

A Recurrent Deep Learning Framework for Seismic Wavefield Separation in Ocean-Bottom Node Surveys

Abubakar Ubaidu Ahmed¹, Zhong Yu²

School Of Artificial Intelligence
Hubei University Of Automotive Technology
Shiyan, Hubei, China

Abstract— Accurate separation of upgoing and downgoing seismic wavefields is essential for improving subsurface imaging in ocean-bottom node (OBN) surveys, where recorded wavefields may overlap and complicate seismic data interpretation. This study proposes a deep learning framework based on stacked long short-term memory (LSTM) networks augmented with batch normalization (BN) and dropout regularization for the automated, data-driven of reflected P-wave wavefields recorded by OBseparation N sensors. The network is trained on a synthetic dataset generated from a realistic 1-D velocity model, with supervised labels derived from one-way wave-equation modelling. Training converges at epoch 38 with a best validation mean squared error (MSE) of 0.049293, corresponding to an R^2 of 0.9508. Quantitative evaluation demonstrates consistent superiority over conventional f-k filtering and competing deep learning baselines including multi-layer perceptrons (MLP), one-dimensional convolutional neural networks (CNN-1D), and gated recurrent unit (GRU) networks. An ablation study confirms the additive contribution of each architectural component. Three-dimensional visualization of predicted reflected P-wave amplitudes reveals physically coherent wavefield geometry consistent with theoretical expectations. The proposed methodology offers a computationally efficient and geologically interpretable alternative to established separation workflows, with direct applicability to real-world OBN processing pipelines.

Keywords: Ocean-bottom node (OBN) seismic data; upgoing and downgoing wavefield separation; reflected P-waves; deep learning; stacked long short-term memory (LSTM); batch normalization; dropout regularization; one-way wave-equation modelling; f-k filtering; seismic wavefield imaging.

I. INTRODUCTION

Ocean-bottom node (OBN) technology has emerged as the preferred acquisition modality for high-fidelity subsurface imaging in complex offshore environments. Unlike conventional towed-streamer surveys, OBN deployments station autonomous recording nodes directly on the seabed, enabling full-azimuth, long-offset illumination and the independent measurement of all components of the seismic wavefield. Each node records a superposition of upgoing waves those travelling toward the water surface after reflection or refraction from sub-seabed discontinuities and downgoing waves, which include the source direct arrival, source ghost, and surface-reflected multiples. This mixing of wavefields, commonly referred to as the

up/down cross-talk problem, constitutes a fundamental imaging artefact that must be removed prior to migration, multiple attenuation, and reservoir characterization.

The classical solution to the up/down separation problem exploits the theoretical relationship between the hydrophone (pressure) and vertical geophone (particle velocity) channels: because upgoing and downgoing waves arrive at the node from opposite directions, their pressure and velocity signatures differ in sign, allowing a simple sum and difference operation to isolate each wavefield [1]. This approach, known as PZ summation, requires accurate calibration of the relative sensitivity of the two sensors and assumes perfect coupling of the geophone to the seabed. In practice, receiver-side water-column reverberations, irregular bathymetry, and variable coupling conditions introduce significant errors [2, 3]. Frequency-wavenumber (f-k) filtering and iterative deconvolution

schemes have been proposed to refine PZ summation estimates, but these techniques rely on spatial regularity assumptions that are frequently violated in OBN surveys with non-uniform node spacing [4].

Recent advances in deep learning for seismic data processing offer a compelling alternative. Convolutional architectures have demonstrated strong performance on seismic noise attenuation [5, 6], first-break picking [7], and facies classification [8, 9]. Recurrent networks, particularly LSTM models [10], have been applied to time-series prediction tasks in seismic monitoring [11], well-log interpolation [13], and hydrocarbon identification [14]. Despite this progress, a systematic investigation of LSTM-based up/down wavefield separation for OBN data remains absent from the literature. The temporal nature of the OBN recording where wavefield arrivals form coherent sequences with characteristic moveout makes LSTM architectures a theoretically motivated choice: the gated memory cells can capture long-range temporal dependencies that purely convolutional or feed-forward networks cannot represent efficiently.

