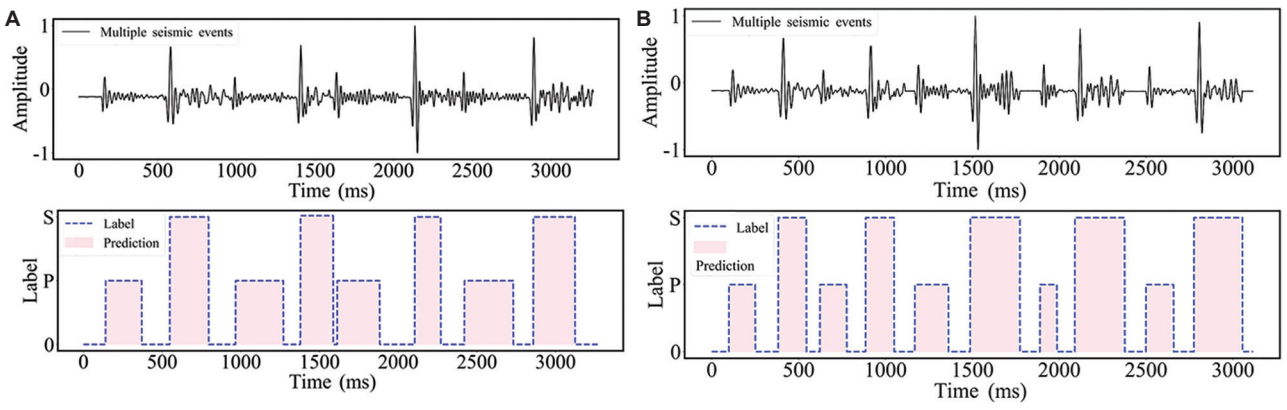


Figure 9. Classification results of the SeisFormer model in scenarios with multiple events occurring within a short time window. Each waveform segment has a duration of 3,000–4,000 ms.

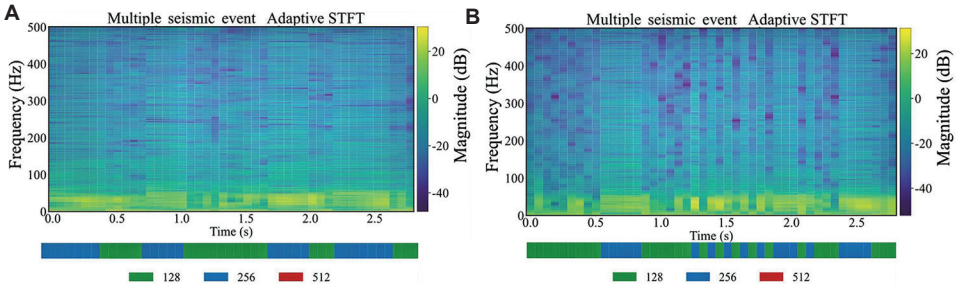


Figure 10. Adaptive short-time Fourier transform time–frequency spectrogram with frame-level window-selection indicator bars (corresponding to Figure 9)

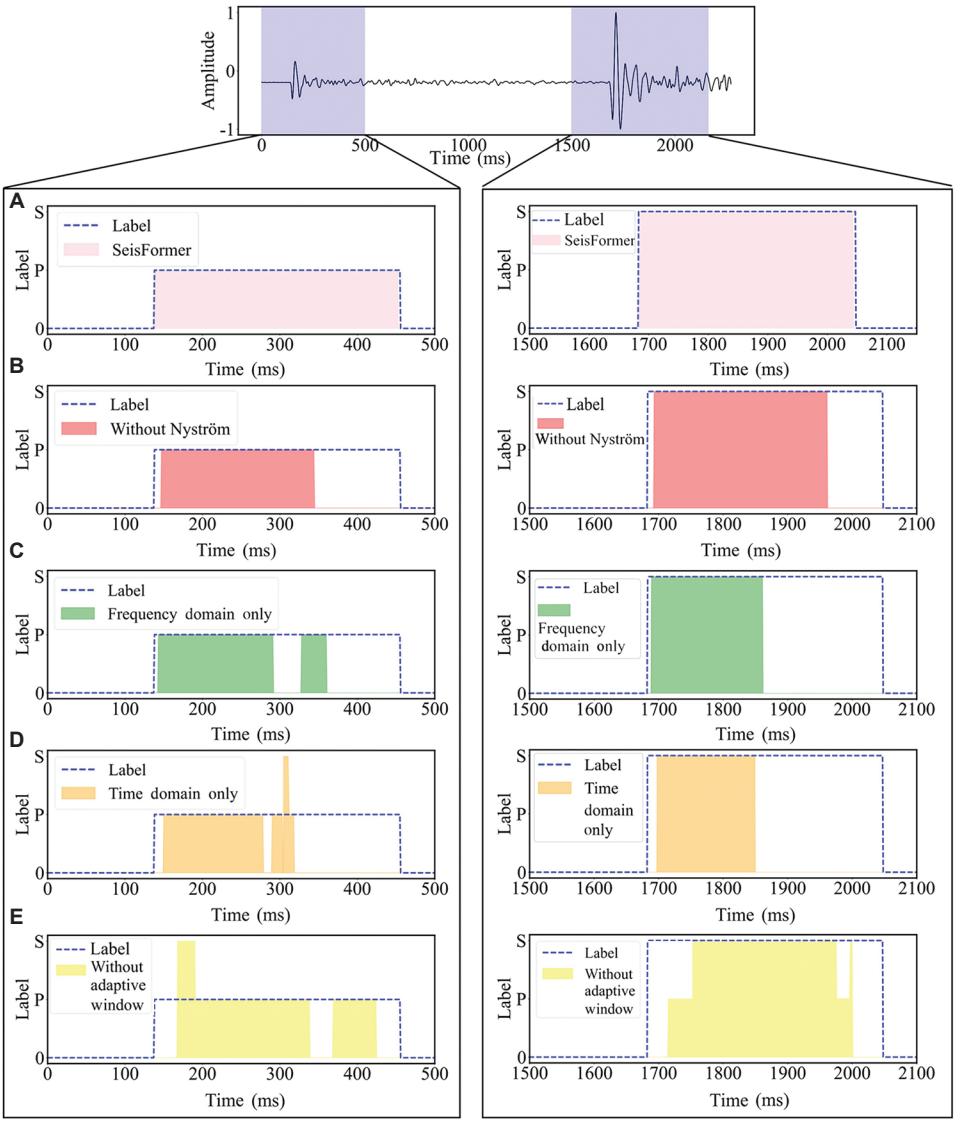


Figure 11. Comparison of classification and picking results across ablation variants. (A) SeisFormer; (B) without Nyström; (C) frequency domain only; (D) time domain only; and (E) without adaptive window.

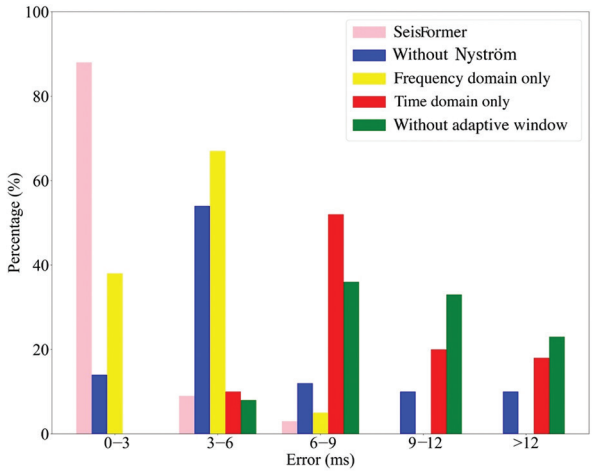


Figure 12. Distribution of average P-wave arrival time errors across different model configurations

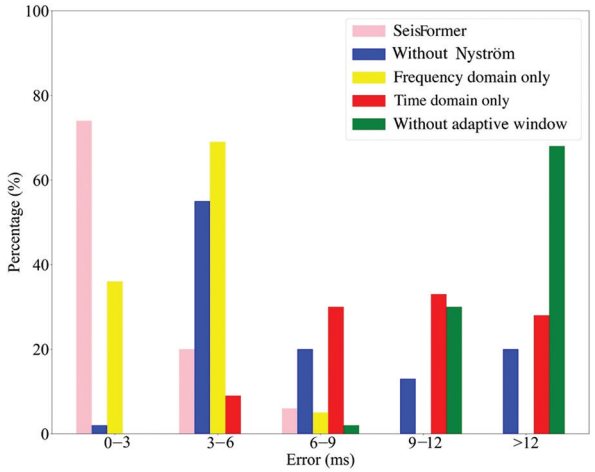


Figure 13. Distribution of average S-wave arrival time errors across different model configurations

(in the full model) constrained small-singular-value directions, suppressed noise amplification, and stabilized the weight distribution. Restricting the representation to a single domain markedly weakened modeling capacity: The frequency domain only variant, while closer to the full model, attained only 90.21%/92.16% (accuracy/F1) with MAE 4.79/5.34 ms; the time domain only variant further degraded to 86.02%/89.18% with MAE 8.81/9.05 ms, underscoring that a single domain cannot simultaneously capture transient onsets and band-limited structure—joint time–frequency modeling is critical for robust picking and classification. Removing the adaptive windowing mechanism (without adaptive windowing) still yielded 93.82% accuracy, but F1 dropped to 90.30% and arrival errors increased to 10.21/14.16 ms, demonstrating that frame-wise window selection, in combination with STFT/convolutional parameters, is essential for mitigating low-frequency drift and harmonic interference while preserving onset alignment.

As shown in Figure 11, the ablated variants exhibited more diffuse probability responses, window spillover, and larger onset drifts relative to the baseline. Figures 12 and 13 further corroborate this trend quantitatively: Per-trace error curves showed systematic increases in P/S arrival errors whenever a component was removed, with the largest growth observed without an adaptive window. Taken together, Nyström with a Tikhonov-regularized pseudoinverse + joint time–frequency representation + adaptive windowing acted synergistically: Adaptive windowing yielded the most significant gains in arrival-time precision, the regularized pseudoinverse secured numerical/training stability, and time–frequency complementarity set the upper bound and robustness of both classification and picking.

4. Conclusion

We proposed SeisFormer, a P/S-wave classification and first-arrival picking network for HDR hydraulic fracturing. SeisFormer combines adaptive multi-scale windowing with joint time–frequency modeling and introduces a stabilized Nyström attention module to enhance long-range dependency modeling and feature discriminability. Evaluated on a joint multi-site dataset constructed from HDR stimulation in the Qinghai Gonghe Basin and unconventional hydraulic fracturing in North China, SeisFormer achieved state-of-the-art performance on real data, noise-augmented data with non-stationary interference, and dense multi-event windows, demonstrating robustness across operating conditions and strong generalization. In field settings, the classification accuracy reached 98.30%, with mean arrival-time errors of 1.42 ms (P) and 2.29 ms (S). Under low SNR and complex signal environments, the model maintained stable classification and picking accuracy. Ablation studies further confirmed the significant contributions of the key components to overall performance.

Based on measured results on a NVIDIA GeForce RTX 4060 (8 GB) + Intel Core i9-13900HX platform—where the method attained P/S arrival-time errors ”.2.5 ms—future work can refine the unified pre-processing and end-to-end inference pipeline and conduct systematic robustness and fault-tolerance evaluations under complex, non-stationary noise, and dynamic operating conditions. In parallel, SeisFormer can be migrated to edge-computing modules and portable platforms to support near-real-time field monitoring and facilitate engineering deployment.

Acknowledgments

None.

Funding

This work was supported by the National Key R&D Program of China (Grant Nos. 2018YFB1501803 and 2019YFC1804805-4) and the China Geological Survey Project (Grant No. DD2019135).

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Mingjun Ouyang, Feng Sun
Formal analysis: Mingjun Ouyang
Investigation: Zenan Leng, Haotian Hu, Zubin Chen, Fa Zhao
Methodology: Mingjun Ouyang
Validation: Mingjun Ouyang
Visualization: Mingjun Ouyang
Writing–original draft: Mingjun Ouyang
Writing–review & editing: Mingjun Ouyang, Feng Sun

Availability of data

The microseismic dataset used in this study, obtained from hydraulic fracturing operations in HDR formations in the Qaidam Basin, Qinghai, China, is subject to a non-disclosure agreement and cannot be made publicly available.

References

1. Dong S, Jiao J, Zhou S, Lu P, Zeng Z. 3-D gravity data inversion based on enhanced dual U-Net framework. *IEEE Trans Geosci Remote Sens*. 2023;61:1-11.

   doi: 10.1109/TGRS.2023.3306980

2. Cloetingh S, Sternai P, Koptev A, *et al*. Coupled surface to deep Earth processes: Perspectives from TOPO-EUROPE with an emphasis on climate- and energy-related societal challenges. *Global Planet Change*. 2023;226:104140.

   doi: 10.1016/j.gloplacha.2023.104140

3. Ma W, Wang Y, Wu X, Liu G. Hot dry rock (HDR) hydraulic fracturing propagation and impact factors assessment via sensitivity indicator. *Renew Energy*. 2020;146:2716-2723.

   doi: 10.1016/j.renene.2019.08.097

4. Xie J, Cao H, Wang D, Peng S, Fu G, Zhu Z. A comparative study on the hydraulic fracture propagation behaviors in hot dry rock and shale formation with different structural discontinuities. *J Energy Eng*. 2022;148(6):04022040.

   doi: 10.1061/(asce)ey.1943-7897.0000856

5. Cheng Y, Zhang Y, Yu Z, Hu Z, Yang Y. An investigation on hydraulic fracturing characteristics in granite geothermal reservoir. *Eng Fract Mech*. 2020;237:107252.

   doi: 10.1016/j.engfracmech.2020.107252

6. Cheng Y, Zhang Y, Yu Z, Hu Z. Investigation on reservoir stimulation characteristics in hot dry rock geothermal formations of China during hydraulic fracturing. *Rock Mech Rock Eng*. 2021;54(8):3817-3845.

   doi: 10.1007/s00603-021-02506-y

7. Cheng Y, Zhang Y. Experimental study of fracture propagation: The application in energy mining. *Energies*. 2020;13(6):1411.

   doi: 10.3390/en13061411

8. Wang H, Alkhalifah T, Waheed UB, Birnie C. Data-driven microseismic event localization: An application to the Oklahoma Arkoma Basin hydraulic fracturing data. *IEEE Trans Geosci Remote Sens*. 2022;60:1-12.

   doi: 10.1109/TGRS.2021.3120546

9. Warpinski N. Microseismic monitoring: Inside and out. *J Petrol Technol*. 2009;61(11):80-85.

   doi: 10.2118/118537-JPT

10. Allen RV. Automatic earthquake recognition and timing from single traces. *Bull Seismol Soc Am*. 1978;68(5):1521-1532.

    doi: 10.1785/bssa0680051521

11. Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control*. 1974;19(6):716-723.

    doi: 10.1109/tac.1974.1100705

12. Allen R. Automatic phase pickers: Their present use and future prospects. *Bull Seismol Soc Am*. 1982;72(6B):S225-S242.

    doi: 10.1785/BSSA07206B0225

13. Álvarez I, Garcia L, Mota S, *et al*. An automatic P-phase picking algorithm based on adaptive multiband processing. *IEEE Geosci Remote Sens Lett*. 2013;10(6):1488-1492.

    doi: 10.1109/lgrs.2013.2260720

14. Earle PS, Shearer PM. Characterization of global seismograms using an automatic-picking algorithm. *Bull Seismol Soc Am*. 1994;84(2):366-376.

    doi: 10.1785/bssa0840020366

15. Akazawa TA. Technique for Automatic Detection of Onset Time of P- and S-Phases in Strong Motion Records. In: *Proceedings 13th Conference on Earthquake Engineering*. Vancouver, Canada. Paper No. 786; 2004.

