Petroleum Science 23 (2026) 4717–4734

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Original Paper

# Diffusion-IDEA: A conditional diffusion framework for well log imputation

Cheng Feng [a,*], Guo-Zeng Zhang [a], Zi-Yan Feng [b], Jie Yu [c], Meng-Ru Hu [a], Ke-Gang Ling [d]

[a] Faculty of Petroleum, China University of Petroleum-Beijing at Karamay, Karamay, 834000, Xinjiang, China
[b] School of Geophysics, China University of Petroleum-Beijing, Beijing, 102249, China
[c] Supervisory Center, PetroChina Xinjiang Oilfield Company, Karamay, 834000, Xinjiang, China
[d] University of North Dakota, Grand Forks, USA

ARTICLE INFO

*Article history:*
Received 7 November 2025
Received in revised form
20 March 2026
Accepted 5 May 2026
Available online 7 May 2026

Edited by Jie Hao

*Keywords:*
Well log reconstruction
Conditional imputation
Thresholded cross attention (TCA)
Adaptive filtering
Interpretability

ABSTRACT

Well log data constitute a primary source of critical information for quantitative reservoir evaluation; however, acquisition failures, wellbore irregularities, and constraints imposed by the log program often result in missing intervals or varying degrees of discontinuity within the recorded well log curves. In recent years, data-driven predictive paradigms have emerged as an effective and increasingly prevalent avenue for reconstructing incomplete well logs; nevertheless, because the process is modulated by stratified lithological heterogeneity, thin-bed phenomena, abrupt depositional interfaces, and inherent petrophysical and tool-response constraints, the task is more rigorously framed as a conditional imputation problem rather than a naive regression formulation. To address this limitation, we propose Diffusion-IDEA, an entirely new framework for well log reconstruction. The framework employs the observed log curves together with geological priors as conditioning constraints and, leveraging a denoising diffusion probabilistic model, performs conditional reconstruction of the missing segments. Unlike existing diffusion-based methods, we implement three key components within our diffusion model: (i) Dual Mamba; (ii) Thresholded Cross Attention (TCA); and (iii) an Adaptive Filtering and Interpretation Architecture (IDEA). We apply adaptive filtering to modulate the signal-to-noise ratio (SNR) of the observed log signals, while the Dual Mamba module fuses forward and backward representations to provide robust conditioning information. During decoding, the cross-attention module filters salient associations using a cumulative probability threshold; the final output is explicitly decomposed into trend and frequency components to enhance interpretability. Experimental results demonstrate that the proposed model significantly outperforms all existing baselines across every evaluation metric. Compared with the strongest competing method, Diffusion-IDEA achieves state-of-the-art (SOTA) performance, with a relative reduction in MAPE of approximately 22% to 58% across various well-logging parameters.


## 1. Introduction

Well logs continuously record formation petrophysical characteristics along the borehole. They provide fundamental data for lithology discrimination, petrophysical parameter estimation, and formation boundary delineation. They also constrain reservoir-scale geological modeling and stratigraphic correlation analyses (Zhou et al., 2025). However, borehole instability, mud filtrate invasion, inadequate tool contact, and operational or recording interruptions often degrade the measured logs. The logs may then exhibit isolated point dropouts, continuous missing intervals, or complete loss of an entire log suite (Al-Fakih et al., 2025; Haritha et al., 2025; Zhang et al., 2024a,b).

In recent years, deep learning–based sequence models have achieved notable progress in well log reconstruction. Current mainstream training–validation strategies fall into three categories: (i) sequential or random partitioning of a single-well dataset for training and testing (Fan et al., 2021; Li and Jiang, 2025; Ren et al., 2023); (ii) joint training on data from multiple wells with evaluation on blind wells in the same area (Li et al., 2023; Zhang et al., 2024a,b); (iii) splits grouped by lithology or similar formations to maintain distributional consistency between training and testing sets (Zhang et al., 2025). Based on the above

* Corresponding author.
*E-mail address:* fcvip0808@126.com (C. Feng).
Peer review under the responsibility of China University of Petroleum (Beijing).

https://doi.org/10.1016/j.petsci.2026.05.001

partitioning strategies, existing methods often directly adapt temporal sequence models into depth-indexed (depth-as-sequence) networks. They therefore perform only pointwise predictions. This makes them unable to capture stratigraphic-scale long-range structure and inter-well distribution differences. Model reliability and physical consistency are limited in cross-well extrapolation and in scenarios with long contiguous missing intervals.

LSTM networks and their variants are widely adopted in well log reconstruction. Their gated memory architecture is inherently suited to depth-varying sequential signals (Cheng et al., 2022; Li et al., 2025). Zhang et al. (2022) enhanced model expressive capacity by deepening an improved BiLSTM network, enabling effective reconstruction of long missing intervals. Shang et al. (2022) jointly modeled domain knowledge with LSTM, constructing a neural network with scenario knowledge constraints and providing a new perspective for well log reconstruction. Liu et al. (2024a,b) proposed a reconstruction method based on component-wise LSTM representations. By decomposing the log signals, it achieved lower prediction error. However, limited by its recursive mechanism, LSTM experiences memory decay and over-smoothing in ultra-long sequence settings. Consequently, subsequent studies introduced the Transformer attention mechanism into well log modeling. Attention does not rely on stepwise propagation; it enables global contextual interaction through weighted aggregation of similarities between arbitrary positions in the sequence (Vaswani et al., 2017). Mei et al. (2025) combined BiLSTM with multi-head attention to further extract key geological response features. Huang et al. (2025) proposed a Transformer–LSTM hybrid model that effectively captures long-range dependencies in well log data. Wang et al. (2024) and Shan et al. (2021) respectively incorporated attention into TCN–BiGRU and CNN–BiLSTM and conducted reconstruction experiments on field-measured missing data from the study area. These attention-based modeling approaches mitigate catastrophic forgetting. However, both time and GPU memory overhead grow exponentially (Lee et al., 2024; Su et al., 2025). Large language models (LLMs) and diffusion models are experiencing a surge in cross-domain applications. X. Chen et al. (2025) fine-tuned a pretrained LLM for well log reconstruction. However, the mapping between the pretraining corpus and blind-well data is opaque. This unconditional cross-regional transfer makes the provenance of the reconstruction results difficult to explain. Han et al. (2024) proposed a conditional denoising diffusion probabilistic model, extending well-log imputation from deterministic prediction to uncertainty quantification. Z. Chen et al. (2025) further employed diffusion models for the generation and augmentation of synthetic seismic-wave datasets, achieving accurate matching under complex physical and geological constraints. Owing to their distinctive training paradigm, diffusion models provide a new modeling framework for intelligent geophysical processing; nevertheless, how to endow such models with explicit, physically grounded interpretability remains a critical open problem.

To overcome the above challenges, this study proposes a reconstruction model. It treats the observed well logs and geological priors as conditional constraints. It conditionally reconstructs the missing intervals using a denoising diffusion probabilistic model (DDPM). Specifically, as illustrated in Fig. 1, the model performs iterative reverse diffusion to recover a stratigraphic distribution that is consistent with observations and satisfies geological priors. The main encoder adopts Dual Mamba to realize a linear-complexity attention-like mechanism, meeting our

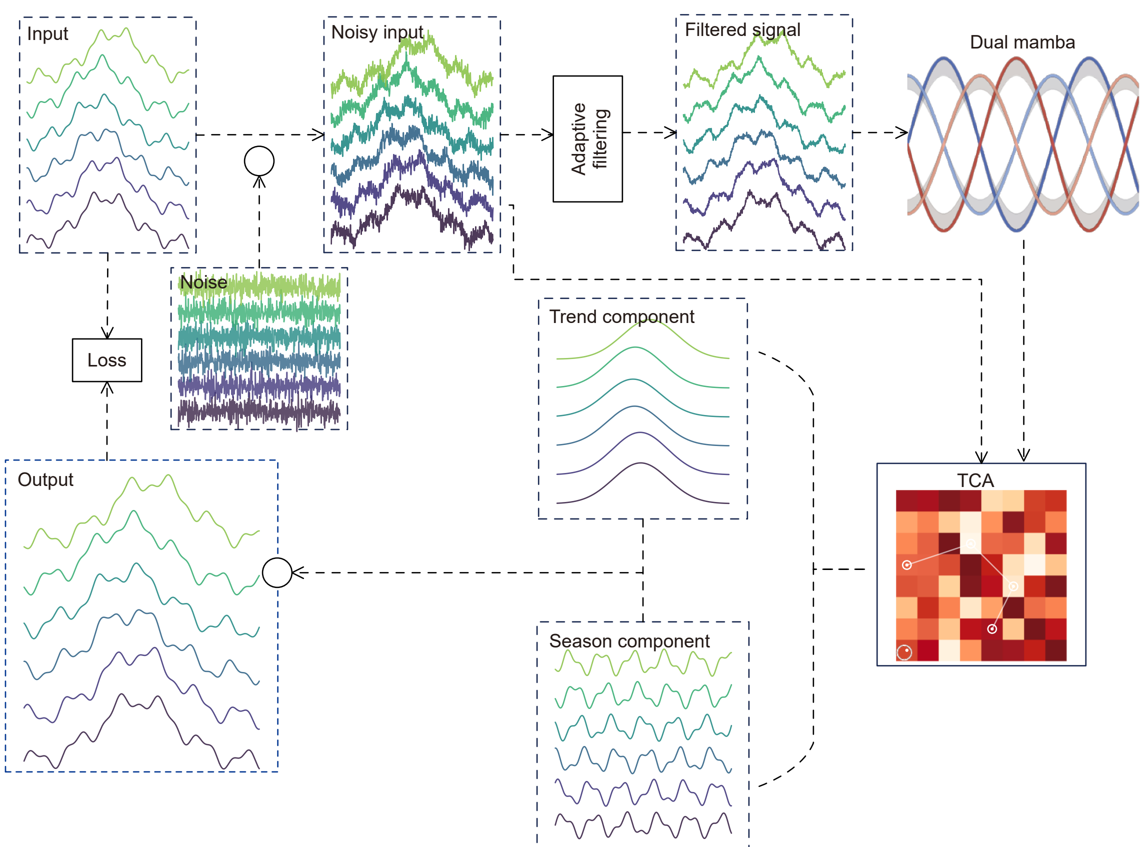


**Fig. 1.** Overall workflow of diffusion-IDEA.

long-sequence modeling requirement. Meanwhile, we introduce a TCA module in the decoder to extract key weights across depth steps and mitigate the adverse impact of redundant elements. An adaptive filtering module then dynamically modulates the SNR ratio of inputs at each denoising stage, providing a more explicit and reliable decomposition pathway for the subsequent interpretability framework. To our knowledge, this is the first end-to-end, interpretable, conditional well log reconstruction approach built on denoising diffusion. Experimental results show that the proposed model attains advanced performance on field data from the study area.