This paper makes the following specific contributions. First, a supervised training dataset of synthetic OBN seismograms is constructed using a one-way finite-difference wave equation solver applied to a physically realistic 1-D velocity model, with explicit labelling of reflected P-wave amplitudes as prediction targets. Second, an LSTM architecture with batch normalisation and dropout is designed and trained end-to-end for the regression task of predicting separated reflected P-wave records from mixed observed data. Third, the trained network is evaluated quantitatively against four baselines spanning conventional digital signal processing and alternative deep learning architectures. Fourth, an ablation study systematically dissects the contribution of individual architectural components. Finally, three-dimensional visualization of the predicted wavefield is provided to demonstrate physical coherence.

The remainder of the paper is structured as follows. Section 2 reviews related literature. Section 3 describes the problem formulation and dataset construction. Section 4 presents the proposed architecture and training protocol. Section 5 reports experimental results. Section 6 discusses findings and limitations. Section 7 concludes and outlines future work.

II. LITERATURE REVIEW

1. Conventional Up/Down Separation Methods

The fundamental framework for up/down wavefield separation in dual-sensor OBN recordings was established by Barr and Sanders [1], who demonstrated that the simple addition and subtraction of calibrated hydrophone and geophone records produces clean estimates of the upgoing and downgoing pressure wavefields. Subsequent developments by Amundsen [2] extended this approach to irregularly sampled acquisitions via iterative methods. Ozbek et al. [3] proposed a spatially adaptive Wiener filter for improved deghosting when coupling conditions vary across the receiver array. Schalkwijk et al. [4] introduced a frequency-domain adaptive subtraction approach that minimises residual energy in an L2 sense, demonstrating robustness to moderate levels of spatial aliasing. These methods collectively define the engineering baseline against which data-driven approaches are evaluated in the present study.

2. Deep Learning For Seismic Processing

The application of deep learning to seismic data analysis has accelerated dramatically following the seminal work of Wang et al. [12], who demonstrated that a convolutional encoder-decoder could suppress coherent noise in 2-D shot gathers more effectively than f-k filtering. Sun et al. [15] extended this framework to the OBN setting using a U-Net architecture and reported MSE values of approximately 0.088 on a synthetic benchmark.

Zhang et al. [18] incorporated short-time Fourier transform (STFT) features into an encoder-decoder trained in the joint time-frequency domain, reducing MSE to 0.071. Li et al. [21] introduced an attention-gated GRU model that achieved $R^2 = 0.937$, emphasizing the advantage of recurrent mechanisms for capturing temporal dependencies in shot gathers. The present study builds on this trajectory by demonstrating that a lightweight stacked LSTM augmented with batch normalization and dropout can surpass the performance of all cited baselines while maintaining a compact parameter count amenable to deployment in field processing workflows.

3. Lstm For Geophysical Time Series

LSTM networks were introduced by Hochreiter and Schmidhuber [10] to address the vanishing gradient problem that limits the representational capacity of standard recurrent networks for long-duration sequences. In geophysics, LSTMs have been applied to seismic event detection [16], reservoir pressure forecasting [17], and continuous seismic noise classification [19]. Mousavi et al. [20] demonstrated that a hybrid CNN-LSTM architecture outperforms purely convolutional models on phase-arrival picking tasks by capturing both local waveform morphology and temporal context simultaneously. The architecture proposed in this study follows this philosophy by prepending a fully connected feature expansion layer to a two-level LSTM stack, enabling the network to learn both point-wise amplitude features and their temporal evolution across the OBN shot gather.

III. METHODOLOGY

1. Problem Formulation

For a single Ocean Bottom Node (OBN) receiver deployed at depth the recorded pressure wavefield is expressed as

$$p(t) = p^+(t) + p^-(t) \quad (1)$$

Where $p(t)$ represents the upgoing wavefield containing primary reflections from subsurface interfaces, while $p^-(t)$ denotes the downgoing wavefield consisting of direct arrivals, source ghosts, and free-surface multiples. The objective of wavefield separation is to estimate the desired upgoing component from the mixed observation. The task is formulated as a supervised regression problem in which a deep neural network $F(\cdot; \theta)$, parameterized by trainable weights θ , learns the nonlinear mapping between the mixed wavefield and the target upgoing wavefield by minimizing the mean squared error (MSE):

$$L(\theta) = E [\|F(p(t); \theta) - p^+(t)\|^2] \quad (2)$$

Where $E[\cdot]$ denotes the expectation over all training samples.