16. Kurz JH, Grosse CU, Reinhardt HW. Strategies for reliable automatic onset time picking of acoustic emissions and of ultrasound signals in concrete. *Ultrasonics*. 2005;43(7):538-546.

    doi: 10.1016/j.ultras.2004.12.005

17. Anikiev D, Birnie C, Waheed UB, *et al*. Machine learning in microseismic monitoring. *Earth Sci Rev*. 2023;239:104371.

doi: 10.1016/j.earscirev.2023.104371

18. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436-444.

doi: 10.1038/nature14539

19. Mousavi SM, Beroza GC. Deep-learning seismology. *Science*. 2022;377:725-729.

doi: 10.1126/science.abm4470

20. Zhu W, Beroza GC. PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophys J Int*. 2019;216(1):261-273.

doi: 10.1093/gji/ggy423

21. He Z, Peng P, Wang L, Jiang Y. PickCapsNet: Capsule network for automatic P-wave arrival picking. *IEEE Geosci Remote Sens Lett*. 2021;18(4):617-621.

doi: 10.1109/lgrs.2020.2983196

22. Zhao, Y, Xu, H, Yang, T, Wang D, Sun D. A hybrid recognition model of microseismic signals for underground mining based on CNN and LSTM networks. *Geomat Nat Hazard Risk*. 2021;12(1):2803-2834.

doi: 10.1080/19475705.2021.1968043

23. Mousavi SM, Ellsworth WL, Zhu W, Chuang LY, Beroza GC. Earthquake transformer: An attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nat Commun*. 2020;11:3952.

doi: 10.1038/s41467-020-17591-w

24. Saad OM, Chen Y, Siervo D, *et al*. EQCCT: A production-ready earthquake detection and phase-picking method using the compact convolutional transformer. *IEEE Trans Geosci Remote Sens*. 2023;61:1-15.

doi: 10.1109/TGRS.2023.3319440

25. Li S, Yang X, Cao A, *et al*. SeisT: A foundational deep-learning model for earthquake monitoring tasks. *IEEE Trans Geosci Remote Sens*. 2024;62:1-15.

doi: 10.1109/tgrs.2024.3371503

26. Li XN, Chen FJ, Lai YP, Tang P, Liang X. ICAT-net: A lightweight neural network with optimized coordinate attention and transformer mechanisms for earthquake detection and phase picking. *J Supercomput*. 2025;81:191.

doi: 10.1007/s11227-024-06664-y

27. Peng P, Lei R, Wang JM. Enhancing microseismic signal classification in metal mines using transformer-based deep learning. *Sustainability*. 2023;15(20):14959.

doi: 10.3390/su152014959

28. Zhang X, Wang X, Zhang Z, Wang Z. CNN-transformer for microseismic signal classification. *Electronics*. 2023;12(11):2468.

doi: 10.3390/electronics12112468

29. Zhu L, Chuang L, McClellan JH, Liu E, Peng Z. *A Multi-Channel Approach for Automatic Microseismic Event Association using RANSAC-Based Arrival Time Event Clustering (RATEC)*. [arXiv Preprint]; 2017.

doi: 10.48550/arXiv.1702.01856

30. Li Z, Gou X, Jin W, Qin N. Frequency features of microseismic signals. *Chin J Geotech Eng*. 2008;30(6):830-834.

31. Aki K, Richards PG. *Quantitative Seismology*. 2nd ed. Sausalito, CA: University Science Books; 2002.

32. Ma C, Ran X, Xu W, *et al*. Fine classification method for massive microseismic signals based on short-time Fourier transform and deep learning. *Remote Sens*. 2023;15(2):502.

doi: 10.3390/rs15020502

33. Xiong Y, Zeng Z, Chakraborty R, *et al*. Nyströmformer: A Nyström-based algorithm for approximating self-attention. *Proc AAAI Conf Artif Intell*. 2021;35(16):14138-14148.

doi: 10.48550/arXiv.2102.03902

34. Piao X, Chen Z, Murayama T, *et al*. Fredformer: Frequency Debiased Transformer for Time Series Forecasting. In: *Proceedings 30th ACM SIGKDD Conference Knowledge Discovery and Data Mining (KDD 2024)*. Long Beach, CA, USA; 2024. p. 1234-1245.

doi: 10.1145/3637528.3671928

35. Soto-Quiros P. A fast method to estimate the Moore-Penrose inverse for well-determined numerical rank matrices based on Tikhonov regularization. *J Math Comput Sci*. 2024;37(1):59-81.

doi: 10.22436/jmcs.037.01.05

36. Longhas PRA, Abdul AM. Application of Tikhonov regularization in generalized inverse of adjacency matrix of undirected graph. *Int J Math Trends Technol*. 2022;68(2):1-6.

doi: 10.14445/22315373/ijmtt-v68i2p501

37. Gulliksson ME, Wedin PÅ, Wei Y. Perturbation identities for regularized Tikhonov inverses and weighted pseudoinverses. *BIT Numer Math*. 2000;40:513-523.

doi: 10.1023/A:1022319830134

38. Ross ZE, Meier MA, Hauksson E, Heaton TH. Generalized seismic phase detection with deep learning. *Bull Seismol Soc Am*. 2018;108(5A):2894-2901.

doi: 10.1785/0120180080

ACCSCIENCE PUBLISHING

Journal of Seismic Exploration

ARTICLE

Full waveform inversion for a long-wavelength velocity model using a regenerated wavefield based on the SWEET method

Seoje Jeong[1], Sumin Kim[2], Woohyun Son[3], and Wookeen Chung[2]*

[1]Department of Convergence Study on the Ocean Science and Technology, Ocean Science and Technology School, National Korea Maritime & Ocean University, Busan, Republic of Korea

[2]Department of Energy and Resources Engineering, National Korea Maritime & Ocean University, Busan, Republic of Korea

[3]Marine Geology and Energy Division, Korea Institute of Geoscience and Mineral Resources, Daejeon, Republic of Korea

Abstract

In full waveform inversion (FWI), long-wavelength velocity models are essential for accurately estimating subsurface physical parameters. However, building long-wavelength velocity models with low-frequency components is challenging due to mechanical limitations in seismic data acquisition. We propose a novel FWI method that utilizes a regenerated wavefield derived from the Suppressed Wave Equation Estimation of Traveltime (SWEET) algorithm. The regenerated wavefield in our approach was obtained by convolving the arbitrary source wavelet with a Green's function, which is represented by the first-arrival traveltime and amplitude extracted from the SWEET algorithm. Our approach can build long-wavelength velocity models, provided that a low-frequency wavelet is used. Furthermore, the potential for multi-scale inversion was demonstrated by gradually increasing the frequency of the source wavelet, leading to the acquisition of high-resolution models. In numerical examples, our proposed algorithm was validated using both synthetic and field data sets. We also assessed the noise sensitivity of the proposed method, confirming its applicability in practical scenarios. These results demonstrate that the proposed method is a robust and versatile tool for constructing long-wavelength and high-resolution velocity models from band-limited seismic data.

Keywords: Full waveform inversion; Long-wavelength velocity model; SWEET method; Regenerated wavefield; Multi-scale inversion

***Corresponding author:**
Wookeen Chung
(wkchung@kmou.ac.kr)

Citation: Jeong S, Kim S, Son W, Chung W. Full waveform inversion for a long-wavelength velocity model using a regenerated wavefield based on the SWEET method. *J Seismic Explor*. 2025;34(6):78-96.
doi: 10.36922/JSE025370071

Received: September 11, 2025

Revised: November 3, 2025

Accepted: November 10, 2025

Published online: December 2, 2025

1. Introduction

Full waveform inversion (FWI) estimates subsurface physical parameters by minimizing the misfit between observed and modeled seismic data.[1-3] However, successful FWI requires addressing nonlinearity, substantial computational demands, and dependence on an appropriate initial velocity model.[4-6] In particular, a reliable initial velocity is essential for avoiding cycle skipping and for robust convergence when applying FWI to field data.[7] One approach to improving the accuracy of the initial velocity model is to build long-wavelength velocity models and use them as the initial velocity model

for FWI. Many studies have proposed constructing long-wavelength velocity models, with first-arrival tomography being a commonly used technique,[8,9] and a reflection tomography approach[10] has typically been employed to generate initial velocity models for waveform inversion.[11-14]

Traveltime tomography has been developed to achieve progressively higher-resolution models starting from long-wavelength models. This method, which relies on ray-tracing theory and the high-frequency approximation, is less sensitive to the initial velocity model.[15,16] However, conventional traveltime tomography often falls short of the resolution requirements for FWI because it does not utilize phase and amplitude information.[17,18] Using the phase and traveltime of the first-arrival wavefield, Liu et al.[19] proposed the first-arrival phase-traveltime tomography (FPT). Nevertheless, FPT has a drawback in that it does not consider the amplitude of the first-arrival wavefield, making it difficult to accurately generate detailed structures.

Another approach involves transforming the observed and modeled seismic data to reconstruct the wavefield by utilizing specific waveform attributes, such as instantaneous phase, envelope, and Laplace-transformed data. For example, Shin and Cha[20] proposed a Laplace-domain waveform inversion to extract low-wavenumber components from high-frequency seismic data, thereby generating an initial velocity model for FWI. Although the Laplace domain, like tomography, does not consider phase information,[21] by utilizing a multi-scale inversion approach with varying damping factors during the Laplace transform, it is possible to build relatively high-resolution initial velocity models. Moreover, various studies have proposed building long-wavelength velocity models using envelope data derived through the application of the Hilbert transform. Similar to Laplace-domain waveform inversion, envelope inversion faces challenges in accurately resolving detailed subsurface structures due to the lack of phase information.[22,23] Notably, recent advancements have introduced methods that incorporate phase information during the envelope inversion process, thereby improving the resolution of the inverted models.[24] These approaches aim to overcome the lack of low-frequency components in the data used for generating long-wavelength velocity models.

Recently, several complementary strategies have been developed to improve FWI convergence. Reflection waveform inversion separates a smooth background model from short-wavelength reflectivity, allowing reflections to contribute low-wavenumber updates to the background model and thereby reducing dependence on very low frequencies in the recorded data.[25-27] In addition, low-frequency extrapolation methods, including global multi-scale deep networks, reconstruct sub-band energy from band-limited recordings.[28] In parallel, improved convexity or robust misfit functions and extended search spaces mitigate nonconvexity and reduce sensitivity to the background model.[29-31] Pretrained or self-supervised frameworks provide informative priors and warm-start initialization, which can be coupled with physics-based optimization.[32-34] Despite the progress, most long-wavelength building strategies remain limited when low frequencies are missing and when phase and amplitude along the first-arrival wavefield are not exploited.

Consequently, we introduce an approach that leverages traveltime and amplitude information extracted under high-damping conditions to reconstruct a reliable long-wavelength background model for FWI. The approach, referred to as FWI using a regenerated wavefield based on the Suppressed Wave Equation Estimation of Traveltime (SWEET) algorithm (FWI-RWS), utilizes the first-arrival traveltime and amplitude to regenerate a wavefield that contains low-frequency components. The SWEET algorithm is typically employed to derive traveltimes for refraction tomography. It uses the characteristics of the observed seismic data transformed under high-damping conditions in the Laplace domain to approximate the first-arrival time and amplitude. By integrating this method, FWI-RWS aims to enhance the accuracy of the long-wavelength velocity model and provide robust results for subsequent multi-scale FWI on field data.

To implement the proposed method, we regenerated the wavefield by convolving the first-arrival traveltimes and amplitudes obtained from each receiver with an arbitrary source wavelet. This process allows for the successful construction of long-wavelength velocity models by generating wavefields that contain low-frequency components. In addition, to effectively mitigate nonlinearity, a multi-scale method that starts the inversion at low frequencies and fits higher-frequency components sequentially was applied.[35] We demonstrated through preliminary tests that multi-scale inversion is possible depending on the frequency characteristics of the convolved source wavelet, thereby enabling the stable acquisition of long-wavelength models.

The theory and methodology of FWI with the SWEET algorithm are presented, and the effectiveness of the proposed method is demonstrated using synthetic data. We also discuss the robustness of the proposed method in noisy environments, which is a known limitation of the SWEET algorithm. Finally, field data results are presented to verify the applicability of FWI-RWS.

2. Methodology

2.1. Regenerated wavefield based on the SWEET method

In this section, the basic theory of the regenerated wavefield based on the SWEET method is reviewed. In the time domain, a wavefield u can be approximated as a series of spikes.[36]

$$u(t)=\sum_{q}A_{q}\delta\left(t-t_{q}\right) \tag{I}$$

where A_q and t_q represent the amplitude and q-th time step (counted from the first-arrival event), respectively, and δ is the Dirac delta function. By multiplying **Equation (I)** by an exponential damping factor e^{-st} and integrating it, the wavefield $\tilde{u}$ in the Laplace domain can be written as follows:[36,37]

$$\tilde{u}(s)=\int_{0}^{\infty}u(t)e^{-st}dt\approx A_{f}e^{-st_{f}} \tag{II}$$

where s is a positive real number known as the Laplace damping constant, t_f is the first-arrival traveltime, and A_f is the amplitude at the first-arrival traveltime. By taking the derivative of **Equation (II)** with respect to s and dividing it by $\tilde{u}$, the equation becomes:

$$\frac{\partial\tilde{u}(s)}{\partial s}=-t_{f}A_{f}e^{-st_{f}}=-t_{f}\tilde{u}(s) \tag{III}$$

Therefore, the first-arrival traveltime and amplitude are defined as follows:

$$t_{f}=-\frac{1}{\tilde{u}(s)}\frac{\partial\tilde{u}(s)}{\partial s} \tag{IV}$$

and

$$A_{f}=\tilde{u}(s)e^{st_{f}} \tag{V}$$

In this study, the wavefield in the time domain was regenerated by convolving the traveltime and amplitude, calculated from **Equations (IV)** and **(V)**, with an arbitrary source wavelet.