## 2. Methods

The proposed Diffusion-IDEA architecture is shown in Fig. 2. The encoder couples an adaptive filter with Dual Mamba. The decoder comprises a cascaded TCA module and an interpretability decomposition module. To inject lithology labels effectively, we employ a label-conditioned adaptive normalization (Ada-LayerNorm). It maps lithology conditioning information to per-channel scale and shift parameters that dynamically modulate the network representations, forming a learnable geological prior.

### 2.1. Adaptive filter (DEA)

The reverse diffusion process essentially recovers signal from noise (Fig. 2(a)). Without spectral priors, the denoising search space becomes vast, which impedes progress toward the true data distribution. Meanwhile, from a signal-processing perspective, well-logging data exhibit pronounced nonstationarity and multi-scale characteristics. The low-frequency components typically capture the large-scale trend (baseline) associated with sedimentary-environment evolution and compaction effects, whereas the high-frequency components are coupled with local stratigraphic heterogeneity, fine-scale responses to lithological variations, and sensor noise. Conventional filtering approaches therefore struggle to suppress noise effectively while preserving high-fidelity reconstruction of the true geological signal. To address this, we apply adaptive frequency-domain filtering before the main encoder to explicitly perform spectral selection and weighting. It guides the model in a sample-adaptive manner by emphasizing high-weight spectral components (Liu et al., 2024a,b; Wu et al., 2025).

The design motivation of DEA is to mimic the visual interpretation habit of geoscience experts when analyzing well logs, i.e., grasping the global trend while being sensitive to diagnostically meaningful local discontinuities (Fig. 2(b)). Compared with conventional signal analysis methods, this mechanism is fundamentally different. Relative to wavelet transform, wavelet denoising relies heavily on a predefined static mother wavelet basis function. Owing to the complexity of stratigraphic lithologic discontinuity interfaces, a fixed mathematical basis often leads to artifact effects because of basis mismatch. In contrast, the present mechanism does not require any predefined basis functions; instead, it relies on the model to adaptively learn the optimal complex weights for each frequency band. Furthermore, compared with empirical mode decomposition (EMD), although EMD is adaptive, it strictly depends on local extrema to construct envelope curves and is therefore highly susceptible to high-frequency logging noise, which can induce mode overlap. By contrast, DEA performs explicit high- and low-frequency separation and learnable truncation in the global Fourier domain, thereby overcoming the theoretical deficiency of mode mixing. The specific implementation of the algorithm is as follows: we perform the Fourier transform (rFFT) on the input signal ($\mathbf{X}_i$) to obtain its frequency-domain representation ($\mathbf{X}_{\text{fft}}$). We then apply adaptive selection and separation, yielding the high- and low-frequency components ($\mathbf{X}_{\text{low}}$,$\mathbf{X}_{\text{high}}$), as shown in Eqs. (1) and (2). Next, each branch applies learnable frequency-domain operators to adjust the spectral responses; the components are then merged, and the inverse transform (irFFT) is performed, as shown in Eqs. (3) and (4). Through the above adaptive filtering and truncation, the model can dynamically learn, according to the SNR of different depth intervals, to distinguish geologically meaningful signals such as thin interbeds from measurement and stochastic noise. Meanwhile, the frequency domain prior guidance effectively constrains the search space of the diffusion model in complex geological settings, and provides the essential domain knowledge prior required for the subsequent interpretable decomposition.

$$\mathbf{X}_{\text{low}} = \mathbf{X}_{\text{fft}} \cdot \mathrm{I}(|f| \le f_{\text{low}}) \tag{1}$$

$$\mathbf{X}_{\text{high}} = \mathbf{X}_{\text{fft}} \cdot \mathrm{I}\left(|f| \ge f_{\text{high}}\right) \tag{2}$$

$$\forall i \in \{\text{fft}, \text{low}, \text{high}\} \parallel \mathbf{X}_{\text{out},i} = f_{\text{cReLU}}\left(\mathbf{X}_i \cdot \mathbf{W}_{i,1}\right) \cdot \mathbf{W}_{i,2} \tag{3}$$

$$\mathbf{X}_{\text{out}} = \text{irFFT}\left(\mathbf{X}_{\text{out,weighted}} = \mathbf{X}_{\text{out,fft}} + \mathbf{X}_{\text{out,low}} + \mathbf{X}_{\text{out,high}}\right) \tag{4}$$

Here, $f, f_{\text{low}}, f_{\text{high}}$ denote the per-bin frequency, the low-pass cutoff, and the high-pass cutoff, respectively. $\mathrm{I}(\cdot)$ serves as an indicator function that selects frequency components meeting the specified conditions. The parameters ($\mathbf{W}_{i,1}$, $\mathbf{W}_{i,2}$) are learnable complex-valued weights. $f_{\text{cReLU}}$ denotes the complex ReLU activation. Concurrently, in Eqs. (3) and (4)., the branch index $\forall i \in \{\text{fft}, \text{low}, \text{high}\}$ denotes the independent processing streams for global, low-frequency, and high-frequency components, respectively. The subscript out further identifies the resultant states of these signals following the adaptive extraction process.

### 2.2. Dual Mamba

Existing work still faces a dilemma in long-sequence modeling (see Introduction, paragraph 3). On the other hand, prior studies indicate that the conventional attention family lacks awareness of recursive reasoning across timesteps, which limits effective learning on long-sequence tasks (Wang et al., 2025). Moreover, stratigraphic depositional sequences typically exhibit strict depthwise causal relationships, and well log interpretation often requires integrated analysis over longer depth intervals, for example by jointly considering the overlying and underlying surrounding formations (Dong et al., 2023a, 2023b). Motivated by the above limitations, we propose a state-space–based framework that achieves near-linear, selective, bidirectional recursive reasoning: Dual Mamba (Fig. 2(c)) (Patro and Agneeswaran, 2025). We define the full hidden-state inference process as $\mathbf{H} = \{\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3, \ldots, \mathbf{h}_n\}$. The state transition matrix $\overline{\mathbf{A}}$ determines the extent of memory transfer across timesteps ($T = \{t_1, t_2, t_3, \ldots, t_n\}$). This enables a form of selective emphasis consistent with the role of attention. The input mapping matrix $\overline{\mathbf{B}}$ projects the current input $\mathbf{x}_t$ into the state space. The complete formulation is given in Eq. (5). Its recursive representation also implies that the depth-varying causal relationships are preserved, while meeting the requirement for efficient modeling of long sequences. Intuitively, bidirectionality yields more comprehensive contextual information than a unidirectional design for learning data distributions. Accordingly, we fuse Mamba's forward and backward branches by summation and take the result as the encoder output $\mathbf{En}_{\text{out}}$, as shown in Eq. (6).