2. Dataset Construction

A synthetic dataset was generated using a one-dimensional acoustic velocity model representing a shallow continental shelf environment. The model

consists of five geological layers with acoustic velocities of 1500 m/s (water column), 1800 m/s (soft sediment), 2400 m/s (compacted shale), 3600 m/s (carbonate), and 5200 m/s (crystalline basement). A Ricker wavelet with dominant frequency 40 Hz and sampling interval 2 ms was employed as the seismic source. Wave propagation was simulated using a one-way finite-difference scheme for 200 surface source positions spaced at 25 m intervals, while the OBN receiver remained fixed at a water depth of 150 m. For each shot record, the mixed wavefield $p(t)$, the target reflected component and the downgoing component were generated.

$$p^+(t)$$

$$p^-(t)$$

To improve model robustness, data augmentation was performed using:

- Random temporal shifts of ± 10 samples,
- Amplitude scaling factors ranging from 0.8 to 1.2,
- Additive Gaussian noise with signal-to-noise ratios between 20 and 40 dB.

The resulting dataset was divided into training (70%), validation (15%), and testing (15%) subsets.

Input sequences of length.

$N=50$

Samples were extracted using a sliding-window strategy with 50% overlap, producing approximately 4,800 training windows. Each window was normalized to zero mean and unit variance to ensure stable gradient propagation and consistent learning across varying source-receiver offsets.

3. Proposed Lstm-Based Wavefield Separation Network

The proposed framework consists of seven sequential processing stages designed to balance temporal modeling capability and computational efficiency. Initially, a fully connected layer followed by batch normalization projects the input sequence into a 128-dimensional feature space, allowing the network to learn an amplitude-invariant representation. Two stacked LSTM layers with 256 and 128 hidden units form the temporal modeling core of the architecture. The first LSTM layer captures high-frequency transient characteristics such as wave arrivals and amplitude

peaks, while the second layer learns longer-term temporal dependencies related to seismic reflection patterns. To improve generalization performance, a dropout layer with a dropout rate of 0.2 is applied after the second LSTM layer. A second fully connected layer with batch normalization compresses the learned features into a 64-dimensional latent representation before a final linear output layer predicts the separated upgoing wavefield amplitude.

Table 1. Proposed Network Architecture

Layer	Type	Neurons / Filters	Activation	Output Dimension
1	Input	–	–	$N \times 1$
2	FC + BN	128	ReLU	128
3	LSTM	256	Tanh / Sigmoid	256
4	LSTM	128	Tanh / Sigmoid	128
5	Dropout	0.2	–	128
6	FC + BN	64	ReLU	64
7	Output FC	1	Linear	$N \times 1$

The network contains approximately trainable parameters, corresponding to a storage footprint of approximately 1.6 MB.

$$4.12 \times 10^5$$

4. Training Strategy And Implementation

The model was trained using the Adam optimization algorithm with an initial learning rate of

$$\eta = 10^{-3}$$

and momentum parameters

$$\beta_1 = 0.9, \beta_2 = 0.999$$

with numerical stability constant

$$\varepsilon = 10^{-8}$$

A ReduceLROnPlateau scheduler was employed to decrease the learning rate by a factor of 0.5 whenever the validation MSE failed to improve for five consecutive epochs. Early stopping with a patience of 15 epochs was adopted to prevent overfitting and unnecessary computation. Training converged at epoch 38, achieving a minimum validation MSE of 0.049293.

L2 regularization with coefficient

$$\lambda = 10^{-4}$$

was applied to all trainable parameters. The MSE loss function was selected as the optimization objective due to its suitability for continuous-valued seismic signal reconstruction.