The regenerated wavefield $\bar{u}$ is given by:

$$\bar{u}(t)=w(t)*\left[A_{f}\delta\left(t-t_{f}\right)\right] \tag{VI}$$

where w is the source wavelet, and $*$ is the convolution operator.

2.2. FWI using a regenerated wavefield based on the SWEET method (FWI-RWS)

In its classical formulation, FWI is defined as an optimization problem that seeks to minimize an objective function, and FWI-RWS has a similar form to FWI. The objective function E, based on the l_2-norm of residuals between the observed and modeled data, can be expressed as[2]:

$$E=\frac{1}{2}\sum_{i}^{N_s}\sum_{j}^{N_r}\left\|\bar{u}_{i,j}-\bar{d}_{i,j}\right\|_{2}^{2} \tag{VII}$$

where $\|\ldots\|_2^2$ is the l_2-norm, $\bar{u}_{i,j}$ and $\bar{d}_{i,j}$ are the regenerated traces both for modeled and observed traces using **Equation (VI)**, i and j indicating the source and the receiver locations, and N_s and N_r are the number of sources and receivers, respectively. The gradient with respect to the k-th subsurface model parameter m_k can be expressed as:

$$\nabla E=\frac{\partial E}{\partial m_{k}}=\sum_{i}^{N_s}\sum_{j}^{N_r}\left(\frac{\partial\bar{u}_{i,j}}{\partial m_{k}}\right)^{T}\cdot\bar{r}_{i,j} \tag{VIII}$$

where $\frac{\partial\bar{u}_{i,j}}{\partial m_k}$ is the partial derivative wavefield using the regenerated wavefield, $\bar{r}_{i,j}$ is the data residual calculated by $\bar{u}_{i,j}-\bar{d}_{i,j}$, T is transpose, and $\cdot$ is the dot product operator.

In conventional FWI, as calculating **Equation (VIII)** leads to substantial computational cost, the adjoint-state method is used to reduce the computational burden associated with the partial derivative wavefield.[38] Similarly, we aim to apply the adjoint-state method to the proposed approach. By substituting **Equation (VI)** into **Equation (VIII)**, the partial derivative wavefield can be expressed as follows:

$$\begin{aligned}\frac{\partial\bar{u}_{i,j}}{\partial m_{k}}&\approx\frac{\bar{u}_{i,j}\left(m_{k}+\Delta m\right)-\bar{u}_{i,j}\left(m_{k}\right)}{\Delta m}\\&=w_{i}(t)*\left(\frac{A_{f}^{p}-A_{f}^{b}}{\Delta m}\right)\delta\left(t-t_{f}\right)\\&=w_{i}(t)*\left(\frac{A_{f}^{p}-A_{f}^{b}}{\Delta m}\right)\delta\left(t-t_{f}\right)\end{aligned} \tag{IX}$$

with

$$\bar{u}_{i,j}\left(m_{k}+\Delta m\right)=w_{i}(t)*\left[A_{f}^{p}\delta\left(t-t_{f}\right)\right] \tag{X}$$

$$\bar{u}_{i,j}\left(m_{k}\right)=w_{i}(t)*\left[A_{f}^{b}\delta\left(t-t_{f}\right)\right] \tag{XI}$$

where f_i is the source wavelet, indicating the use of the i-th source wavelet. In **Equation (IX)**, the first-arrival traveltime t_f remains the same regardless of Δm, while A_f^p and A_f^b denote the first-arrival amplitudes computed from the perturbed model $m_k+\Delta m$ and background model m_k, respectively. Therefore, the partial derivative wavefield can be further simplified as follows:

$$\Delta A_f = A_f^p - A_f^b \frac{\partial \overline{\mathrm{u}}_{i,j}}{\partial m_k} = w_i(t) * \frac{\Delta A_f}{\Delta m}\delta(t-t_f) = \left(\frac{\mathrm{u}_{i,j}(m_k+\Delta m)-\mathrm{u}_{i,j}(m_k)}{\Delta m}\right) = \left(\overline{\frac{\partial \mathrm{u}_{i,j}}{\partial m_k}}\right) \quad \text{(XII)}$$

Through **Equations (VIII), (IX),** and **(XII),** we can confirm that applying **Equation (VI)** to the partial derivative wavefield yields results identical to the regenerated wavefield. To verify **Equation (XII),** we used a homogeneous velocity model of 1.5 km/s on a 201 × 101 grid (Figure 1A). The grid size was 10 m and the time interval was 1 ms. Figure 1A shows a single shot at 800 m and a single receiver at 1,200 m, both positioned on the surface. The point-scatterer was located at (1,000 m, 500 m) at the center of the velocity model. The source signature was a Ricker wavelet with a dominant frequency of 5 Hz. The partial derivative wavefield at the point-scatterer within a simple subsurface model was compared with results obtained from numerical differentiation. In Figure 1B, the comparison of the results from convolving the same sources with **Equation (XII)** indicates excellent agreement between the two approaches.

To reduce the computational burden of FWI-RWS, the adjoint-state method was applied. The partial derivative wavefield for the regenerated wavefield in **Equation (VIII)** can be expressed as follows:

$$\frac{\partial \overline{\mathrm{u}}_{i,j}}{\partial m_k} = \left(\overline{\frac{\partial \mathrm{u}_{i,j}}{\partial m_k}}\right) = \mathrm{G} * \overline{\mathrm{v}}_{i,j} \quad \text{(XIII)}$$

where G is Green's operator, $\overline{\frac{\partial \mathrm{u}_{i,j}}{\partial m_k}}$ is the regenerated partial derivative wavefield, and $\overline{\mathrm{v}}_{i,j}$ is the regenerated virtual source wavefield. To verify **Equation (XIII),** the gradient to a perturbation point within the subsurface model (Figure 1A) was examined. This yielded two distinct gradient estimates: one calculated using the finite-difference method and the other obtained using the virtual source. Figure 2 shows that the two gradients closely match. By substituting **Equation (XIII)** into **Equation (VIII),** the gradient using the regenerated wavefield can be expressed as follows:

$$\begin{aligned}
&\nabla E = \\
&\sum_i^{N_s}\sum_j^{N_r}\int_0^{\mathrm{T_{max}}}\left(\frac{\partial \overline{u}_{i,j}(t)}{\partial m_k}\right)\Delta \overline{r}_{i,j}(t)dt = \sum_i^{N_s}\sum_j^{N_r}\left(\frac{\partial \overline{u}_{i,j}(t)}{\partial m_k}\right)\otimes \overline{r}_{i,j}(t) \\
&= \sum_i^{N_s}\sum_j^{N_r}\left[G(t) * \overline{v}_{i,j}(t)\right]\otimes \overline{r}_{i,j}(t) = \sum_i^{N_s}\sum_j^{N_r}\overline{v}_{i,j}(t) \\
&*\left[G(t) * \overline{r}_{i,j}(\mathrm{T_{max}}-t)\right] \\
&= \sum_i^{N_s}\sum_j^{N_r}\overline{v}_{i,j}(t) * B_{i,j}(t) = \sum_i^{N_s}\sum_j^{N_r}\overline{v}_{i,j}(t)\otimes B_{i,j}(\mathrm{T_{max}}-t) \\
&= \sum_i^{N_s}\sum_j^{N_r}\int_0^{\mathrm{T_{max}}}\overline{v}_{i,j}(t)B_{i,j}(\mathrm{T_{max}}-t)dt \text{ with } B_{i,j} \\
&(t) = G(t) * \overline{r}_{i,j}(\mathrm{T_{max}}-t)
\end{aligned} \quad \text{(XIV)}$$

where $B_{i,j}$ is the backpropagated wavefield with the regenerated residual wavefield, and ⊗ denotes the zero-lag cross-correlation.

Based on the steepest-descent method, the model update at each iteration is expressed as[38]:

$$\mathrm{m}^{l+1} = \mathrm{m}^l - \alpha \times \Delta \mathrm{m} \quad \text{(XV)}$$

where m^l is the subsurface model parameter for l-th iteration, α is the step length, and Δm is the update direction.

In this study, the diagonal term of a pseudo-Hessian was used as a preconditioner for the gradient[21]. Furthermore, to mitigate model dependence in seismic imaging, a multi-scale strategy was employed, applying the FWI method across a spectrum of frequencies, progressing from low-frequency to high-frequency components.[39] The conventional multi-scale FWI approach enhances resolution by progressively incorporating the low-frequency to high-frequency components of the observed data.[1-3] Overall, a multi-scale FWI-RWS approach was introduced by gradually increasing the maximum frequency of the source wavelet used in **Equation (VI)**. Figure 3 shows the workflow of multi-scale FWI-RWS, which can be divided into eight steps:

(i) Set initial velocity model parameters and source wavelet for low frequencies.
(ii) Perform forward propagation modeling.
(iii) Extract the traveltime and amplitude of the shot gathers.
(iv) Convolve the arbitrary source wavelet with the extracted traveltime and amplitude to regenerate the wavefield.
(v) Use the calculated regenerated wavefield, virtual source, and residual to perform backward propagation modeling.

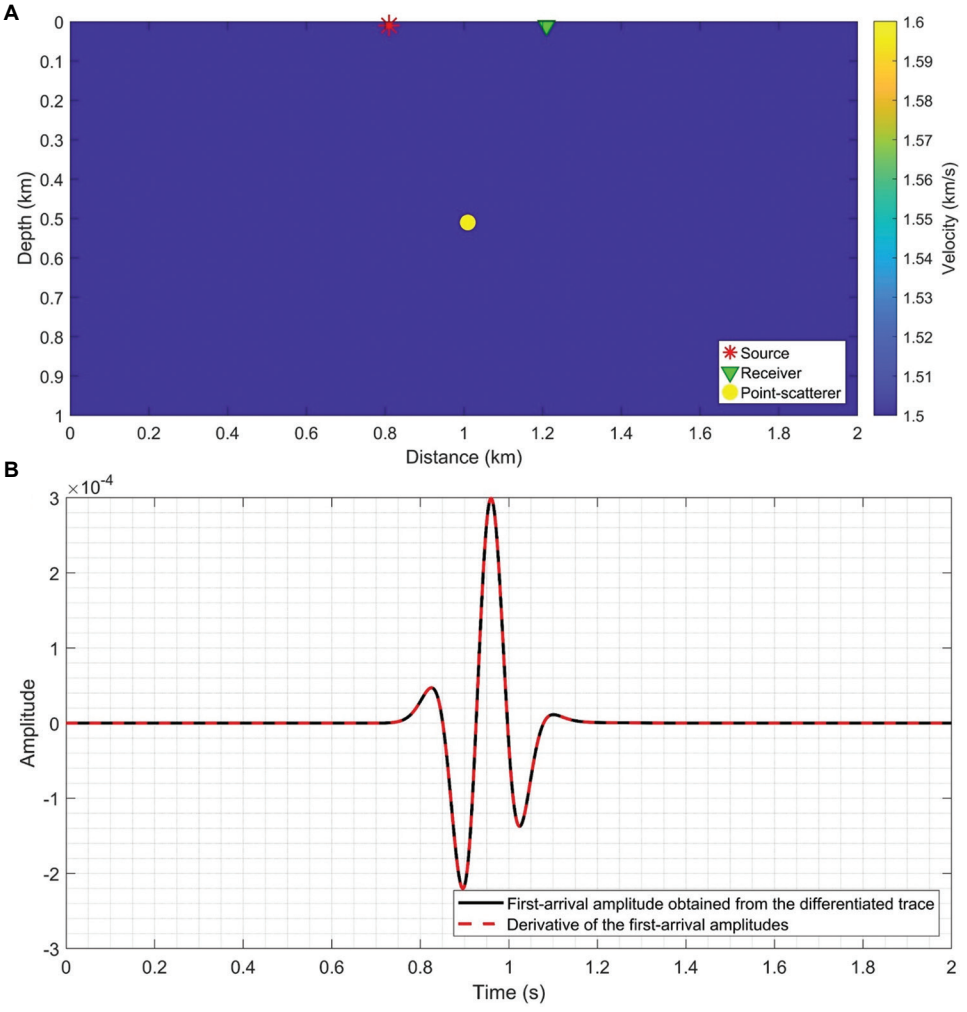


Figure 1. Geometry of the homogeneous model and validation of **Equation [XII].** (A) The homogeneous velocity model illustrates the source, receiver, and point-scatterer locations. (B) Comparison of results obtained by convolving the same source wavelet with the first-arrival amplitude obtained from the differentiated trace (solid black line) and the derivative of the first-arrival amplitudes from the point-scatterer and background models (dashed red line).

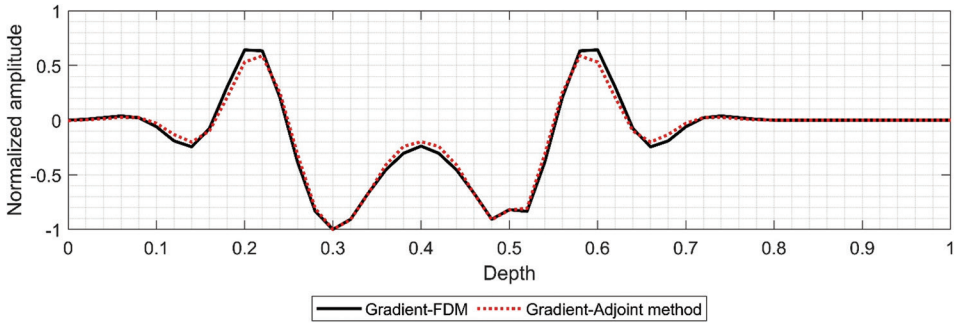


Figure 2. A comparison of the gradient between the finite-difference method (**Equation [IX]**; solid black line) and the adjoint formulation (**Equation [XIII]**; dashed red line)
Abbreviation: FDM: Finite-difference method.