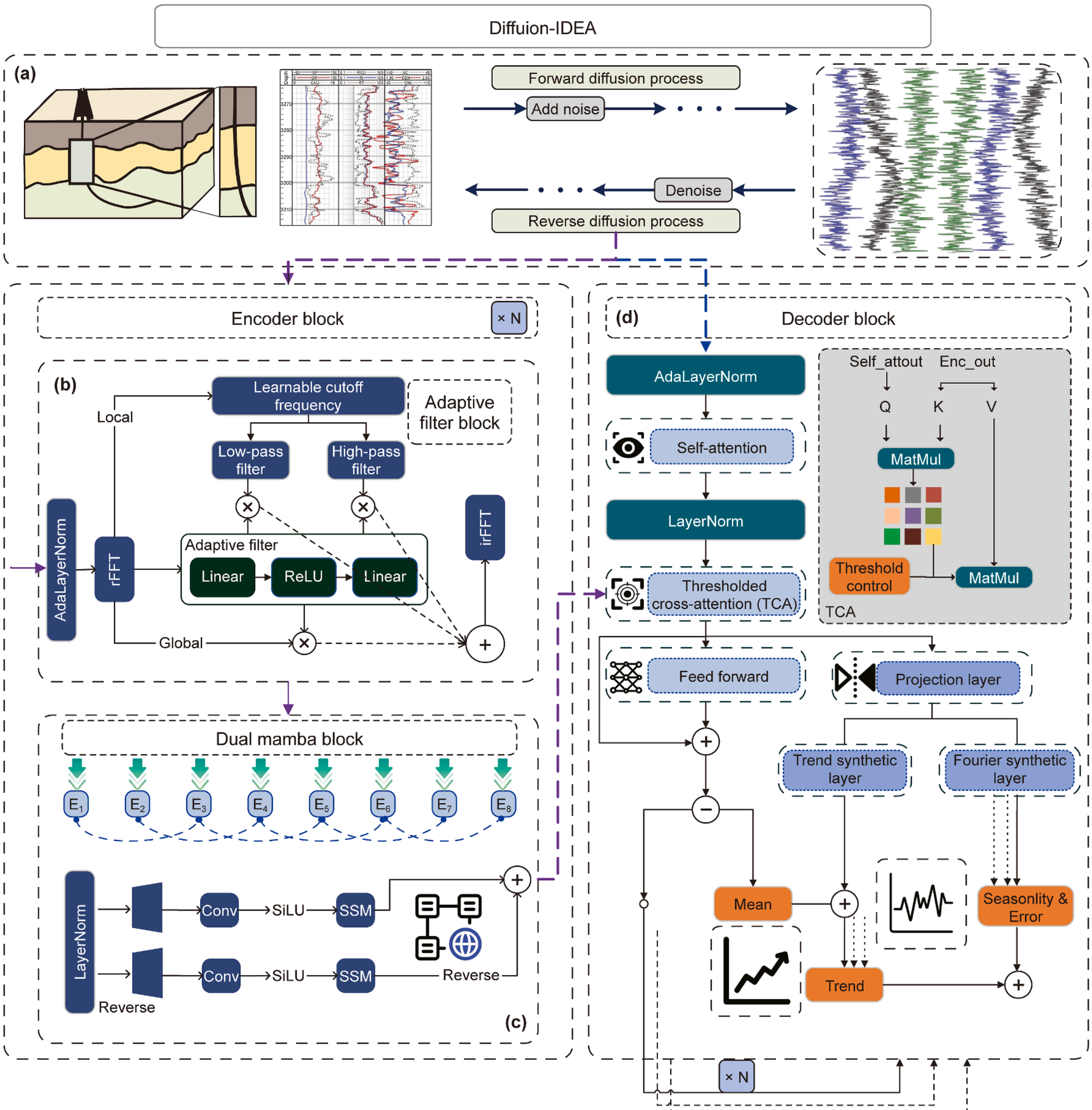


**Fig. 2.** Architecture of the diffusion IDEA model: (**a**) diffusion process; (**b**) Adaptive filter block; (**c**) Dual Mamba block; and (**d**) decoder architecture comprising the TCA and interpretable decomposition modules.

$$\mathbf{h}_t = \overline{\mathbf{A}} \cdot \mathbf{h}_{t-1} + \overline{\mathbf{B}} \cdot \mathbf{x}_t \quad (5)$$

$$\mathbf{En}_{\text{out}} = \text{Mamba}(\mathbf{X}_{\text{out}}) + \text{Mamba\_r}(\mathbf{X}_{\text{out}}) \quad (6)$$

Among them, Mamba$(\cdot)$ represents the forward pass during training, Mamba_r$(\cdot)$ represents the backward pass during training, and $\mathbf{X}_{\text{out}}$ denotes the feature representation output by the DEA.

## 2.3. TCA and interpretability decomposition

Intuitively, Mamba preserves effective long-term memory under aggressive compression in long-sequence modeling. However, within the decoder-side decomposition framework, we seek to retain complete full-sequence information. Inspired by prior work, we retain in the decoder the mechanism that routes $\mathbf{X}_{\text{out}}$ back into self-attention. We use it as the query ($\mathbf{Q}$) in cross-attention, while $\mathbf{En}_{\text{out}}$ provides the key and value vectors ($\mathbf{K},\mathbf{V}$) that are fed into TCA (Fig. 2(d)) (Tong et al., 2024; Qin et al., 2024). However, in complex stratigraphy, the logging response at a given depth is typically strongly correlated with only a few genetically related horizons, while exhibiting negligible dependence on the background strata. Conventional full attention introduces a large number of spurious low-weight correlations, leading to oversmoothing in the feature representations. We therefore introduce two key modifications to prune low-confidence long-tail connections and retain only sparse nodes with statistically significant genetic associations: First, we accumulate attention weights and apply thresholding. This prunes low-contribution weights that are highly irrelevant to the query. Second, we propagate accumulated attention weights across layers and mask sub-threshold weights. Subsequent layers attend to and compute only on nodes retained by preceding layers. This yields sparsified attention-score computation. Formally, given the query matrix ($\mathbf{Q}^{(l)}$) at the $l$-th layer and the key and value matrices output by the encoder, the raw attention weight matrix of the current layer ($\mathbf{A}^{(l)}$) is computed as shown in Eq. (7), where $L$ and $S$

denote the sequence lengths of the query and key, respectively, and $d$ is the scaling factor. Meanwhile, we define the average distribution over all attention heads in the previous layer ($\overline{A}_{ij}^{(l-1)}$) as the structural prior for the current layer, as given in Eq. (8).

$$\mathbf{A}^{(l)} = \mathrm{softmax}\left(\frac{\mathbf{Q}^{(l)}\mathbf{K}^{\mathrm{T}}}{\sqrt{d}}\right) \in \mathbb{R}^{L\times S} \tag{7}$$

$$\overline{A}_{ij}^{(l-1)} = \frac{1}{H}\sum_{h=1}^{H}\widehat{A}_{(h,ij)}^{(l-1)} \tag{8}$$

Subsequently, for each query $i$, the prior distribution from the previous layer ($\overline{A}_{ij}^{(l-1)}$) is sorted in descending order, denoted by $\overline{A}_{i,(1)}^{(l-1)}, \overline{A}_{i,(2)}^{(l-1)}, \ldots, \overline{A}_{i,(S)}^{(l-1)}$. We then define the minimal retention set $\Omega_i$, i.e., the original indices corresponding to the top-$k_i$ keys after sorting, such that the predefined cumulative probability threshold ($\tau \in (0, 1]$) is satisfied, as formulated in Eq. (9). On this basis, a binary mask is further constructed and the attention weights are renormalized, yielding the sparsified weights ($\widehat{\mathbf{A}}$), as shown in Eq. (10). Here, $\mathrm{I}(\cdot)$ denotes the indicator function, and $\epsilon = 10^{-8}$ represents the numerical stability term. The final output is given by $\mathbf{Y}^{(l)} = \widehat{\mathbf{A}}^{(l)}\mathbf{V}$.

$$k_i = \min\left\{k : \sum_{j=1}^{k}\overline{A}_{i,(j)}^{(l-1)} \geq \tau\right\} \tag{9}$$

$$\widehat{A}_{ij}^{(l)} = \frac{A_{ij}^{(l)}\cdot \mathrm{I}(j \in \Omega_i)}{\sum\limits_{m\in\Omega_i} A_{im}^{(l)} + \epsilon} \tag{10}$$

The decomposition module guides Diffusion-IDEA to capture semantic information in time series. Specifically, we linearly project the TCA output into two representations: they are fed into the Trend module to extract the trend component ($\mathbf{trend}_i$) and into the Seasonal module to extract the seasonal component ($\mathbf{Season}_i$). In parallel, at each layer we retain the mean over the depth dimension of that layer's hidden representation and apply a linear mapping to it (**Mean_stack**) (Yuan and Qiao, 2024). Finally, we fuse $\mathbf{trend}_i$ and **Mean_stack** from all layers to obtain the final trend output (**Trend_total**). For the seasonal component, we perform a cumulative sum over layers of $\mathbf{Season}_i$, add the final de-meaned residual (**res**), and map the result back to the original dimension via an inverse projection to produce the final frequency output (**Season_error**). In this process, the Trend module absorbs the global baseline trend induced by sedimentary-environment evolution and compaction effects, whereas the Season module focuses on learning representations of relative lithologic fluctuations and high-frequency bed-boundary responses that are independent of absolute depth. The model's final reconstruction (**Model_out**) is obtained by summing **Trend_total** and **Season_error**, yielding an interpretable decomposition and reconstruction of the sequence, as shown in Eqs. (11) and (12) and (13).

$$\mathbf{Trend\_total} = \mathrm{Combine\_m}(\mathbf{Mean\_stack}) + \sum \mathbf{trend}_i \tag{11}$$

$$\mathbf{Season\_error} = \mathrm{Combine\_s}\left(\sum \mathbf{Season}_i\right) + \mathbf{res} \tag{12}$$

$$\mathbf{Model\_out} = \mathbf{Trend\_total} + \mathbf{Season\_error} \tag{13}$$

Here, Combine_m is used to project along the layer dimension to obtain the final mean, and Combine_s is used to project the seasonal signal in the embedding space back to the original feature space. $\sum \mathbf{trend}_i$ represents the cumulative sum of trends across layers, while $\sum \mathbf{Season}_i$ represents the cumulative sum of seasonal components across layers.

## 3. Experiments

In this section, we first describe the data processing methodology and experimental setup, report the main experimental results, and present comparative experiments, ablation analyses, and other implementation-related visualizations.

### 3.1. Data description

The well log dataset used in this study comes from the Mahu Depression, situated in the Junggar Basin of the Xinjiang Uygur Autonomous Region, northwestern China (Fig. 3(a)). The basin evolved through three stages-a foreland basin, a downwarped intracontinental depression, and a reactivated foreland basin-and hosts prolific hydrocarbon accumulations (Fig. 3(b)). As shown in Fig. 3(c), the study area is bounded to the north by the Wuxia Fault Zone and to the northwest by the Kebai Fault Zone; the Zhongguai Uplift lies to the southwest, the Dabasong Uplift to the south, the Xiayan Uplift to the east, and the Luliang Uplift to the northeast. The overall structure is a gently southeast-dipping monocline, with locally developed low-amplitude structural platforms, anticlines, and structural noses. The target interval investigated in this work is the Triassic succession within the northern slope belt of the Mahu Depression. Within the Triassic, the stratigraphic units exhibit fundamental differences in sedimentary facies, diagenetic evolution, and downhole log environments. Consequently, the raw log responses show varying degrees of systematic distortion and information loss across stratigraphic levels.