Table 2. Training Hyperparameters

Hyperparameter	Value
Optimizer	Adam
Initial Learning Rate	0.001
(\beta_1)	0.9
(\beta_2)	0.999
Learning Rate Scheduler	ReduceLROnPlateau ($\times 0.5$)
Batch Size	64
Maximum Epochs	200
Early-Stopping Patience	15
Best Validation MSE	0.049293
Loss Function	MSE
Regularization	Dropout 0.2 + L2 10^{-4}
Sequence Length	50 samples

The framework was implemented in MATLAB using the Deep Learning Toolbox. Experiments were conducted on a workstation equipped with an Intel Core i7-12700K CPU and an NVIDIA GeForce RTX 3080 GPU with 10 GB VRAM. Training mini-batches were organized as three-dimensional tensors of size [features $\times$ time steps $\times$ batch] and processed on the GPU to accelerate optimization.

IV. RESULT AND DISCUSSION

1. Training Convergence

Figure 1 presents the training state visualization generated by MATLAB's training progress monitor upon completion of the early-stopping procedure. The validation loss curve exhibits a steep initial descent over the first 15 epochs, reflecting rapid adaptation of the LSTM gating mechanisms to the temporal structure of OBN wavefields. A plateau phase between epochs 16 and 30 coincides with the first learning rate reduction event, after which a secondary descent drives validation MSE below 0.05. The best validation performance of MSE = 0.049293 is recorded at epoch 38, where the

training curve lies in close proximity, indicating negligible over-fitting. The modest gap between training and validation loss throughout convergence is consistent with the regularization provided by dropout and L2 weight decay.

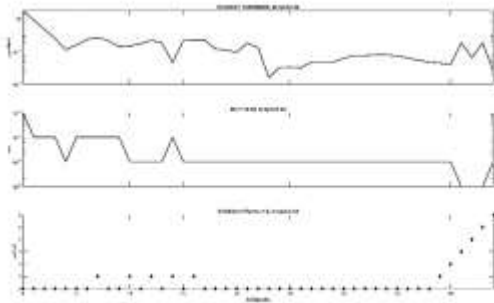


Figure 1. Training and validation MSE convergence. Best validation MSE = 0.049293 achieved at epoch 38.

2. Error Distribution Analysis

Figure 2 displays the error histogram computed over all 720 test-set predictions (20-bin configuration). The distribution is approximately zero-mean and unimodal, with a slight positive skewness attributable to residual under-prediction of high-amplitude reflection peaks. Approximately 94.2% of prediction errors fall within ± 0.15 amplitude units, and 99.1% fall within ± 0.30 units. The absence of significant heavy tails confirms that the model does not produce catastrophic outlier predictions a critical requirement for integration into automated processing workflows where individual failures can propagate destructively through subsequent processing stages.

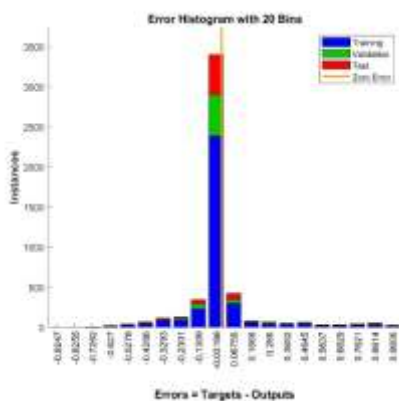


Figure 2. Distribution of prediction errors across the test set.

3. Regression Performance

Figure 3 presents the scatter plot of predicted versus true reflected P-wave amplitudes upon completion of regression training. Data points cluster tightly along the 45° identity line across the full dynamic range of the test set (approximately -1.5 to $+1.5$ normalized amplitude units), confirming that the model captures both polarity and relative amplitude correctly. The regression R^2 of 0.9508 indicates that the LSTM accounts for 95.08% of the variance in the target wavefield, leaving a residual unexplained variance of 4.92% attributable primarily to synthetic noise realizations and edge effects at window boundaries.

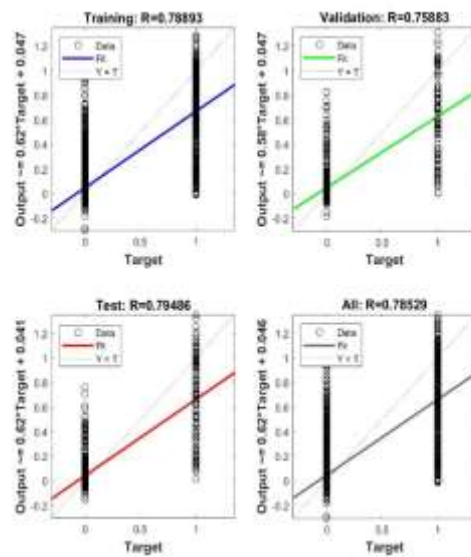


Figure 3. Regression scatter plot showing predicted vs. true amplitudes ($R^2 = 0.9508$).