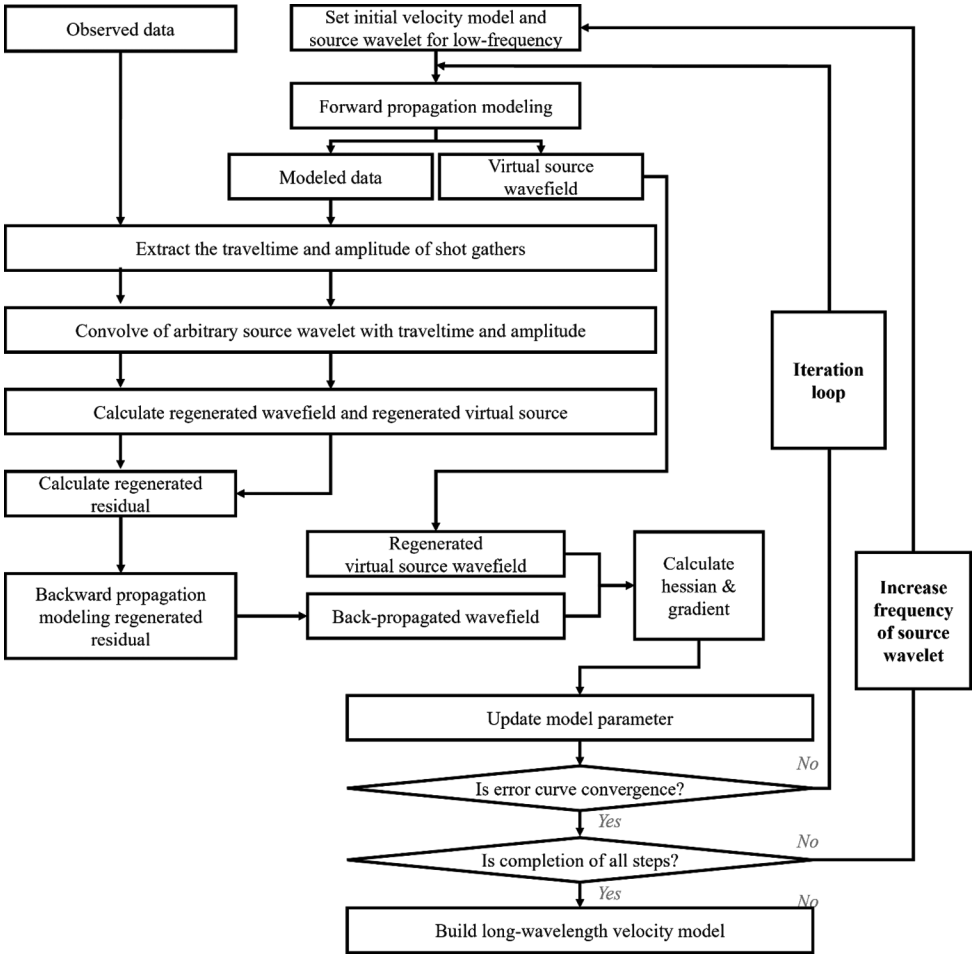


Figure 3. Workflows of multi-scale FWI-RWS

(vi) Update the velocity model using the regenerated wavefield.
(vii) Check the error curve convergence and increase the frequency of the source wavelet.
(viii) Build a long-wavelength velocity model.

3. Results

3.1. Analysis of the regenerated wavefield and gradient using a three-layer velocity model

In the numerical tests, we solved the 2D isotropic, constant-density acoustic wave equation in the time domain. The propagator employed second-order central differences in space and a second-order three-level time scheme. To analyze how the maximum frequency of the arbitrary source wavelet in **Equation (VI)** affects the shot gathers and the corresponding gradients, we designed a synthetic experiment using a three-layer model (Figure 4). The initial velocity model was homogeneous with a velocity of 1.5 km/s. The top boundary was treated as a free surface boundary, while the remaining sides employed convolutional perfectly matched layers.[40] The model grid comprised 601 × 201 points with 10.0 m spacing in both directions. The time sampling interval was 1 ms, and the recording length was 5.0 s. A total of 57 shots were deployed from x = 0.2 to x = 5.8 km at 100 m spacing. For each shot, 601 receivers were distributed along the velocity model surface, with a receiver interval of 10 m. The source signature was a Ricker wavelet with a dominant frequency of 5 Hz. Shot gathers were generated by propagating wavefields induced by the arbitrary source wavelet and subsequently used to compute and analyze the gradients within the multi-scale FWI-RWS workflow.

Figure 5A-C shows the modeled data, observed data, and residuals obtained by forward modeling. Figure 5D-L illustrates the regenerated modeled data, observed data, and residuals after convolution with Ricker wavelets at three different maximum frequencies: 15 Hz (Figure 5D-F), 1 Hz (Figure 5G-I), and 0.5 Hz (Figure 5J-L). As frequency decreased, the observed and modeled data became smoother, and the residuals grew broader and less detailed.

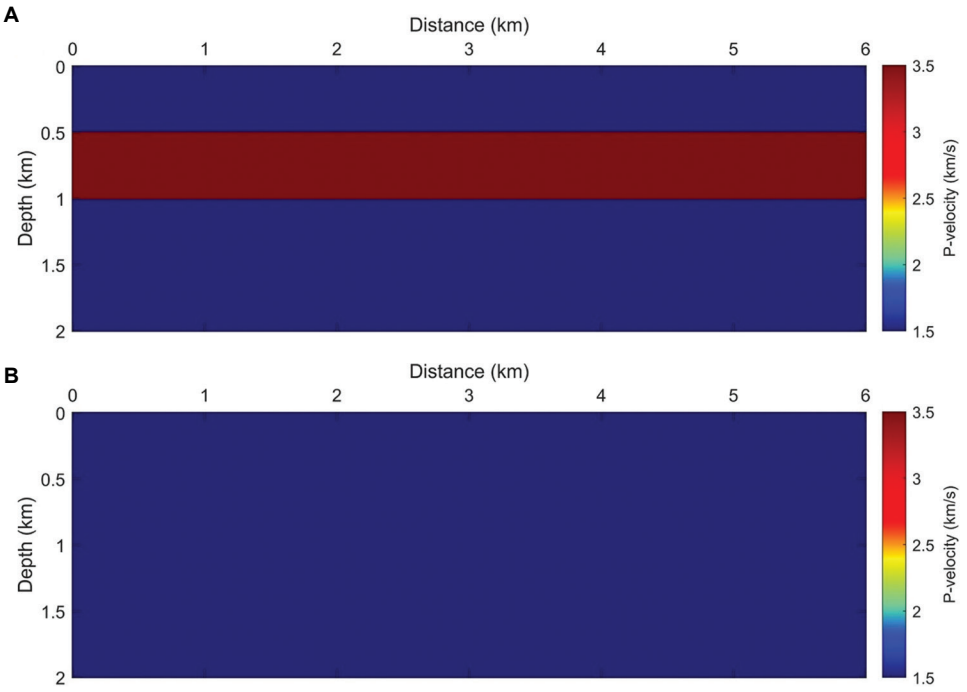


Figure 4. Test velocity models. (A) The three-layer velocity model. (B) The initial velocity model.

Based on **Equations (IV)** and **(V)**, the first-arrival traveltime and amplitude of the wavefield, as shown in Figure 6, were obtained with a Laplace damping constant of 15. As the maximum frequency of the convolved source decreased, the regenerated wavefield contained more low-frequency information.

In FWI-RWS, the gradient was calculated by regenerating the virtual source and the backward-propagated wavefields. To compare the inversion capabilities of conventional FWI and FWI-RWS on the three-layer model, we examined gradients at various frequencies: the gradient of FWI (Figure 7A) and the gradient of FWI-RWS using maximum frequencies of 15 Hz (Figure 7B), 1 Hz (Figure 7C), and 0.5 Hz (Figure 7D). The gradient of conventional FWI significantly updated the upper part of the high-velocity layer. In contrast, FWI-RWS, utilizing arbitrary source wavelets at low frequencies, updated the lower and internal parts of the high-velocity layer.

3.2. FWI-RWS using the SEG/EAGE salt model

The FWI-RWS was tested using a 2D section of the Society of Exploration Geophysicists/European Association of Geoscientists and Engineers (SEG/EAGE) salt model.[41] The model size was 676 × 210 points with 10.0 m spacing (Figure 8A). The synthetic dataset consisted of 186 shots, with a shot interval of 400 m. For each shot, 676 receivers were distributed along the velocity model surface, with a receiver interval of 10 m. The data recording time was 5.0 s, with a time interval of 1.0 ms. The source signature was a Ricker wavelet with a dominant frequency of 5 Hz. The initial velocity model was a linear model, with a velocity range of 1.5 km/s at the surface to 2.5 km/s at the bottom, as shown in Figure 8B. The FWI-RWS algorithm used a Laplace damping constant of 15 to extract first-arrival traveltime and amplitude.

In this case, the multi-scale FWI-RWS was performed in three steps to update the velocity model, where each step involved repeating the process across a specified frequency band. It progressively increased the maximum frequency of the source wavelet used in **Equation (VI)** from 0.05 Hz to 1 Hz, with a frequency interval of 0.05 Hz. Figure 8C shows the inverted velocity model for the 103rd iteration. The FWI-RWS results enabled the updating of broad regions of the salt body and the acquisition of a more reliable initial velocity model. When the iteration termination ε is satisfied:

$$\varepsilon < \frac{E^{l} - E^{l-1}}{E^{1}} \tag{XVI}$$

where E^{l} represents the evaluated objective function at the l-th iteration. In our test, we set ε to 0.02. The updated images for FWI-RWS from these inversion results (Figure 8C) demonstrated the potential of a multi-scale strategy. The inversion quality improved by iteratively processing the inversion from low to high frequencies of the source. Then, multi-scale FWI-RWS was used to improve

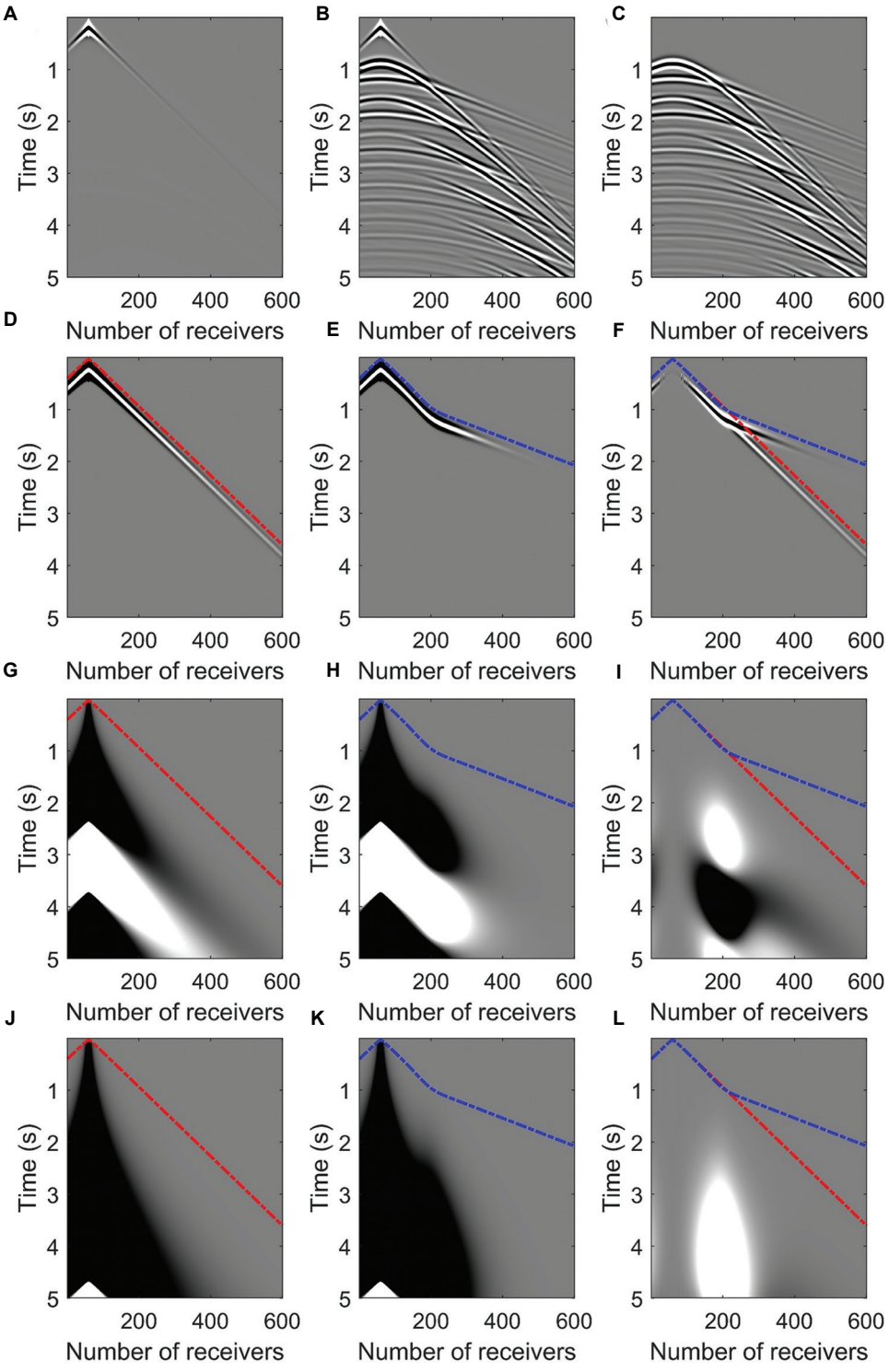


Figure 5. Comparison of the modeled data, observed data, and residuals. (A–C) Original wavefield. Regenerated wavefield using a maximum frequency of (D–F) 15 Hz, (G–I) 1 Hz, and (J–L) 0.5 Hz. The red dashed lines denote the first-arrival traveltime of the modeled data, and the blue dashed lines denote the first-arrival traveltime of the observed data.

the resolution of inverted salt structures. To evaluate the dependency of FWI results on the initial velocity model, FWI results obtained using a linear velocity model (Figure 8B) were compared with those obtained using the

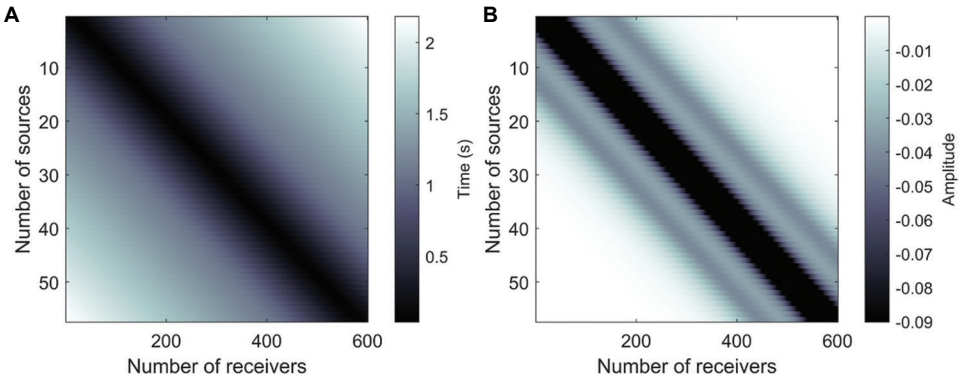


Figure 6. Spatial patterns of first-arrival traveltime and amplitude across sources and receivers. (A) First-arrival traveltime, and (B) amplitude.