Specifically, the Baijiantan Formation consists of thick mudstone packages and shows high natural gamma ray (GR), density (DEN), and compensated neutron log (CNL) readings; the clay-rich background readily causes density–neutron overlap (DEN-CNL overlap) and porosity masking; at the top of the Karamay Formation, GR decreases markedly, forming a clear stratigraphic contrast, but variations in lamination and sandbody scale introduce thin-bed averaging and shoulder effects; the Baikouquan Formation develops three sand packages, and the reservoirs display a blocky, medium-resistivity character accompanied by coarse to conglomeratic textures, which tend to induce borehole instability, tool eccentricity, and mud filtrate invasion; in the Lower Wuerhe Formation, shale resistivity is notably low, and differences in the distribution of conductive phases and invasion depth lead to nonuniform electrical responses. These differences affect petrophysical interpretation within individual stratigraphic units and, at layer interfaces, give rise to boundary effects and cross-layer model discontinuities. The resulting complexity increases the difficulty of interpretation while simultaneously imposing stricter demands on the generalization of well log data imputation and reconstruction algorithms. Consequently, this dataset is well suited to rigorously validating the proposed model's adaptability and effectiveness under real, geologically complex conditions.

In the main experiments, we used Triassic well logs from MH133 and randomly split them into training and test sets for the imputation task. The blind-well experiment jointly trained on MH133 together with four nearby offset wells (five wells in total) and tested on the Triassic data from MH13. The details of the training dataset are summarized in Table 1. The dataset spans a total stratigraphic thickness of 6844 m and contains 544 labeled stratigraphic intervals. We treated the following conventional logs as variable units: depth, shallow resistivity (RXO), density (DEN),

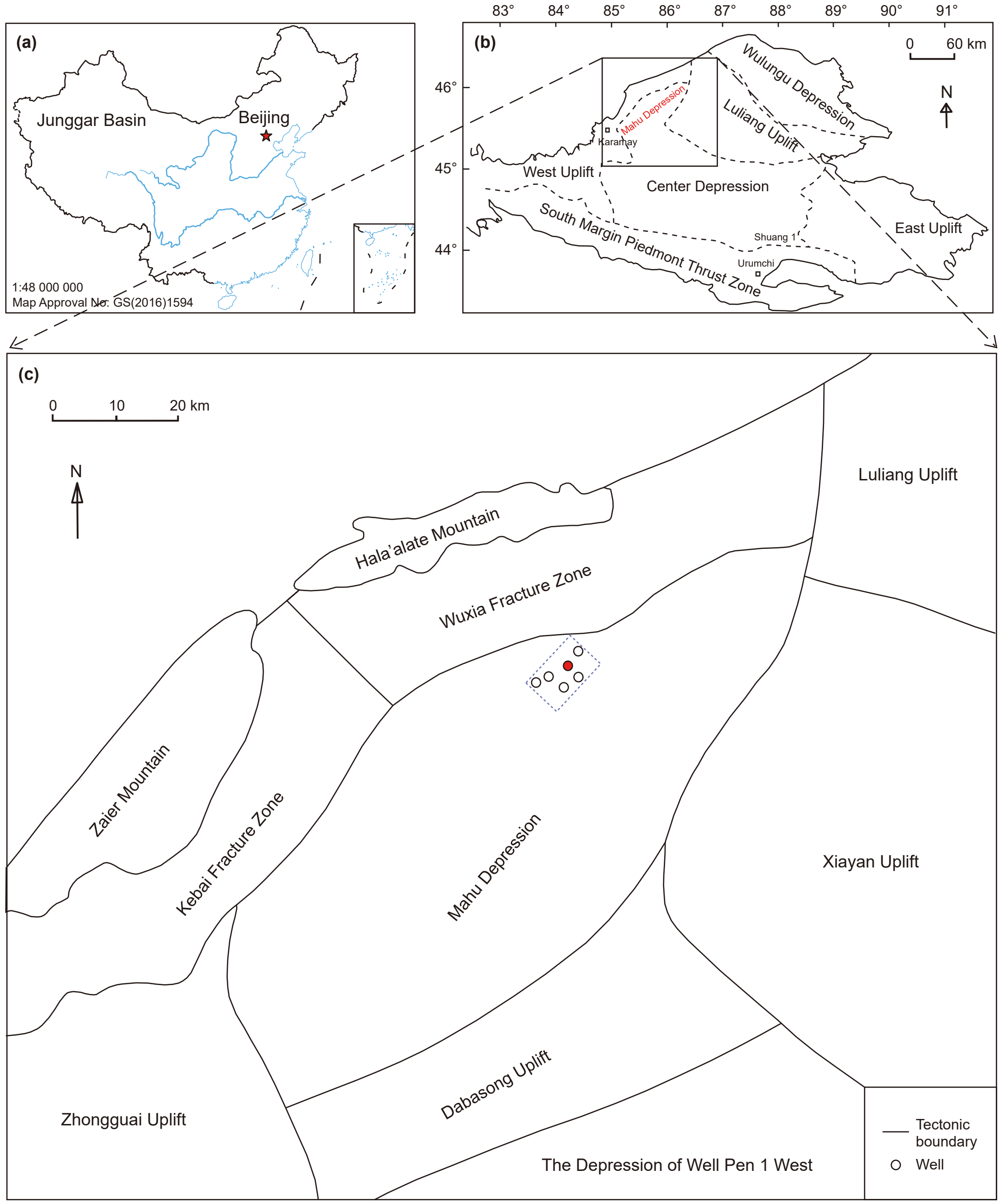


**Fig. 3.** Geological overview of the Mahu Depression (Red circles indicate blind wells): (**a**) shows the regional geographic location of the Junggar Basin in northwestern China; (**b**) displays the secondary tectonic units within the Junggar Basin; and (**c**) illustrates the specific geological boundaries of the Mahu Depression and its adjacent areas.

**Table 1**
Summary of the training set samples.

| Well | Stratigraphic thickness, m | No. of samples, layers | Key lithology features |
|---|---|---|---|
| MH133 | 1270 | 98 | Mudstone, Siltstone |
| Neighbor well 1 | 1150 | 85 | Mudstone, Fine sandstone |
| Neighbor well 2 | 1320 | 105 | Siltstone, Conglomerate |
| Neighbor well 3 | 1490 | 118 | Sandy mudstone, Conglomerate |
| Neighbor well 4 | 1614 | 138 | Mudstone, Siltstone |

compensated neutron log (CNL), natural gamma ray (GR), deep resistivity (RT), medium resistivity (RI), acoustic (AC), and caliper (CALI). In addition, a sliding window approach is employed to randomly mask the original data and perform local slicing. The window length is set to 48, with a stride of 1. Compared with some related studies, a length of 48 is longer and can cover one or more layer boundaries as well as intra-layer features, thereby providing richer stratigraphic context for the model. A stride of 1 ensures a

**Table 2**
Breakdown of observed versus missing samples.

| Metric | RXO | DEN | CNL | GR | RT | RI | AC | CALI |
|---|---|---|---|---|---|---|---|---|
| Observed points | 230 | 230 | 221 | 243 | 260 | 227 | 239 | 242 |
| Missing points | 154 | 154 | 163 | 141 | 124 | 157 | 145 | 142 |

high degree of overlap between adjacent windows, substantially increasing the number of training samples and preserving continuity along depth, which helps the model better identify local anomalies and transition zones and achieve smoother reconstructions.

### 3.2. Experimental details

During training, we adopt an extended sequence-conditional diffusion generative framework. For feature extraction, we employ a 4-layer DEA together with Dual Mamba as the encoder. The decoder uses a 3-layer TCA (threshold set to 0.8) for interpretable decomposition. The diffusion process follows a cosine beta schedule with 1000 steps. The initial learning rate is set to 1e-5 with the Adam optimizer; training is capped at 25,000 steps with gradient accumulation every 2 steps. We use a ReduceLROnPlateau scheduler combined with a warmup mechanism: a warmup learning rate of 8e-4 for 500 steps, a reduction factor of 0.5, and a patience of 5000 steps. We also apply an exponential moving average (EMA) with a decay factor of 0.995, updating the EMA weights every 10 steps. The loss function comprises an L1 loss and a weighted Fourier-domain loss (i.e., computing the L1 error in the frequency domain; optional). All experiments were conducted on an NVIDIA RTX 3090 GPU.

### 3.3. Main experiments

We first report, in Table 2, the sample counts of observed versus missing entries in the test set, with missing values accounting for approximately 40%. The comparison between Diffusion-IDEA and the other baseline models is summarized in Table 3, with all metrics de-normalized to the original physical scale. Evidently, diffusion-based, conditionally constrained reconstruction methods are more competitive than purely algorithmic models, which is consistent with our earlier view that curve reconstruction should be treated as conditional reconstruction rather than prediction. Moreover, owing to a series of improvements made to the baselines, Diffusion-IDEA achieves superior results. For a fair comparison, the baselines' data preprocessing and common training hyperparameters were kept consistent with those of Diffusion-IDEA.