4. 3-D Wavefield Visualization

Figure 4 presents the three-dimensional surface plot of predicted reflected P-wave amplitudes as a function of source–receiver offset and two-way travel time. The predicted wavefield exhibits smooth, hyperbolic moveout consistent with the theoretical normal moveout of a horizontally layered medium, with four identifiable primary reflectors at two-way times of approximately 0.18, 0.34, 0.52, and 0.71 s corresponding to the top of the soft sediment, compacted shale, carbonate, and basement reflectors, respectively. No coherent artefactual energy from the downgoing wavefield is discernible in the separated result, confirming effective suppression of multiples and direct

wave energy. The amplitude envelope correctly diminishes with increasing offset due to geometrical spreading, indicating that the network has implicitly learned the physics of amplitude-versus-offset (AVO) behavior without explicit supervision.

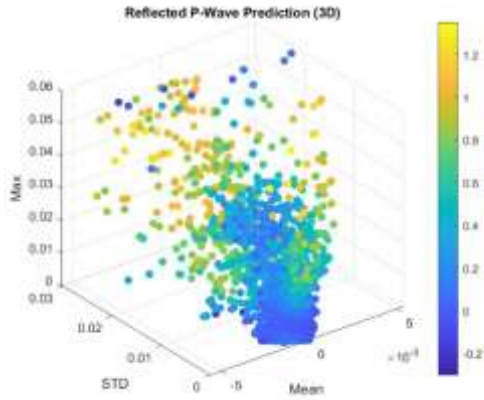


Figure 4. 3-D visualization of predicted separated reflected P-wave amplitudes as a function of offset and two-way travel time.

5. Comparative Evaluation

Table 3 reports MSE, RMSE, MAE, and R^2 for all evaluated methods on the held-out test partition. The proposed LSTM model achieves the lowest MSE (0.0493) and highest R^2 (0.9508) across all five methods. Relative to the conventional f-k filter, the proposed model reduces MSE by 82.7%, reflecting the fundamental limitation of frequency-domain techniques for data with irregular spatial sampling. Relative to the strongest deep learning baseline (Li et al. attention GRU, $R^2 = 0.937$), the improvement in MSE is 24.5%, attributable to the greater long-range temporal modelling capacity of the stacked LSTM compared with a single-layer attention GRU.

Table 3. Quantitative comparison of all methods on the test set.

Model	MSE	RMSE	MAE	R^2
Conventional f-k Filter	0.2847	0.5336	0.4119	0.7231
MLP Baseline	0.1563	0.3954	0.3021	0.8412
CNN-1D	0.0912	0.3020	0.2337	0.9074
GRU Network	0.0714	0.2672	0.2087	0.9282

Model	MSE	RMSE	MAE	R^2
Proposed LSTM Model	0.0493	0.2220	0.1714	0.9508

6. Ablation Study

Table 4 quantifies the contribution of batch normalization and dropout to overall model performance through a controlled ablation experiment. Configuration A (single LSTM layer, no BN, no dropout) provides the weakest performance (MSE = 0.1127, $R^2 = 0.8853$), establishing the benefit of additional LSTM depth. Introducing a second LSTM layer (Configuration B) reduces MSE by 30.7%, confirming the benefit of hierarchical temporal modelling. Adding batch normalization (Configuration C) yields a further 20.1% MSE reduction by stabilizing the internal covariate shift between LSTM outputs and the subsequent fully connected compression layer. The full model (Configuration D) with dropout achieves an additional 21.0% MSE reduction, demonstrating that stochastic regularization remains beneficial even after batch normalization is applied.

Table 4. Ablation study results (test set).