FWI-RWS results (Figure 8C). The model parameters used for modeling and FWI were the same as those used in the previous example.

After 200 iterations, the FWI results using the linear velocity model as the initial velocity model only recovered the shallow part of the salt due to a lack of low-frequency information (Figure 9A). It is difficult to obtain the reflection information of the salt bottom and subsalt when there is insufficient low-frequency information. However, FWI using FWI-RWS results as the initial velocity model showed improvement with the same number of iterations (Figure 9B), showing the effectiveness of the proposed method. The proposed method allows for estimating long-wavelength components when using low-frequency source wavelets, and, as the frequency increases, it progressively recovers finer details, such as the salt boundaries shown in Figure 9B.

The analysis of the depth–velocity profiles (Figure 10) demonstrated that the proposed algorithm not only updates the model to closely resemble the true velocity but also accurately adjusts both the boundaries and internal velocities of the high-velocity layers. Figure 11 shows the history of root mean square error (RMSE) with respect to the number of iterations to assess convergence. The RMSE values were normalized by dividing the error obtained from the first iteration by the total RMSE value. As shown in Figure 11, we noted that the RMSE of the FWI using FWI-RWS results reached 0.2 of the initial value, whereas the RMSE of the FWI using the linear velocity model reached over 0.6 compared to the initial result.

3.3. Field data test

The application of the multi-scale FWI-RWS strategy was further extended to a 2D marine profile to evaluate its robustness. The 2D active seismic dataset was acquired offshore Yeosu in the South Sea by the Korea Institute of Geoscience and Mineral Resources. Seismic data acquisition was performed using a 1,050 m streamer and an airgun source. Additional acquisition parameters are shown in Table 1.

Field seismic data were first conditioned by estimating an effective source wavelet from high-signal-to-noise ratio (SNR) early arrivals. We then applied an eighth-order, zero-phase Butterworth low-pass filter (80 Hz cutoff) to limit bandwidth and resampled the data from 0.5 ms to 0.25 ms to match the modeling grid. Finally, to stabilize the multi-scale FWI-RWS updates and to comply with the regenerated-wavefield formulation, the records were convolved with a 40 Hz dominant Ricker wavelet to achieve a controlled low-frequency target wavelet. All steps were applied consistently to all shots.

Finite-difference modeling used a 0.25 ms time step with second-order accuracy in space and time. The initial velocity model was a homogeneous model with a velocity of 1.5 km/s. The model size was 8,951 × 151 grid points with 1.25 m spacing in both directions. The towed streamer had 84 channels at a 12.5 m group interval, towed at 7 ± 1 m depth, and shots were spaced 12.5 m apart with 3 s records sampled at 1 ms.

Figure 12 presents the estimated source and the Ricker wavelets (with maximum frequencies of 1, 5, and 10 Hz) used for constructing the long-wavelength velocity model. Figure 13A showcases the results of the FWI-RWS, demonstrating the construction of a long-wavelength velocity model. To achieve this, wavefields were regenerated using arbitrary sources with maximum frequency components ranging from 1 Hz to 10 Hz, incremented at 1 Hz intervals. Then, the applicability of the proposed method to field data was evaluated by applying reverse time migration (RTM) with different initial velocity models, including a homogeneous velocity model and a model derived from the FWI-RWS method. Figure 13B and C shows migration images from RTM obtained using a homogeneous model and FWI-RWS

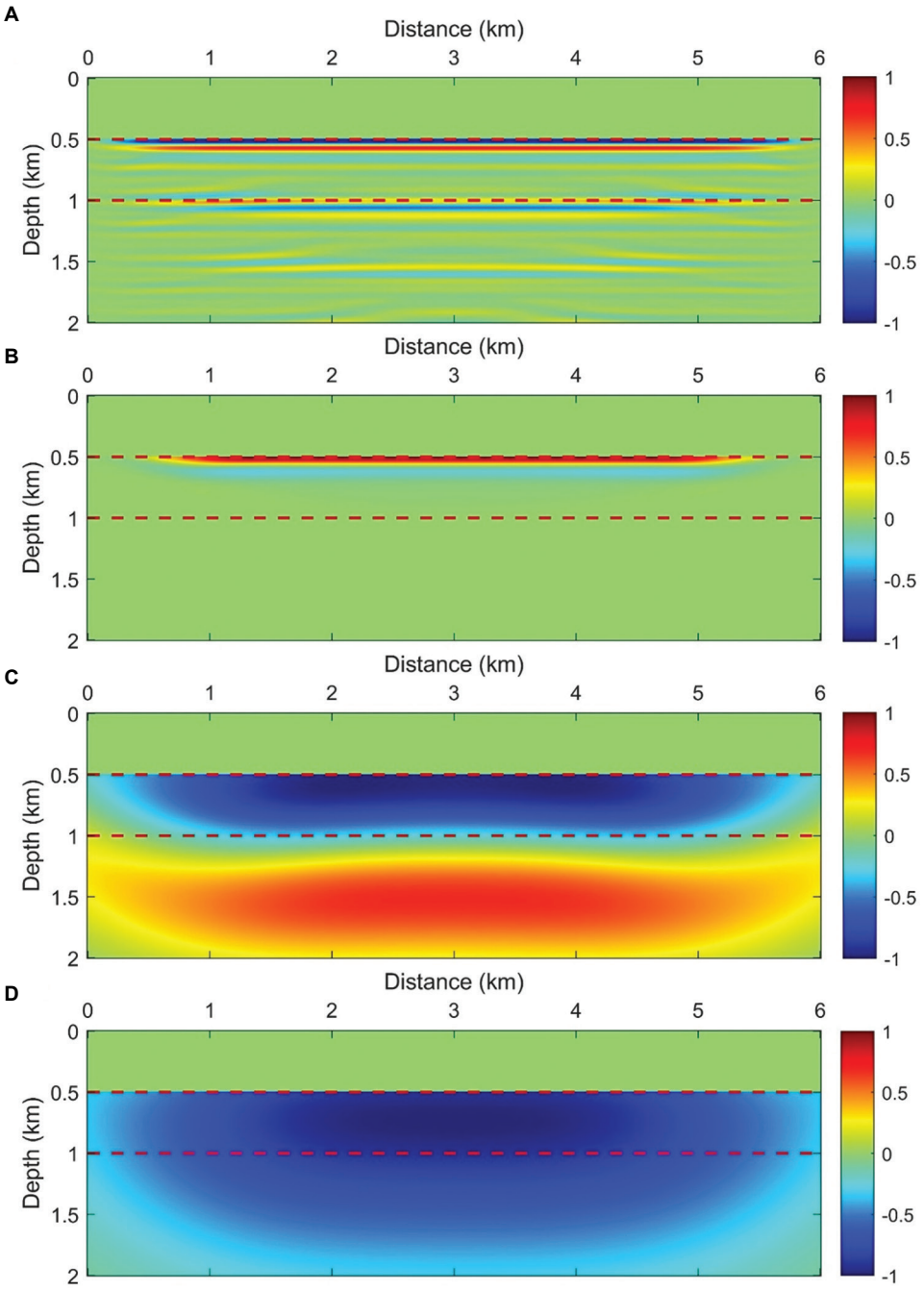


Figure 7. Gradients of (A) FWI using a maximum frequency of 15 Hz, FWI-RWS using a maximum frequency of (B) 15 Hz, (C) 1 Hz, or (D) 0.5 Hz. The red dashed lines denote the first-arrival traveltime of the modeled data, and the blue dashed lines denote the first-arrival traveltime of the observed data.

results. The RTM images generated using the FWI-RWS results (Figure 13C) showed improved amplitude balance compared to those using a homogeneous model. Compared with the RTM results using a homogeneous model, this approach also produced more accurate and high-resolution images with better continuity of subsurface structures.

For an additional analysis, the subsurface offset domain common image gather (SODCIG) obtained by RTM using the homogeneous model and the FWI-RWS results as initial velocity models were compared. Common image gathers (CIGs) are generally used as a primary criterion for validating the velocity model, with moveout guiding

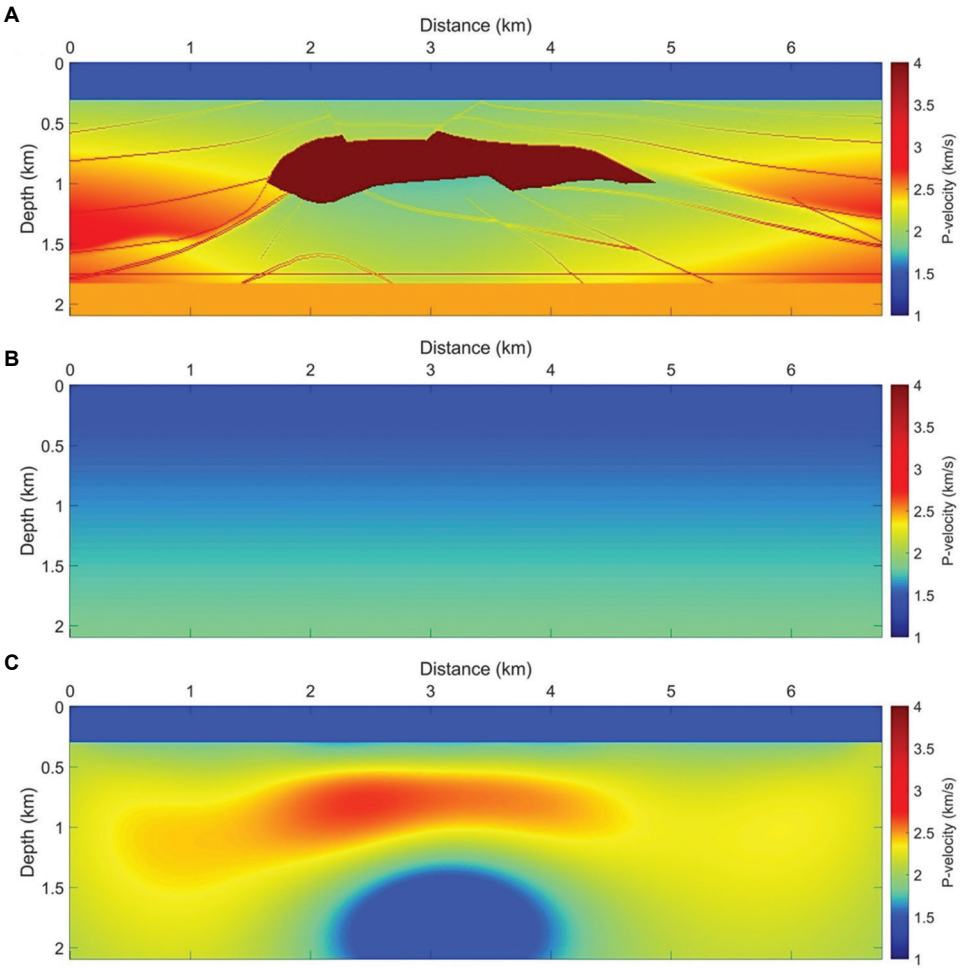


Figure 8. The 2D SEG/EAGE salt model. (A) True, (B) initial velocity model, and (C) long wavelength model of FWI-RWS results.
Abbreviation: SEG/EAGE: Society of Exploration Geophysicists/European Association of Geoscientists and Engineers.

Table 1. Seismic data acquisition parameters

Parameters	Description	Value
Streamer	Streamer length (m)	1,050
	Number of channels	84
	Group interval (m)	12.5
	Streamer depth (m)	7±1
Recording	Recording length (s)	3
	Sampling rate (ms)	1
Source	Source type	Bolt long-life air guns
	Source volume (cu. in.)	269
	Source depth (m)	5
	Source interval (m)	12.5

the velocity updates.[42-45] One type of CIG, known as the SODCIG, is characterized by accurately migrated transmitted events that are vertically aligned at zero-offset along the depth axis, while any defocusing indicates inaccuracies in the migration velocity.[46-49] Figure 14 presents SODCIGs at depths of 1, 2, and 5 km in the migration image. When using a homogeneous velocity model, most reflectors shift upward as the offset increases, indicating that the background velocity is slower than the actual velocity. In contrast, the FWI-RWS velocity model focuses the reflection energy near zero offset, though some reflectors still exhibit curvature with increasing offset. These results suggest that using the FWI-RWS model as the background velocity enhances the recovery of long-wavelength structures, leading to more accurate alignment of reflectors than using the homogeneous velocity model.