Table 3 tabulates the performance of seven models (Diffusion-IDEA, Diffusion-TS, Transformer, Dual Mamba, Kriging, Co-Kriging, and PINN) across three error metrics-RMSE, MAE, and MAPE-on eight curves (RXO, DEN, CNL, GR, RT, RI, AC, CALI). Figs. 4–6 depict reconstruction results for representative curves (RXO, GR, AC): the top and bottom panels respectively display the complete reconstructions by Diffusion-IDEA and Diffusion-TS, with the corresponding scatter plots centered between them. Diffusion-IDEA attains the lowest errors on all eight curves. Relative to the second-best baseline (Diffusion-TS), it reduces the overall average RMSE by 35.61% and the MAE by approximately 37.87%. In addition, the Diffusion family exhibits consistent error attenuation across test sequences, attributable to its stepwise diffusion sampling strategy used during training. Through progressive training, the model mitigates inter-layer heterogeneity and captures window-level contextual dependencies; conditioning imposed during sampling preserves observed information and steers the reverse diffusion process, yielding high-fidelity reconstructions. Notably, comparisons with traditional geostatistical baselines (Kriging and Co-Kriging) and PINN show that the reconstruction accuracy of Kriging is highly prone to degradation when confronted with complex missing patterns. Although Co-Kriging slightly reduces the error through spatial co-constraints, it is still constrained by linear variogram functions and therefore cannot capture the highly nonlinear mapping manifold among multiple physical logging variables, causing its performance to remain far inferior to that of deep learning baselines. In addition, although PINN, which incorporates classical petrophysical empirical formulas as penalty terms, performs significantly better than traditional mathematical interpolation methods, its overall performance still falls short of methods with high-precision reconstruction capability. This phenomenon can be attributed to stratigraphic heterogeneity: the parameters in various empirical formulas vary drastically across different lithologies and depth intervals, such that forcibly imposing a single

**Table 3**
Model comparison for missing-value imputation by well log curve.

| Model | Metric | RXO | DEN | CNL | GR | RT | RI | AC | CALI |
|---|---|---|---|---|---|---|---|---|---|
| Diffusion-IDEA | RMSE | **2.4500** | **0.0184** | **0.5422** | **0.9309** | **0.6628** | **0.3100** | **0.7728** | **0.0958** |
| | MAE | **1.5476** | **0.0092** | **0.3759** | **0.6691** | **0.4348** | **0.2198** | **0.5029** | **0.0542** |
| | MAPE | **23.95%** | **0.40%** | **1.48%** | **0.78%** | **2.49%** | **1.98%** | **0.63%** | **0.46%** |
| Diffusion-TS | RMSE | <u>3.2887</u> | <u>0.0200</u> | <u>0.9851</u> | <u>1.7384</u> | <u>0.8329</u> | <u>0.5762</u> | <u>1.2442</u> | <u>0.2150</u> |
| | MAE | <u>2.0963</u> | <u>0.0143</u> | <u>0.7347</u> | <u>1.3352</u> | <u>0.5965</u> | <u>0.3902</u> | <u>0.8436</u> | <u>0.1268</u> |
| | MAPE | <u>30.67%</u> | <u>0.60%</u> | <u>2.91%</u> | <u>1.62%</u> | <u>3.49%</u> | <u>3.37%</u> | <u>1.05%</u> | <u>1.10%</u> |
| Transformer | RMSE | 4.0501 | 0.0515 | 2.1640 | 2.9612 | 1.5246 | 0.9137 | 3.4387 | 0.3108 |
| | MAE | 2.8642 | 0.0288 | 1.5024 | 2.1278 | 1.0431 | 0.6674 | 1.8907 | 0.1846 |
| | MAPE | 49.00% | 1.24% | 5.62% | 2.53% | 6.32% | 5.86% | 2.32% | 1.63% |
| Dual Mamba | RMSE | 3.5840 | 0.0303 | 1.5924 | 2.1262 | 1.8500 | 0.7084 | 1.5747 | 0.2171 |
| | MAE | 2.4153 | 0.0216 | 1.1831 | 1.6142 | 1.0262 | 0.5481 | 1.0818 | 0.1309 |
| | MAPE | 45.80% | 0.91% | 4.33% | 1.90% | 4.83% | 4.76% | 1.37% | 1.15% |
| Kriging | RMSE | 10.5142 | 0.1724 | 7.1049 | 12.6415 | 6.8421 | 4.1052 | 9.2545 | 1.2415 |
| | MAE | 8.1524 | 0.1415 | 5.5214 | 9.8514 | 5.1054 | 3.1225 | 7.3512 | 0.9541 |
| | MAPE | 115.6% | 5.23% | 18.52% | 12.65% | 24.51% | 22.15% | 9.45% | 8.65% |
| Co-Kriging | RMSE | 7.8421 | 0.1250 | 4.8215 | 8.4512 | 4.9105 | 2.6514 | 6.7512 | 0.8521 |
| | MAE | 5.9214 | 0.0874 | 3.6541 | 6.5103 | 3.5218 | 1.8420 | 5.1054 | 0.6124 |
| | MAPE | 85.42% | 3.51% | 12.84% | 8.52% | 16.45% | 14.82% | 6.84% | 5.92% |
| PINN | RMSE | 3.8215 | 0.0412 | 1.8463 | 2.5408 | 1.9234 | 0.8152 | 2.1580 | 0.2643 |
| | MAE | 2.6518 | 0.0265 | 1.3142 | 1.8865 | 1.2511 | 0.6033 | 1.4506 | 0.1622 |
| | MAPE | 47.15% | 1.05% | 4.92% | 2.21% | 5.56% | 5.24% | 1.82% | 1.38% |

The bold and underlined values represent the best and second-best results, respectively.

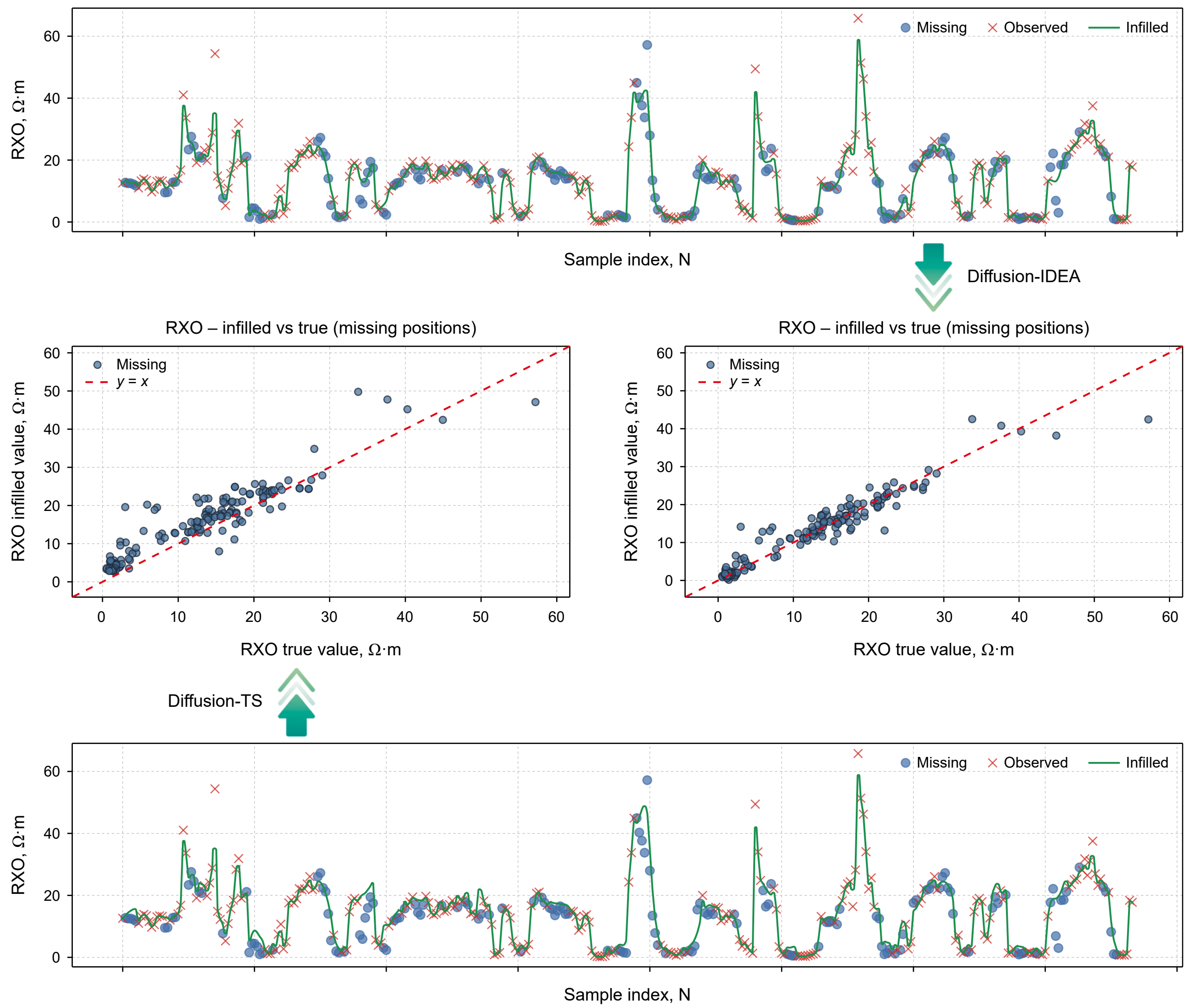


**Fig. 4.** Reconstruction results for missing RXO points.

physical-equation constraint globally can easily introduce rigid physical bias, thereby limiting the network's ability to further optimize under complex lithofacies variations.

### 3.4. Diffusion experiments

To assess whether the diffusion model can adequately learn rapidly varying lithology distributions and to evaluate the effectiveness of our proposed improvements for long-sequence modeling, we compare the denoised reconstructions of the training samples with the original data from three perspectives (Figs. 7 and 8): (i) principal component analysis (PCA) to examine global variance structure; (ii) t-distributed stochastic neighbor embedding (t-SNE) to assess local neighborhood similarity; and (iii) kernel density estimation (KDE) to compare the variables' probability-density functions. We further quantify performance with two evaluation metrics (Table 4): Context FID and Correlational Score. Context-FID is used to verify whether the distribution learned by the model is consistent with that of the training set, and whether the model preserves geological signal fluctuations across different frequency bands. The Correlational Score is used to assess whether the model has captured the physical coupling relationships among different logs, namely their joint distribution, which is critical for well log interpretation.