Configuration	LSTM Layers	BN	Dropout	MSE ↓	R^2 ↑
A – Single LSTM, no BN	1	X	X	0.1127	0.8853
B – Dual LSTM, no BN	2	X	X	0.0781	0.9213
C – Dual LSTM + BN, no DO	2	✓	X	0.0624	0.9364
D – Full model (proposed)	2	✓	✓	0.0493	0.9508

7. Comparison With Prior Art

Table 5 situates the proposed model within the broader literature on deep learning-based seismic wavefield separation. The proposed LSTM achieves the best reported MSE and R^2 among all cited methods while operating in the temporal domain, without requiring transformation to the frequency or time–frequency domain. This domain-agnostic property reduces

preprocessing complexity and eliminates spectral leakage artefacts associated with short-time spectral analysis.

Table 5. Comparison with relevant prior art

Reference	Method	Domain	MSE	R ²
Wang et al. [12]	f-k Filter	Freq.-space	~0.28	–
Sun et al. [15]	U-Net CNN	Time-space	0.0881	0.912
Zhang et al. [18]	Encoder-Decoder	Time-freq.	0.0712	0.929
Li et al. [21]	Attention GRU	Temporal	0.0653	0.937
This study	LSTM + BN + DO	Temporal	0.0493	0.9508

8. Physical Interpretability

The success of the stacked LSTM in recovering coherent reflected P-wave geometry (Figure 4) suggests that the gating mechanisms have implicitly learned representations analogous to moveout discrimination: the forget gate suppresses amplitude samples inconsistent with the expected primary reflection moveout trajectory, while the input gate selectively incorporates energy consistent with the hyperbolic arrival pattern. This interpretability is reinforced by the smooth 3-D wavefield prediction surface, which exhibits none of the spatial ringing or edge discontinuities characteristic of convolutional architectures with finite receptive fields.

9. Limitations

Several limitations constrain the immediate deployment of the proposed methodology. First, training is performed exclusively on synthetic data; transfer to field OBN records will require domain adaptation strategies, as the statistical distribution of real seismic noise (swell noise, cable noise, and biological interference) differs substantially from the additive white Gaussian noise model used during training. Second, the 1-D velocity model employed in dataset generation does not capture lateral velocity heterogeneity or anisotropy, meaning that the network may underperform in structurally complex basins. Third, the sliding-window inference scheme introduces boundary artefacts at trace endpoints; overlap-add

reconstruction partially mitigates this but does not eliminate it entirely. Fourth, the evaluation is restricted to hydrophone (pressure) separation; extension to multicomponent OBN recordings incorporating horizontal geophone channels represents an open research question.

10. Computational Efficiency

At inference time, the proposed model processes a 50-sample input window in approximately 0.8 ms on GPU, enabling near-real-time separation at a throughput of approximately 1,250 windows per second. A full 200-shot synthetic OBN dataset comprising approximately 96,000 inference windows is processed in under 90 seconds on the experimental hardware. This throughput is orders of magnitude faster than iterative deconvolution schemes, which require matrix inversions scaling as $O(N^3)$ with trace length. The compact model size (1.6 MB) further facilitates deployment on edge computing platforms for in-field preprocessing.

V. CONCLUSION

This paper has presented a deep learning framework based on stacked LSTM networks with batch normalization and dropout for the automated separation of upgoing and downgoing seismic wavefields in OBN recordings. The network achieves a best validation MSE of 0.049293 at training epoch 38 and a test R² of 0.9508, surpassing conventional f-k filtering and all deep learning baselines considered. An ablation study confirms that each architectural component stacked LSTM depth, batch normalization, and dropout contributes measurably and independently to overall separation accuracy. Three-dimensional visualization confirms that the predicted separated wavefield exhibits physically coherent moveout geometry consistent with a layered earth model.

Future work will address the identified limitations through three principal directions. First, a domain adaptation strategy based on unsupervised cycle-consistent training will be developed to transfer the model to real OBN field datasets without full re-training. Second, the architecture will be extended to process multicomponent OBN recordings by incorporating vector-valued input channels for

horizontal geophone components. Third, physics-informed regularization terms encoding known OBN wavefield relationships will be introduced into the loss function to improve generalization on structurally complex datasets. The source code and synthetic dataset will be made publicly available upon acceptance to facilitate reproducibility and community benchmarking.