4. Discussion

When estimating the first-arrival traveltime, the Laplace transformation is sensitive to the first-arrival noise, as noted by Shin and Cha.[20] Therefore, if noise precedes the

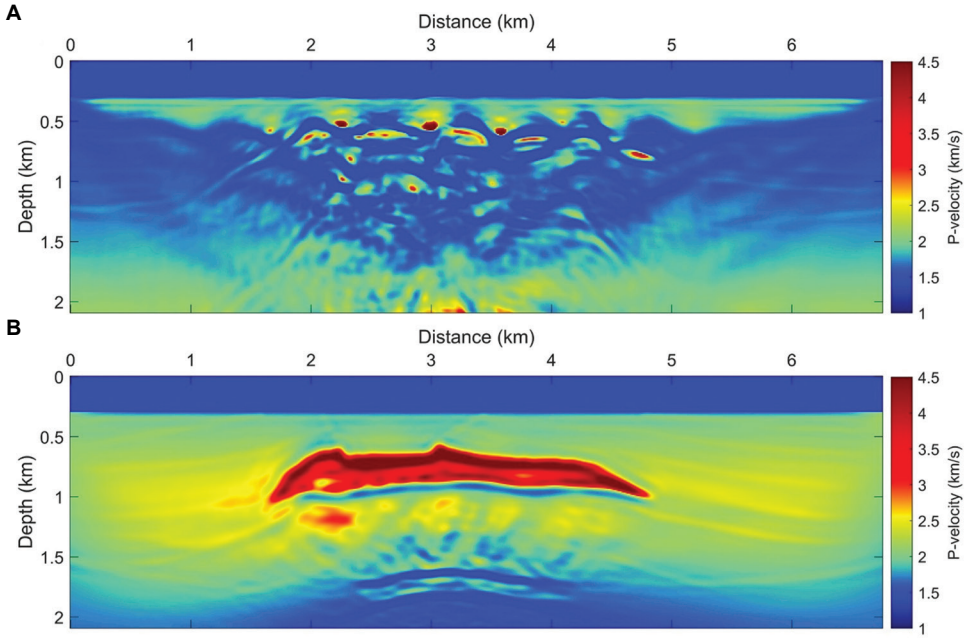


Figure 9. Full waveform inversion (FWI) results using (A) linear velocity model, and (B) FWI-RWS result as initial velocity model

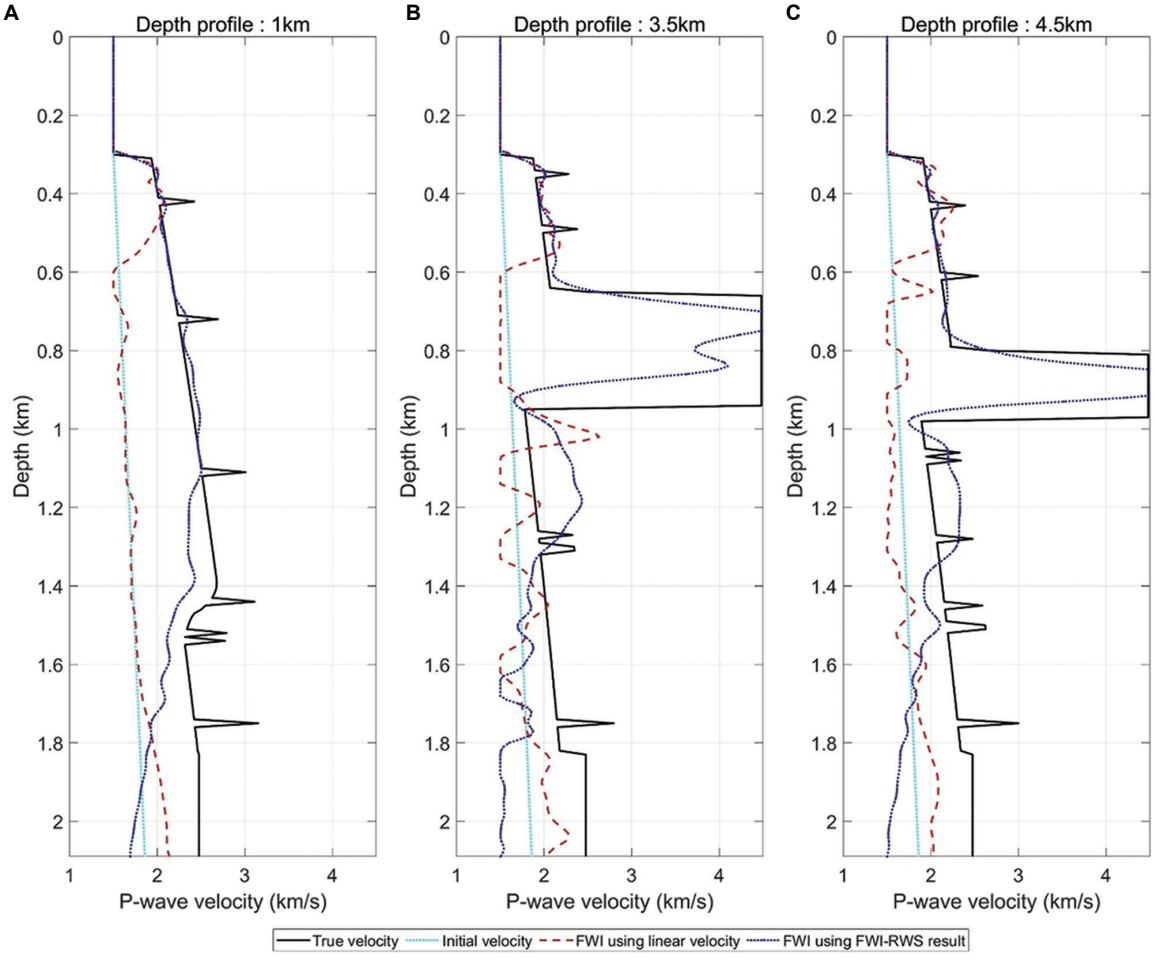


Figure 10. Comparison of depth-velocity profiles at horizontal position of (A) 1 km, (B) 3.5 km, and (C) 4.5 km. The true velocity (solid black lines), the initial velocity model (doted cyan lines), full waveform inversion (FWI) result obtained using linear velocity model (dashed red lines) and FWI-RWS result.

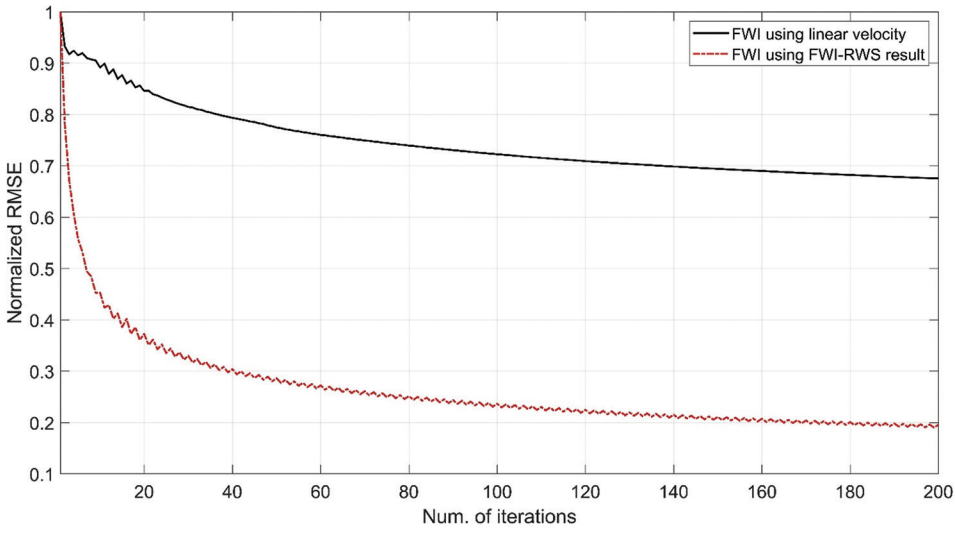


Figure 11. Root mean square error (RMSE) history of the full waveform inversion (FWI) started from the linear velocity model (solid black line) and that of the FWI started from the FWI-RWS results (dashed red line)

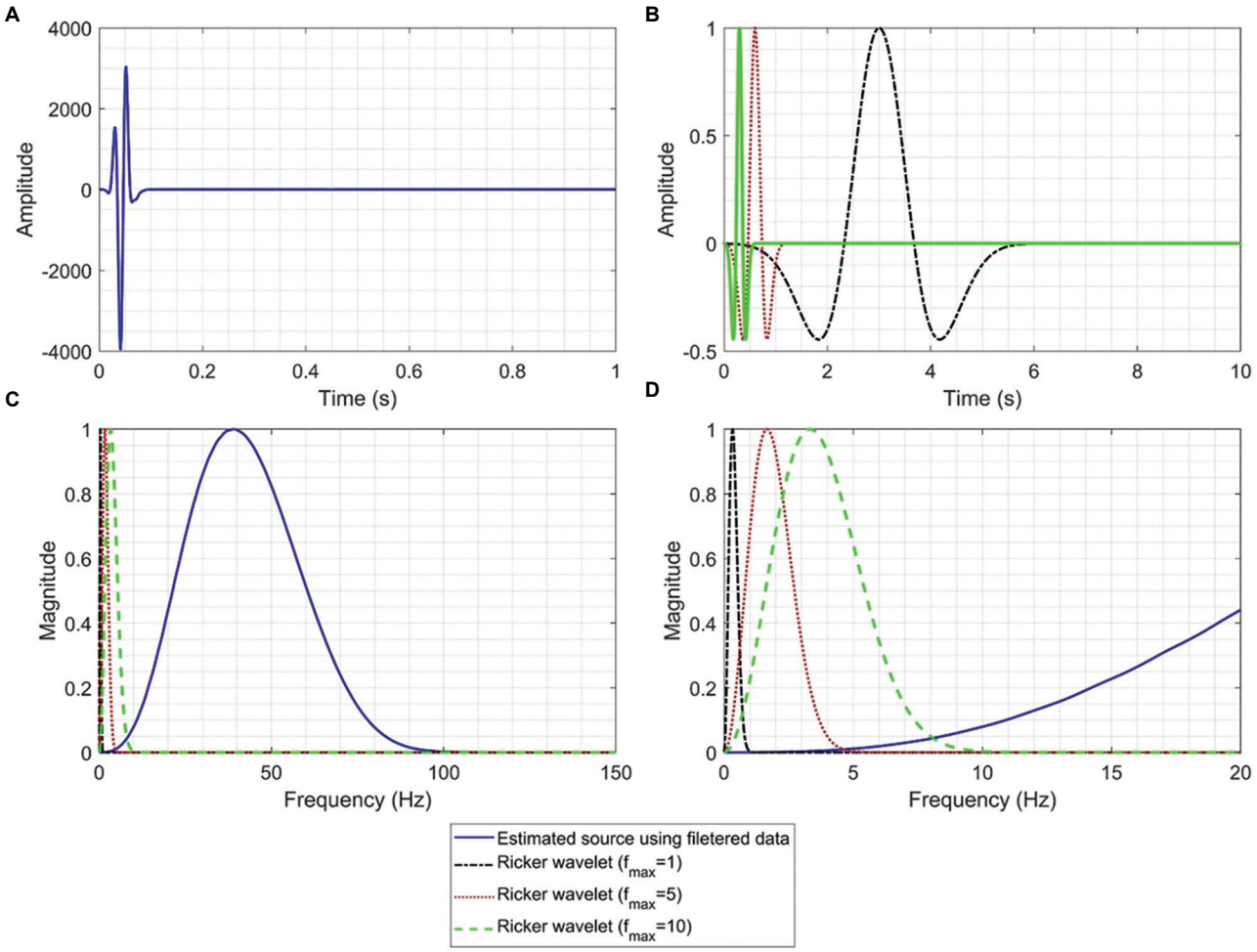


Figure 12. Estimated source analysis and band-limited convolution in time and frequency domains. (A) Estimated source using filtered data in the time domain. (B) Convolved sources with maximum frequencies of 1, 5, and 10 Hz in the time domain. (C) Frequency domain analysis of the estimated source shown in (A) and the convolved sources in (B). (D) Zoomed-in view of the frequency domain up to 20 Hz.

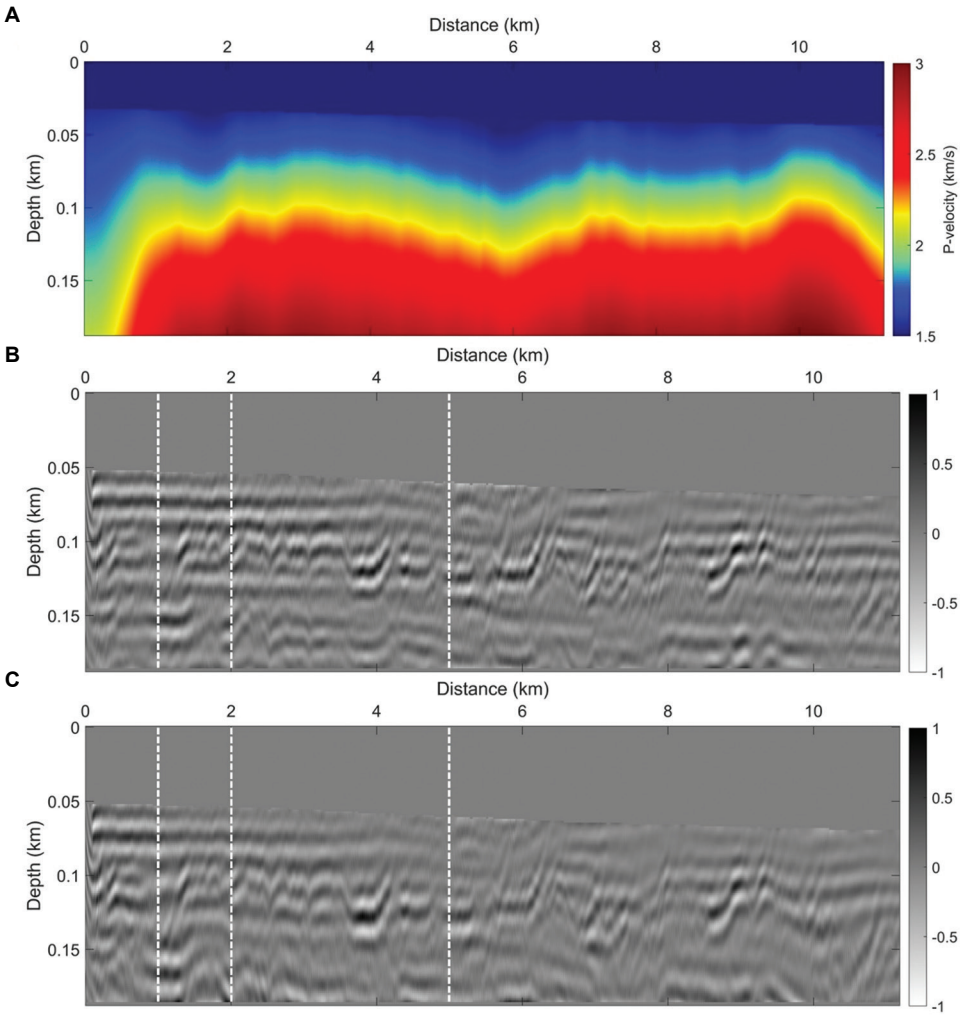


Figure 13. Imaging results from field data. (A) FWI-RWS results using field data. Reverse time migration results post-processing obtained using a (B) homogenous model, and (C) FWI-RWS results as the background velocity model.

first-arrival signal, the transform assigns it significant weight, necessitating the removal of first-arrival noise before applying the Laplace transform. To further test the applicability of the algorithm to noisy data, we analyzed the first-arrival traveltime and amplitude as a function of the SNR to assess noise sensitivity. The first-arrival traveltime and amplitude of the observed data using a strong Laplace damping factor e^{-st} are calculated as follows:

$$\tilde{d}(s)=\int_0^{\infty} d(t)e^{-st}dt \approx A_f e^{-st_f} \quad \text{(XVII)}$$

where $\bar{d}$ is the Laplace-transformed observed data.