The results summarized in Table 4 and Figs. 7 and 8 demonstrate that both diffusion-based approaches effectively capture the underlying training distribution. However, Diffusion-IDEA exhibits superior fidelity, as evidenced by its higher evaluation scores. While Diffusion-TS achieves a satisfactory global fit, it manifests a slight deficiency in capturing fine-grained details compared to Diffusion-IDEA. Visually, Diffusion-IDEA significantly outperforms Diffusion-TS in the Context-FID metric, underscoring its enhanced accuracy in approximating the training data distribution. Furthermore, Diffusion-IDEA achieves superior results in terms of Correlational Score, further validating its proficiency in characterizing the underlying correlation structures. Concurrently, the KDE plots reveal that the data generated by Diffusion-IDEA highly coincides with the ground truth within the main peak region. In contrast, the Diffusion-TS curve exhibits a discernible deviation in the transition zones flanking the main peak.

### 3.5. Ablation study

To examine the effectiveness of each component, we report the ablation results in Table 5. Relative to all ablated variants, the

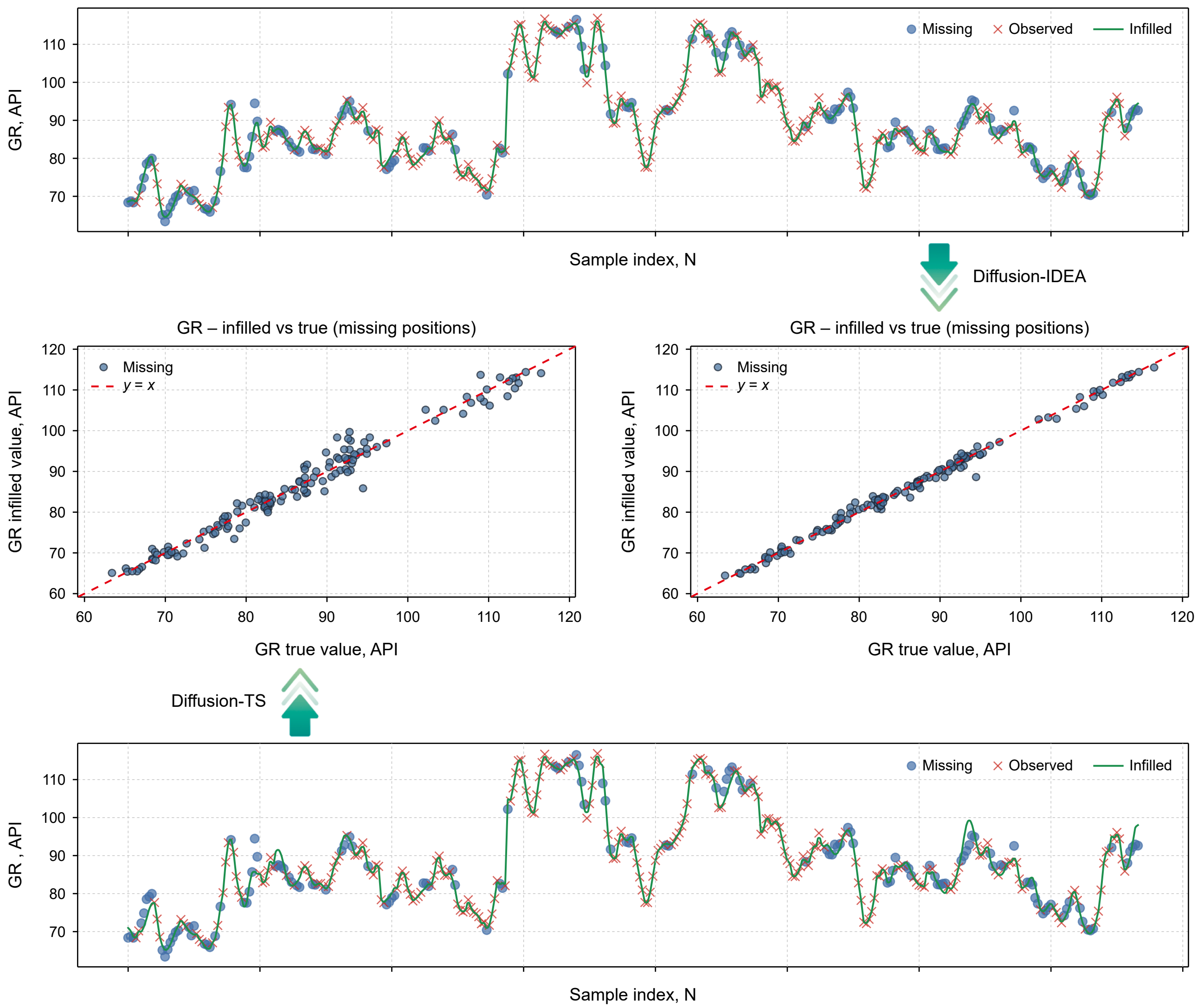


**Fig. 5.** Reconstruction results for missing GR points.

complete Diffusion-IDEA framework achieves the best errors across all logging curves. Removing the Dual Mamba backbone causes a pronounced error increase, indicating that this backbone is pivotal for modeling inter-stratal lithologic dependencies and for capturing intra-layer feature representations. By contrast, removing TCA leads to RMSE increases of varying magnitude across different curves. For example, the RMSE on RXO, CNL, and GR rises by 8.10%, 11.73%, and 7.30%, respectively (when TCA is replaced with standard cross-attention). This suggests that the introduced thresholding mechanism effectively filters out low-correlation noisy matches, preventing the attention module from overfitting to non-physical inter-channel couplings. Removing the DEA module likewise results in a systematic degradation across the full set of logging parameters. To quantify this effect, we compute the overall mean metrics over eight logs. The mean RMSE deteriorates from 0.7228 to 0.9277 (+28.35%), while the mean MAE increases from 0.4767 to 0.5962 (+25.07%). Collectively, these results provide compelling evidence that DEA promotes beneficial convergence toward more informative representations of the underlying geophysical signal. Finally, when the decomposition framework was replaced with a setting that directly predicts the true values, the consistent rebound observed across the reconstruction results of all parameters after the loss of decomposition guidance indicates that the interpretable decomposition module proposed in this study is not merely introduced for transparent training visualization. Rather, it is incorporated into the optimization process as a form of structured physical prior. By explicitly decoupling the low-frequency trend from the high-frequency seasonal components, the model effectively avoids feature entanglement caused by cross-band energy interactions during gradient backpropagation.

To further elucidate the individual contributions of the proposed TCA and DEA modules within the Diffusion-IDEA framework, a qualitative comparative analysis was conducted on typical intervals characterized by complex responses (see Fig. 9). In the absence of the TCA module, the reconstructed curves manifest significant high-frequency artifacts and stochastic fluctuations throughout the sequence. This phenomenon is particularly pronounced in the relatively quiescent segments of the AC and GR curves in Fig. 9(a) and (b). Conversely, configurations lacking the DEA module yield a discernible over-smoothing effect, as highlighted in the magnified views. These observations underscore that the synergistic interplay between the TCA and DEA modules enables the framework to maintain global coherence while simultaneously preserving high fidelity in local detail restoration.

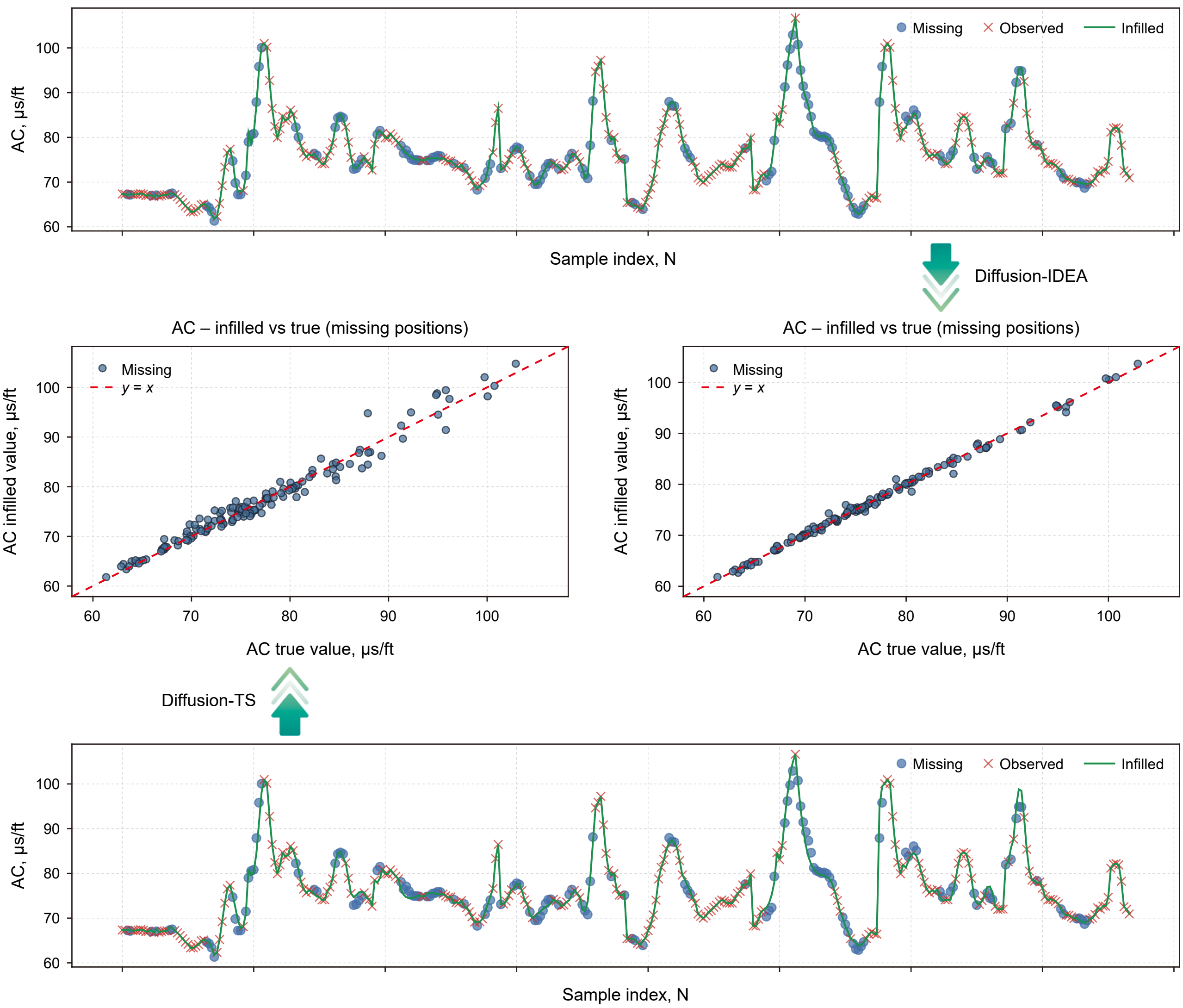


**Fig. 6.** Reconstruction results for missing AC points.

### 3.6. Interpretability validation

The proposed model achieves an effective physics-level disentanglement of well-logging signals by incorporating an explicit trend–season decomposition architecture. As illustrated in Fig. 10, we first apply the diffusion-model forward noising process to the original logging curves (Fig. 10(a)) to construct training samples (Fig. 10(b)). The model then learns to denoise and decomposes the signal into two approximately orthogonal feature subspaces, namely Trend and Seasonal. This decomposition not only preserves reconstruction fidelity but, more importantly, offers an interpretable perspective that is consistent with geological reasoning and cognition.