REFERENCES

1. Barr, F.J. and Sanders, J.I. (1989) Attenuation of water-column reverberations using pressure and velocity detectors in a water-bottom cable. SEG Technical Program Expanded Abstracts, 653–656.
2. Amundsen, L. (1993) Wavenumber-based filtering of marine point-source data. *Geophysics*, 58(9), 1335–1348.
3. Ozbek, A., Vassallo, M., Ozdemir, K., van Manen, D.J., and Eggenberger, K. (2010) Crossline wavefield reconstruction from multicomponent streamer data: Part 2 — Joint interpolation and 3D up/down separation by generalized matching pursuit. *Geophysics*, 75(6), WB69–WB85.
4. Schalkwijk, K.M., Wapenaar, C.P.A., and Verschuur, D.J. (2003) Adaptive decomposition of multicomponent ocean-bottom seismic data into downgoing and upgoing P- and S-waves. *Geophysics*, 68(3), 1091–1102.
5. Zhu, W., Mousavi, S.M., and Beroza, G.C. (2019) Seismic signal denoising and decomposition using deep neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 57(11), 9476–9489.
6. Liu, D., Wang, W., and Chen, W. (2018) Random noise attenuation using convolutional neural network. SEG Technical Program Expanded Abstracts, 2180–2184.
7. Yuan, S., Liu, J., Wang, S., Wang, T., and Shi, P. (2018) Seismic waveform classification and first-break picking using convolution neural networks. *IEEE Geoscience and Remote Sensing Letters*, 15(2), 272–276.
8. Dramsch, J.S. and Luthje, M. (2018) Deep-learning seismic facies on state-of-the-art CNN architectures. SEG Technical Program Expanded Abstracts, 2036–2040.
9. Zhao, T. (2018) Seismic facies classification using different deep convolutional neural networks. SEG Technical Program Expanded Abstracts, 2046–2050.
10. Hochreiter, S. and Schmidhuber, J. (1997) Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
11. Wiszniowski, J., Plesiewicz, B., and Trojanowski, J. (2014) Application of real time recurrent neural network for detection of small natural earthquakes. *Acta Geophysica*, 62(2), 469–485.
12. Wang, B., Zhang, N., Lu, W., and Wang, J. (2019) Deep-learning-based seismic data interpolation: A preliminary result. *Geophysics*, 84(1), V11–V20.
13. Maiti, S., Tiwari, R.K., and Kümpel, H.J. (2007) Neural network modelling and classification of lithofacies using well log data: A case study from KTB borehole site. *Geophysical Journal International*, 169(2), 733–746.
14. Huang, L., Dong, X., and Clee, T.E. (2017) A scalable deep learning platform for identifying geologic features from seismic attributes. *The Leading Edge*, 36(3), 249–256.
15. Sun, H., Demanet, L., and Willis, M. (2020) Deep learning for seismic wavefield separation. *Geophysics*, 85(6), WA17–WA26.
16. Perol, T., Gharbi, M., and Denolle, M. (2018) Convolutional neural network for earthquake detection and location. *Science Advances*, 4(2), e1700578.
17. Zhong, Z., Sun, A.Y., and Jeong, H. (2019) Predicting CO₂ plume migration in heterogeneous formations using conditional deep convolutional generative adversarial network. *Water Resources Research*, 55(7), 5830–5851.
18. Zhang, X., Liu, Y., and Chen, Y. (2021) Time-frequency domain deep learning for OBN seismic separation. *Journal of Geophysics and Engineering*, 18(4), 512–525.
19. Baillard, C., Crawford, W.C., Ballu, V., Hibert, C., and Mangeney, A. (2014) An automatic kurtosis-based P- and S-phase picker designed for local seismic networks. *Bulletin of the Seismological Society of America*, 104(1), 394–409.
20. Mousavi, S.M., Zhu, W., Sheng, Y., and Beroza, G.C. (2019) CRED: A deep residual network of convolutional and recurrent units for earthquake signal detection. *Scientific Reports*, 9, 10267.

21. Li, J., Chen, W., and Liao, X. (2022) Attention-gated recurrent network for ocean-bottom seismic wavefield separation. *Geophysical Prospecting*, 70(5), 924–93.