Therefore, to analyze the noise sensitivity of the proposed method, tests were conducted with varying Laplace damping constants and intervals. For the noise test, arbitrary seismic signals were generated, and noise was added to achieve SNRs of 10 dB, 20 dB, and 30 dB (Figure 15). When the first-arrival traveltime was unknown, the error in first-arrival traveltime and amplitude increased at low SNR levels (Figure 16). The synthetic test showed results nearly identical to situations where the first-arrival traveltime was known (Figure 17). When the first-arrival traveltime was established, the process tended to yield similar outcomes, largely unaffected by the level of noise present. Successful application of this technique is anticipated if

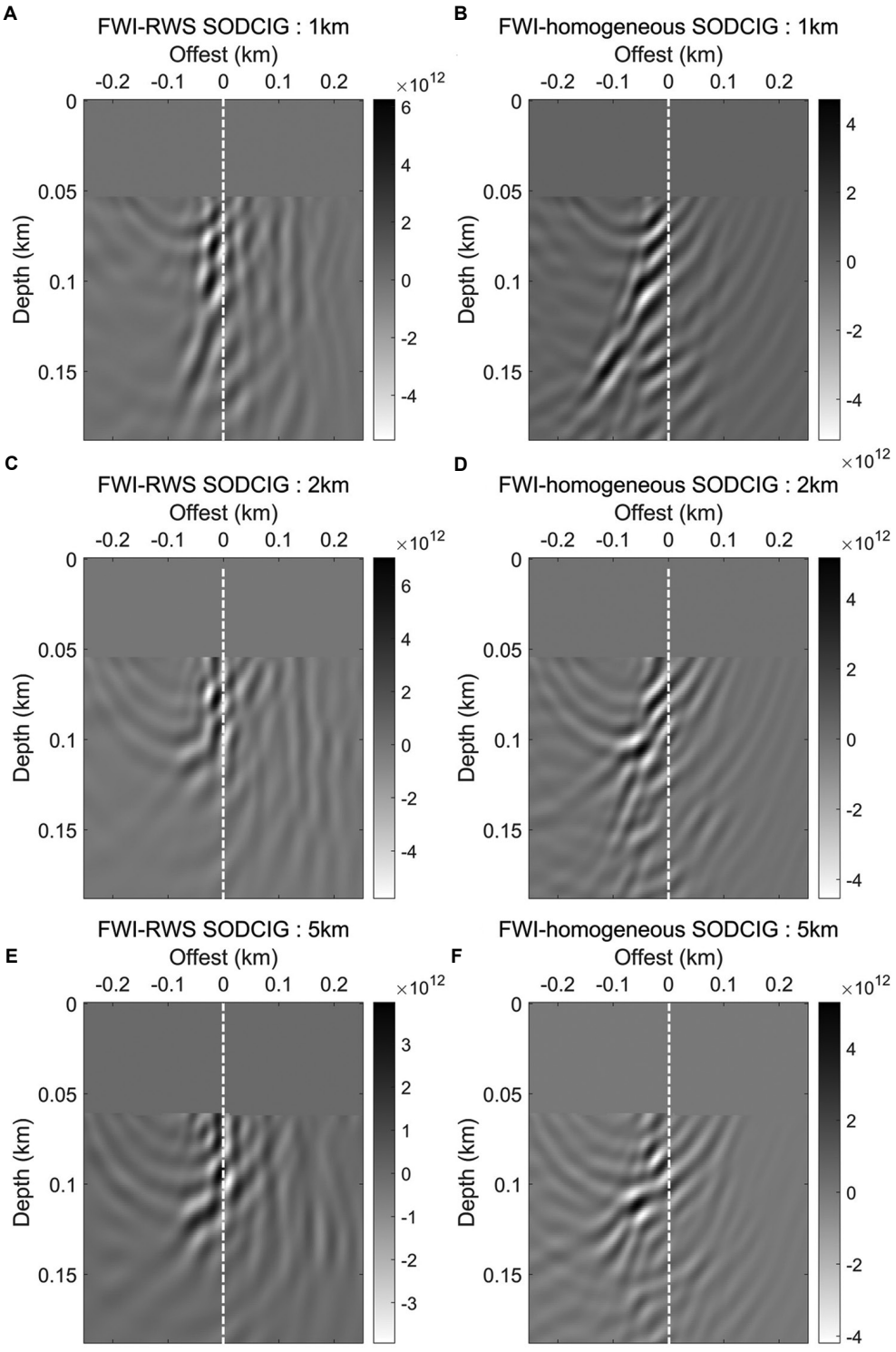


Figure 14. Subsurface offset domain common image gathers at (A and B) 1 km, (C and D) 2 km, and (E and F) 5 km. (A), (C), and (E) are obtained by reverse time migration (RTM) using FWI-RWS results, while (B), (D), and (F) are obtained by RTM using a homogeneous model. The offset range of each common depth point is from −0.2 km to 0.2 km. The white dashed lines mark the common depth point (CDP) location (1 km in [A and B], 2 km in [C and D], and 5 km in [E and F]).

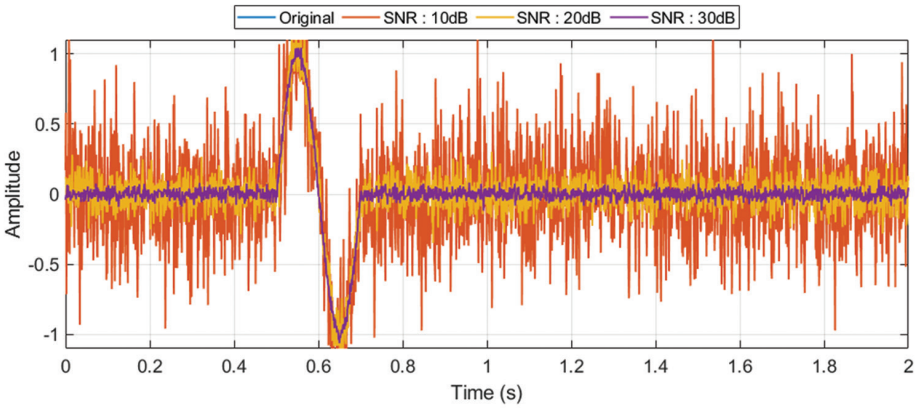


Figure 15. Seismic trace with low signal-to-noise ratio (SNR; 10–30 dB)

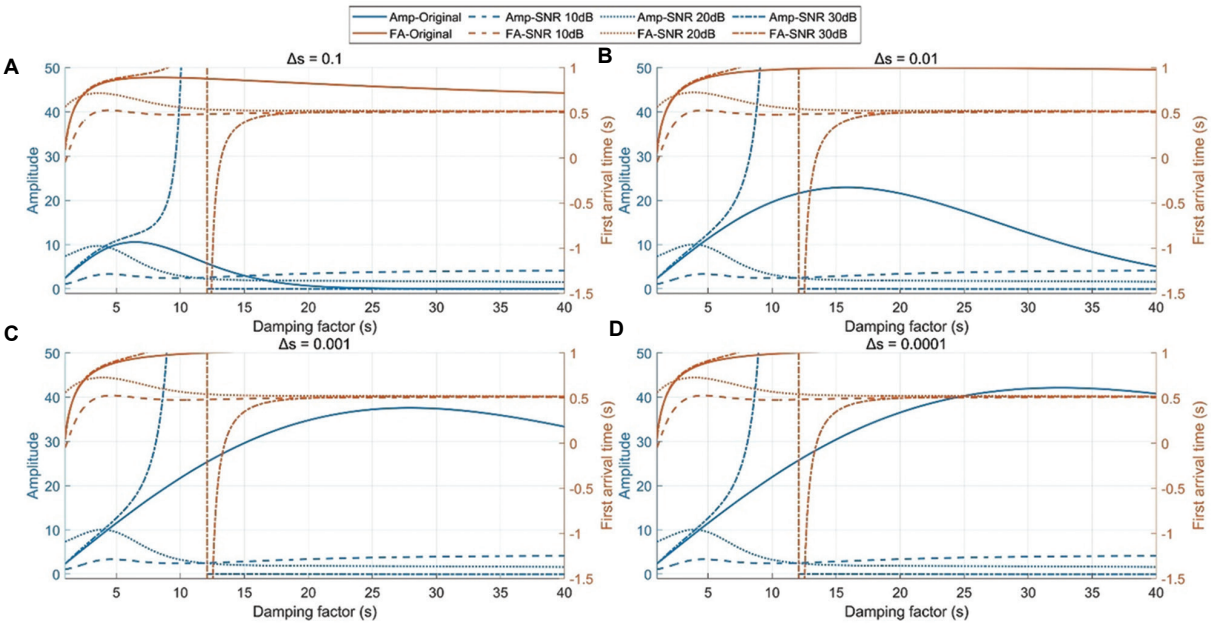


Figure 16. The scenarios with unknown first-arrival traveltime. Variations in the first-arrival traveltime and amplitude in relation to the Laplace damping constant, with a Laplace damping constant interval of (A) 0.1, (B) 0.01, (C) 0.001, or (D) 0.0001.
Abbreviations: Amp: Amplitude; FA: First-arrival traveltime; SNR: Signal-to-noise ratio.

picking is performed for each source, ensuring accurate determination of first-arrival traveltime. At very low SNR levels, our analysis showed that the first-arrival traveltime and amplitude remained consistent across SNRs, provided the first-arrival traveltime of the observed data was accurately known.

5. Conclusion

We presented FWI-RWS, a physics-guided framework that leverages the SWEET method-derived first-arrival traveltime and amplitude. This is achieved by convolving a target source wavelet and gradually increasing its maximum frequency. The potential of the proposed method lies in its ability to enhance resolution incrementally through multi-scale inversion. Both synthetic and field experiments show that FWI-RWS recovers a reliable long-wavelength background from poor initial models and delivers clear RTM focusing improvements when used as the migration velocity model. When applied to field data and used as the background velocity model for RTM, FWI-RWS yields more accurate structural results than a homogeneous velocity model. In addition, we examined the sensitivity of

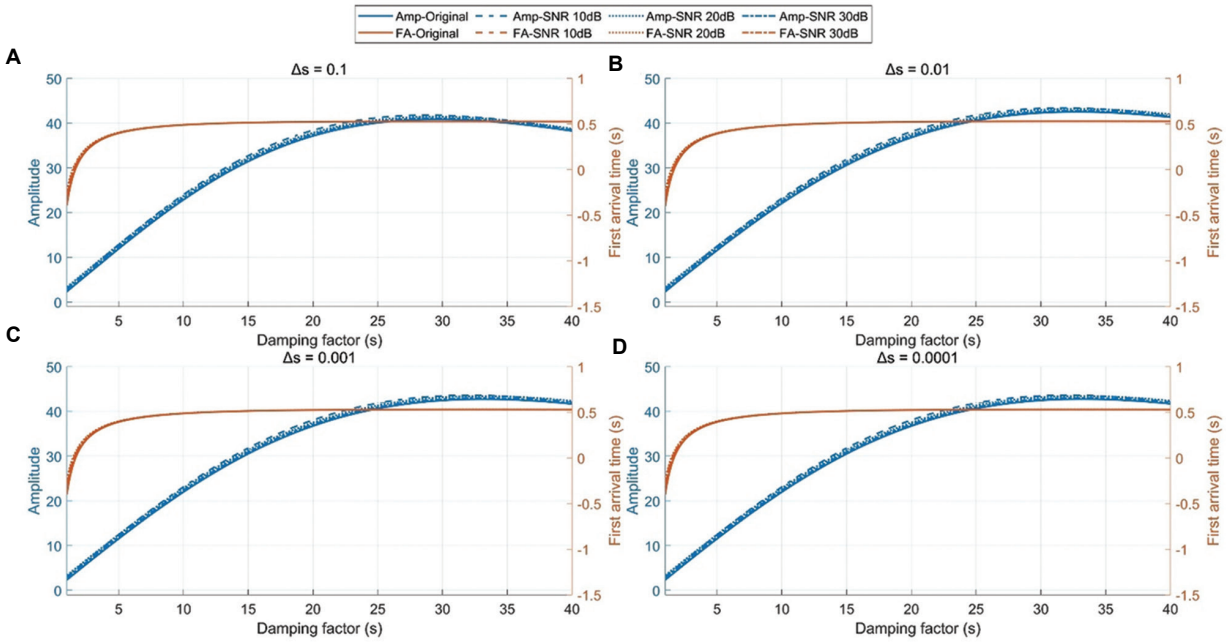


Figure 17. The scenarios with known first-arrival traveltime. Variations in the first-arrival traveltime and amplitude in relation to the Laplace damping constant, with a Laplace damping constant intervals of (A) 0.1, (B) 0.01, (C) 0.001, or (D) 0.0001.
Abbreviations: Amp: Amplitude; FA: First-arrival traveltime; SNR: Signal-to-noise ratio.

the SWEET algorithm to noise, confirming that accurate first-arrival traveltimes and amplitudes can be extracted, provided first-arrival picking is performed beforehand, regardless of noise. Future research will focus on improving the efficiency and applicability of the proposed method by analyzing the impact of different source wavelets and determining the optimal frequency components for the model, with the goal of producing background models that better support attribute analysis, diffraction imaging, and machine learning-based interpretation. Future work will explore adaptive selection of the target-wavelet schedule and extensions to anisotropic and elastic settings.

Acknowledgments

We would like to thank the Korea Institute of Geoscience and Mineral Resources (KIGAM) for permission to publish this work and to show data examples.

Funding

This research was supported by the Korea Institute of Marine Science & Technology Promotion (KIMST), funded by the Ministry of Oceans and Fisheries, Korea (RS-2023-00259633); and the Basic Research Project "Development of operation management infrastructure for TAMHAE 3 and seamless seismic technology connecting coastal areas (25-3322)" of the Korea Institute of Geoscience and Mineral Resources (KIGAM), funded by the Ministry of Science and ICT of Korea.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Seoje Jeong, Wookeen Chung
Formal analysis: Woohyun Son
Funding acquisition: Wookeen Chung
Investigation: Woohyun Son
Methodology: Sumin Kim, Wookeen Chung
Validation: Sumin Kim, Seoje Jeong
Writing–original draft: Seoje Jeong
Writing–review & editing: Seoje Jeong, Wookeen Chung

Availability of data

All data analyzed have been presented in the paper.