The Trend component characterizes the low-frequency, macroscopic stratigraphic background (Fig. 10(e)), and the extracted trend curves exhibit excellent smoothness and continuity. From a geophysical perspective, this behavior is consistent with regional-scale compaction trends or long-period depositional evolution. Notably, RT and RI show a high degree of morphological consistency in the extracted Trend component, suggesting that the model successfully recovers an electrical baseline associated with the formation. In contrast, the Seasonal component primarily captures local heterogeneity and responses to lithologic interfaces. It displays pronounced high-frequency oscillations, with energy concentrated in intervals where petrophysical properties undergo abrupt changes. As evidenced by the comparison between Fig. 10(a) and (d), within Index 80–90 the original curves exhibit sharp transitions, and the Seasonal component accurately reconstructs the corresponding high-amplitude impulses. Geological information confirms that this interval corresponds to an interbed transition from argillaceous siltstone to sandy conglomerate. Together with the overall model output (Fig. 10(c)), these observations corroborate the validity and geological plausibility of the proposed interpretable decomposition framework.

In addition, to rigorously quantify the physical significance of the model's interpretable decoupling, we introduced the Mutation Capture Correlation (MCC) and the Spectral Decoupling Index (SDI) for validation. Specifically, MCC measures the Pearson correlation coefficient between the extracted high-frequency component and the first-order difference of the original well-log signal, thereby quantifying the ability of the decomposed component to anchor real physical structural boundaries. The p-value is employed as an auxiliary verification for MCC to assess the probability under which the null hypothesis holds. SDI is defined as the proportion of the overlapping area between the power spectral density curves of the low-frequency and high-frequency components.

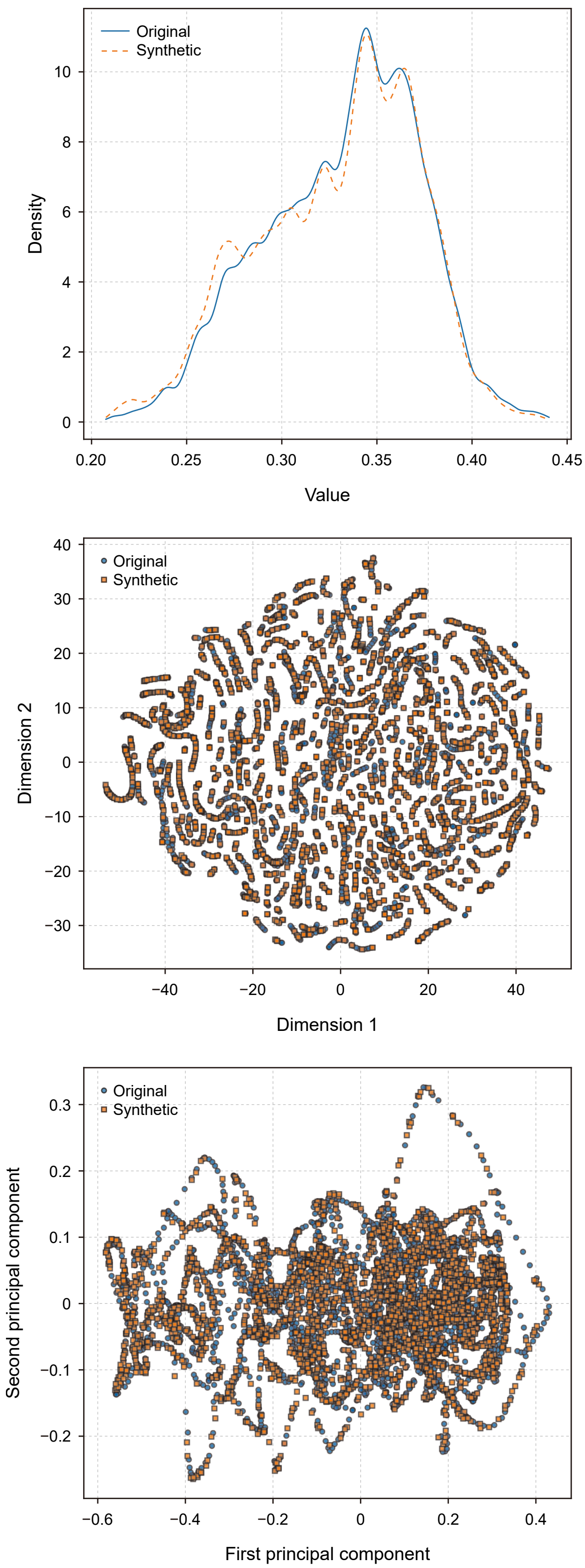


**Fig. 7.** Diffusion IDEA training outcomes (left: KDE; middle: t SNE; right: PCA).

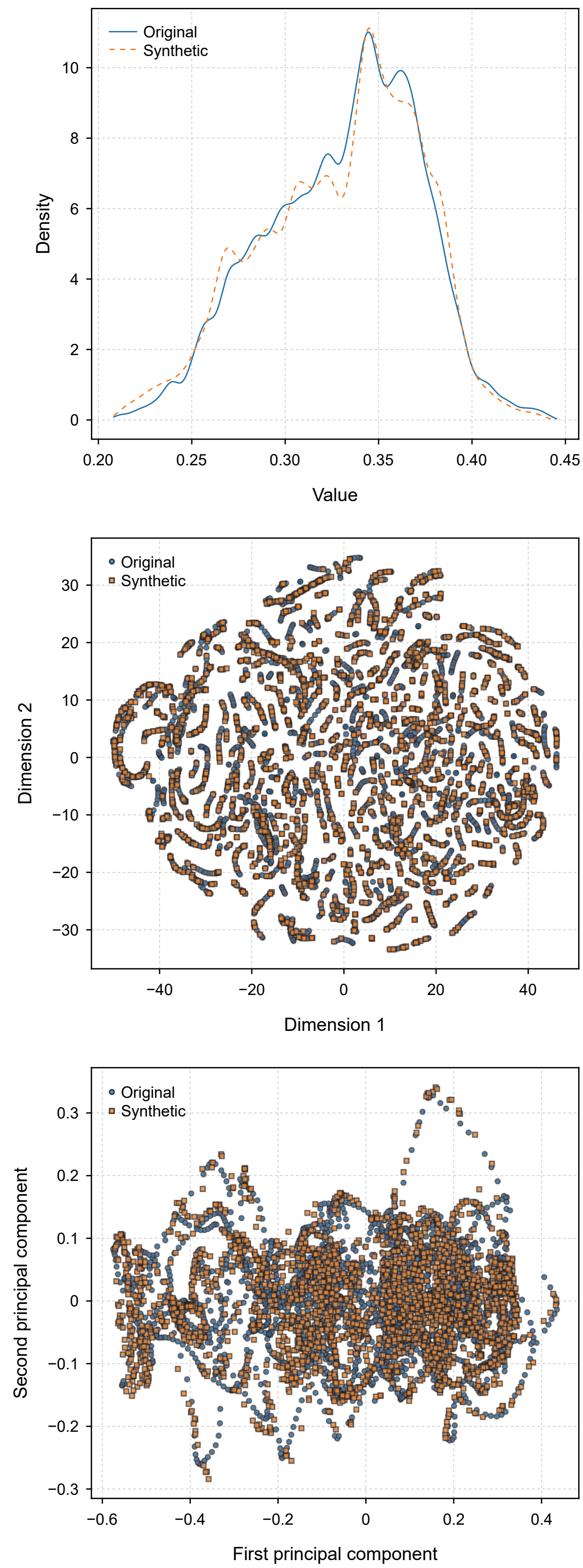


**Fig. 8.** Diffusion TS training outcomes (left: KDE; middle: t SNE; right: PCA).

The experimental results presented in Table 6 show that the EMD method yields an SDI as high as 78.68%, indicating a severe mode-mixing defect, in which high-frequency noise and low-frequency trends are highly entangled in the frequency domain. Combined with the MCC metric, the results further confirm that the high-frequency component decomposed by EMD does not exhibit a statistically significant correlation with the true physical lithologic discontinuities. The SDI of Diffusion-TS is higher than that of our proposed model, suggesting its insufficient sensitivity to thin interbedded geological layers, with the MCC remaining at

**Table 4**
Results for modeling the training distribution.

| Model | Context-FID (Lower the better) | Correlational score (Lower the better) |
|---|---|---|
| Diffusion-IDEA | **0.098±0.012** | **0.100±0.030** |
| Diffusion-TS | 0.175 ± 0.010 | 0.106 ± 0.030 |

The bold values represent the best results.

only 0.2645. In contrast, Diffusion-IDEA improves the MCC to the best value of 0.3692 ($p < 10^{-37}$), demonstrating that our method effectively captures valid lithologic interfaces with geological origins. Meanwhile, the SDI decreases to a reasonable 43.57%. This proportion indicates that, while the model allows benign energy sharing between the trend and seasonal components, it preserves to the greatest extent the physical sharpness and fidelity of the physical discontinuity interfaces.