References

1. Lailly P, Bednar J. *The Seismic Inverse Problem as a Sequence of Before-Stack Migrations. In: Proceedings SIAM Conference Inverse Scattering*; 1983. p. 206-220.
2. Tarantola A. Inversion of seismic reflection data in the acoustic approximation. *Geophysics*. 1984;49(8):1259-1266. doi: 10.1190/1.1441754

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

   doi: 10.1190/1.3238367

4. Mora P. Inversion = migration + tomography. *Geophysics*. 1989;54(12):1575-1586.

   doi: 10.1190/1.1442625

5. Bunks C, Saleck FM, Zaleski S, Chavent G. Multiscale seismic waveform inversion. *Geophysics*. 1995;60(5):1457-1473.

   doi: 10.1190/1.1443880

6. Pratt RG, Shin C, Hicks GJ. Gauss-newton and full newton methods in frequency-space seismic waveform inversion. *Geophys J Int*. 1998;133(2):341-362.

   doi: 10.1046/j.1365-246X.1998.00498.x

7. Gauthier O, Virieux J, Tarantola A. Two-dimensional nonlinear inversion of seismic waveforms: Numerical results. *Geophysics*. 1986;51(7):1387-1403.

   doi: 10.1190/1.1442188

8. Zelt CA, Smith RB. Seismic traveltime inversion for 2-D crustal velocity structure. *Geophys J Int*. 1992;108(1):16-34.

   doi: 10.1111/j.1365-246X.1992.tb00836.x

9. Zhang J, Ten Brink US, Toksöz MN. Nonlinear refraction and reflection travel time tomography. *J Geophys Res Solid Earth*. 1998;103(B12):29743-29757.

   doi: 10.1029/98JB01981

10. Woodward MJ, Nichols D, Zdraveva O, Whitfield P, Johns T. A decade of tomography. *Geophysics*. 2008;73(5):VE5-VE11.

    doi: 10.1190/1.2969907

11. Brenders AJ, Pratt RG. Full waveform tomography for lithospheric imaging: Results from a blind test in a realistic crustal model. *Geophys J Int*. 2007;168(1):133-151.

    doi: 10.1111/j.1365-246X.2006.03156.x

12. Brenders AJ, Pratt RG, Kamei R, Charles S. Waveform Tomography-Marine vs Land: Targets, Challenges and Opportunities. *In: 72nd EAGE Conference and Exhibition-Workshops and Fieldtrips*; 2010. p. 162.

13. Ravaut C, Operto S, Improta L, Virieux J, Herrero A, Dell'Aversana P. Multiscale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-waveform tomography: Application to a thrust belt. *Geophys J Int*. 2004;159(3):1032-1056.

    doi: 10.1111/j.1365-246X.2004.02442.x

14. Operto S, Virieux J, Dessa JX, Pascal G. Crustal seismic imaging from multifold ocean-bottom seismometer data by frequency-domain full-waveform tomography: Application to the eastern Nankai Trough. *J Geophys Res Solid Earth*. 2006;111(B9):B09306.

    doi: 10.1029/2005JB003835

15. Bording RP, Gersztenkorn A, Lines LR, Scales JA, Treitel S. Applications of seismic travel-time tomography. *Geophys J Int*. 1987;90(2):285-303.

    doi: 10.1111/j.1365-246X.1987.tb00728.x

16. Moser TJ. Shortest path calculation of seismic rays. *Geophysics*. 1991;56(1):59-67.

    doi: 10.1190/1.1442958

17. Rawlinson N, Sambridge M. Seismic traveltime tomography of the crust and lithosphere. *Adv Geophys*. 2003;46:81-199.

    doi: 10.1016/S0065-2687(03)46002-0

18. Leung S, Qian J. An adjoint-state method for three-dimensional transmission traveltime tomography using first arrivals. *Commun Math Sci*. 2006;4:249-266.

    doi: 10.4310/CMS.2006.v4.n1.a10

19. Liu Y, Wu Z, Geng Z. First-arrival phase-traveltime tomography. *In: SEG 2017 Workshop: Full-Waveform Inversion and Beyond*. United States: SEG; 2017. p. 83-86.

20. Shin C, Cha YH. Waveform inversion in the Laplace domain. *Geophys J Int*. 2008;173(3):922-931.

    doi: 10.1111/j.1365-246X.2008.03768.x

21. Ha W, Pyun S, Yoo J, Shin C. Acoustic full-waveform inversion of synthetic land and marine data in the Laplace domain. *Geophys Prospect*. 2010;58(6):1033-1047.

    doi: 10.1111/j.1365-2478.2010.00884.x

22. Bozdağ E, Trampert J, Tromp J. Misfit functions for full-waveform inversion based on instantaneous phase and envelope measurements. *Geophys J Int*. 2011;185(2):845-870.

    doi: 10.1111/j.1365-246X.2011.04970.x

23. Wu R, Luo J, Wu B. Seismic envelope inversion and modulation-signal model. *Geophysics*. 2014;79(3):WA13-WA24.

    doi: 10.1190/geo2013-0294.1

24. Xiong K, Lumley D, Zhou W. Improved seismic-envelope full-waveform inversion. *Geophysics*. 2023;88(4):R421-R437.

    doi: 10.1190/geo2022-0444.1

25. Xu S, Wang D, Chen F, Lambaré G, Zhang Y. Inversion on reflected seismic wave. *In: SEG Annual International Meeting, Expanded Abstracts*. United States: SEG; 2012. p. 1-7.

    doi: 10.1190/segam2012-1473.1

26. Zhou H, Amundsen L, Zhang G. Fundamental issues in full-waveform inversion. *In: SEG Annual International Meeting, Expanded Abstracts*. United States: SEG; 2012. p. 1-5.

    doi: 10.1190/segam2012-0878.1

27. Berkhout AJ. Combining full wavefield migration and full-waveform inversion: A glance into the future of seismic imaging. *Geophysics*. 2012;77(2):S43-S50.

doi: 10.1190/geo2011-0148.1

28. Dong S, Dong X, Zhang R, Cong Z, Zhong T, Wang H. Global-feature-fusion and multiscale network for low-frequency extrapolation. *IEEE Trans Geosci Remote Sens*. 2024;62:1-14.

doi: 10.1109/TGRS.2024.3408949

29. Operto S, Gholami A, Aghamiry H, Guo G, Beller S. Extending the search space of full-waveform inversion beyond the single-scattering born approximation: A tutorial review. *Geophysics*. 2023;88(6):R671-R702.

doi: 10.1190/geo2022-0758.1

30. Métivier L, Brossier R, Mérigot Q, Oudet E, Virieux J. Measuring the misfit between seismograms using an optimal-transport distance: Application to full-waveform inversion. *Geophys J Int*. 2016;205(1):345-377.

doi: 10.1093/gji/ggw014

31. Van Leeuwen T, Herrmann FJ. Mitigating local minima in full-waveform inversion by expanding the search space. *Geophys J Int*. 2013;195(1):661-667.

doi: 10.1093/gji/ggt258

32. Dong X, Yuan Z, Lin J, Dong S, Tong X, Li Y. *PreAdaptFWI: Pretrained-Based Adaptive Residual Learning for full-Waveform Inversion without Dataset Dependency*. [Preprint]; 2025.

doi: 10.48550/arXiv.2502.11913

33. Muller AP, Costa JC, Bom CR, *et al*. Deep pre-trained FWI: Where supervised learning meets physics-informed neural networks. *Geophys J Int*. 2023;235(1):119-134.

doi: 10.1093/gji/ggad215

34. Cheng S, Wang Y, Zhang Q, Harsuko R, Alkhalifah T. A self-supervised learning framework for seismic low-frequency extrapolation. *J Geophys Res Mach Learn Comput*. 2024;1(3):e2024JH000157.

doi: 10.1029/2024JH000157

35. Wu RS, Toksöz MN. Diffraction tomography and multisource holography applied to seismic imaging. *Geophysics*. 1987;52(1):11-25.

doi: 10.1190/1.1442237

36. Shin C, Min DJ, Lim HY, *et al*. Traveltime and amplitude calculations using the damped-wave solution. *Geophysics*. 2002;67(5):1637-1647.

doi: 10.1190/1.1512811

37. Shin C, Min DJ. Waveform inversion using a logarithmic wavefield. *Geophysics*. 2006;71(3):R31-R42.

doi: 10.1190/1.2194523

38. Shin C, Jang S, Min DJ. Improved amplitude preservation for prestack depth migration by inverse scattering theory. *Geophys Prospect*. 2001;49(5):592-606.

doi: 10.1046/j.1365-2478.2001.00279.x

39. Boonyasiriwat C, Valasek P, Routh P, Zhu X. Application of multiscale waveform tomography for high-resolution velocity estimation in complex geologic environments: Canadian Foothills synthetic-data example. *Lead Edge*. 2009;28(4):454-456.

doi: 10.1190/1.3112764

40. Pasalic D, McGarry R. Convolutional perfectly matched layer for isotropic and anisotropic acoustic wave equations. *In: SEG International Exposition and Annual Meeting*. United States: SEG; 2010.

doi: 10.1190/1.3513453

41. Aminzadeh F, Burkhard N, Nicoletis L, Rocca F, Wyatt K. SEG/EAEG 3-D modeling project: Second update. *Lead Edge*. 1994;13(9):949-952.

doi: 10.1190/1.1437054

42. Symes WW. A differential semblance criterion for inversion of multioffset seismic reflection data. *J Geophys Res*. 1993;98(B2):2061-2073.

doi: 10.1029/92JB01304

43. Prucha ML, Biondi BL, Symes WW. Angle-domain common-image gathers by wave-equation migration. *In: SEG Technical Program Expanded Abstracts*. United States: SEG; 1999. p. 824-827.

doi: 10.1190/1.1821156

44. Xu S, Chauris H, Lambaré G, Noble M. Common-angle migration: A strategy for imaging complex media. *Geophysics*. 2001;66(6):1877-1894.

doi: 10.1190/1.1487131

45. Zhou H, Gray SH, Young J, Pham D, Zhang Y. Tomographic residual curvature analysis: The process and its components. *In: SEG Technical Program Expanded Abstracts*. United States: SEG; 2003. p. 666-669.

doi: 10.1190/1.1818018

46. Sava PC, Fomel S. Angle-domain common-image gathers by wavefield-continuation methods. *Geophysics*. 2003;68(3):1065-1074.

doi: 10.1190/1.1581078

47. Valenciano AA, Biondi B. Wave-equation angle-domain Hessian. *In: 68th EAGE Conference and Exhibition Incorporating SPE EUROPEC*. 2006. p. 2.

doi: 10.3997/2214-4609.201402166

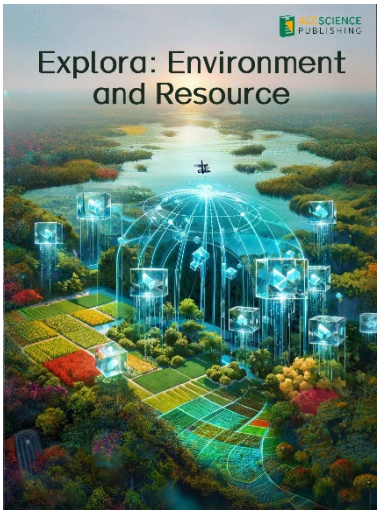
48. Chauris H, Lameloise CA, Donno D. Migration velocity analysis with reflected and transmitted waves. In: *75th EAGE Conference and Exhibition Incorporating SPE EUROPEC*. 2013. p. 348.

doi: 10.3997/2214-4609.20130936

49. Kalita M, Alkhalifah T. Common-image gathers using the excitation-amplitude imaging condition. *Geophysics*. 2016;81(4):S261-S269.

doi: 10.1190/geo2015-0413.1

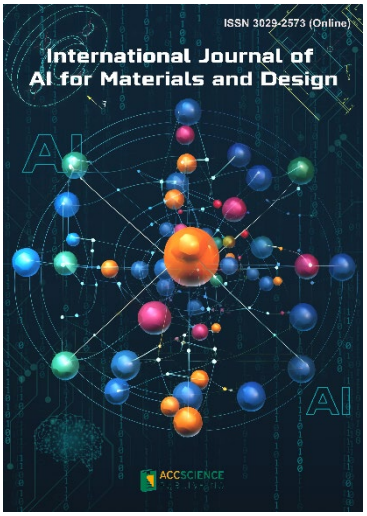
Our Journals



Explora: Environment and Resource (EER) is an international and multidisciplinary journal covering all aspects of the environmental impacts of socio-economic development. It is concerned with the complex interactions among society, development, and the environment, aiming to explore ways and means of achieving sustainability in all human activities related to development.

EER covers subject areas, including but not limited to the following:

- Water reclamation, wastewater treatment and waste management for eco-environment sustainability
- Innovative nanotechnology, catalysis, photocatalysis and nano materials for decontamination practices and renewable energy resources.
- Climate change-related sustainable resource and environment quality analysis
- Microorganisms and plants-oriented low-carbon processes and strategies for clean energy, water and resource management



International Journal of AI for Materials and Design is an international, peer-reviewed open-access journal that aims to bridge the cutting-edge research between AI and materials, AI and design. In recent years, the tremendous progress in AI is leading a radical shift of AI research from a mainly academic endeavor to a much broader field with increasing industrial and governmental investments. The maturation of AI technology brings about a step change in the scientific research of various domains, especially in the world of materials and design. Machine learning (ML) algorithms enable researchers to analyze extensive datasets on material properties and accurately predict their behavior in different conditions. This subsequently impact the industry to leverage on big data and advanced analytics to build scientific strategies, scale operational performance of processes and drive innovation. In addition, AI and ML are uniquely positioned to enable advanced manufacturing technologies across the value chain of different industries. Integration of multiple and complementary AI techniques, such as ML, search, reasoning, planning, and knowledge representation, will further accelerate advances in scientific discoveries, engineering excellence and the future of cyber-physical systems manufacturing.

International Journal of AI for Materials and Design covers the following topics: AI or machine learning for material discovery, AI for process optimization, AI and data-driven approaches for product or systems design, application of AI in advanced manufacturing processes such as additive manufacturing, IoT, sensors, robotics, cloud-based manufacturing, intelligent manufacturing for various applications, autonomous experiments, material intelligence, energy intelligence, and AI-linked decarbonization technologies.

Start a new journal

Write to us via email if you are interested to start a new journal with AccScience Publishing. Please attach your CV, professional profile page and a brief pitch proposal in your email. We shall inform you of our decision whether we are interested to collaborate in starting a new journal.

Contact: info@accscience.com

https://accscience.com/journal/JSE

Access Science Without Barriers

Contact

www.accscience.com
9 Raffles Place, Republic Plaza 1 #06-00 Singapore 048619
E-mail: editorial@accscience.com
Phone: +65 8182 1586