### 3.7. Hyperparameter sensitivity analysis

To evaluate the stability of the proposed model and to identify the optimal configuration for log reconstruction, we conduct a sensitivity analysis with respect to key hyperparameters. Specifically, we examine how the sliding-window length ($w$) affects the model's ability to simultaneously capture local details and long-range stratigraphic dependencies, with ($w \in \{24, 48, 64\}$). In addition, to further assess the impact of the sample overlap ratio (determined by the sliding stride) on fitting stability and generalization performance, we set the sliding stride to ($s \in \{1, 2, 4\}$). And the experimental results for the validation of the

**Table 5**
Ablation study results.

| Model | Metric | RXO | DEN | CNL | GR | RT | RI | AC | CALI |
|---|---|---|---|---|---|---|---|---|---|
| Diffusion-IDEA | RMSE | **2.4500** | **0.0184** | **0.5422** | **0.9309** | **0.6628** | **0.3100** | **0.7728** | **0.0958** |
| | MAE | **1.5476** | **0.0092** | **0.3759** | **0.6691** | **0.4348** | **0.2198** | **0.5029** | **0.0542** |
| w/o Dual Mamba | RMSE | 8.1816 | 0.0570 | 5.8043 | 9.9853 | 2.3107 | 1.4906 | 4.5893 | 0.3556 |
| | MAE | 5.0640 | 0.0363 | 4.7012 | 7.8821 | 1.8590 | 1.2594 | 2.3177 | 0.2530 |
| w/o TCA | RMSE | 2.6484 | 0.0208 | 0.6058 | 0.9989 | 0.7268 | 0.3620 | 0.8306 | 0.1171 |
| | MAE | 1.6343 | 0.0136 | 0.4396 | 0.7087 | 0.5330 | 0.2771 | 0.5845 | 0.0657 |
| w/o DEA | RMSE | 2.9653 | 0.0383 | 0.6560 | 1.0717 | 0.9561 | 0.6367 | 0.9770 | 0.1201 |
| | MAE | 1.7250 | 0.0242 | 0.4782 | 0.7869 | 0.5969 | 0.4148 | 0.6814 | 0.0626 |
| w/o Decomposition | RMSE | 2.4635 | 0.0192 | 0.5614 | 0.9488 | 0.6774 | 0.3318 | 0.8010 | 0.0970 |
| | MAE | 1.5612 | 0.0096 | 0.3994 | 0.6800 | 0.4457 | 0.2355 | 0.5268 | 0.0560 |

The bold values represent the best results.

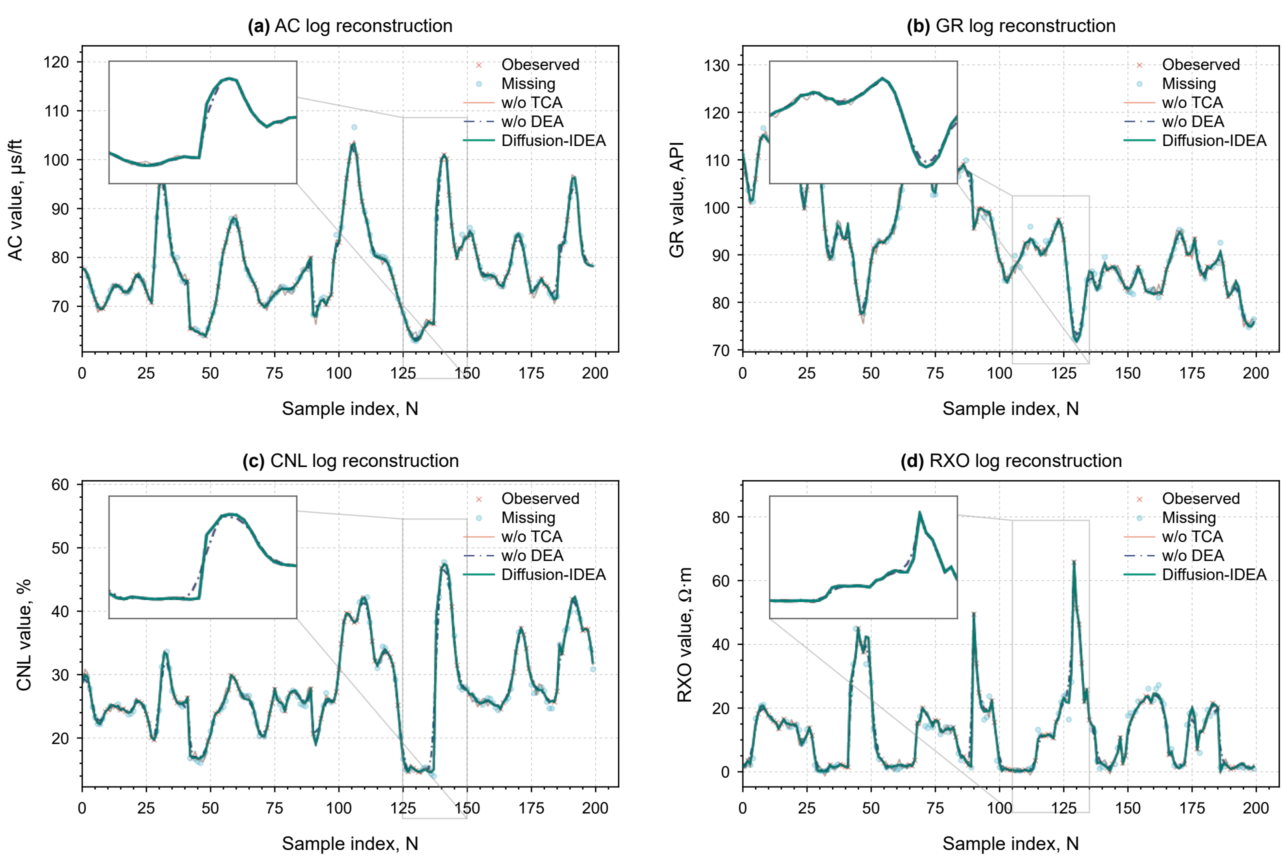


**Fig. 9.** Ablation study comparison (Visual results of Diffusion-IDEA versus the removal of TCA and DEA). For clearer comparison, additional localized magnified views are provided.

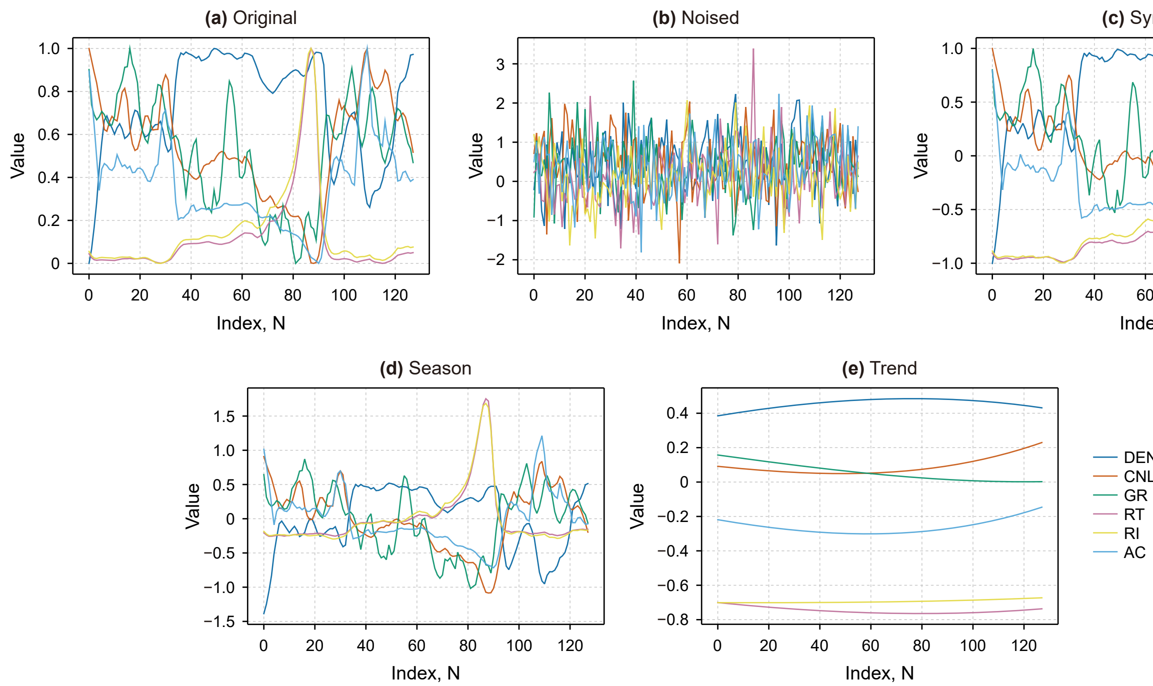


**Fig. 10.** Visualization for interpretability validation: (**a**) original well-logging curves; (**b**) noised samples generated by applying the diffusion forward process to the original data; (**c**) model outputs; (**d**) extracted seasonal component; (**e**) extracted trend component. (For clarity, only representative curves are shown).

**Table 6**
Quantitative evaluation of decomposition quality.

| Model | MCC | p-value | SDI, % |
|---|---|---|---|
| Diffusion-IDEA | **0.3692** | **1.61e-38** | **43.57%** |
| Diffusion-TS | 0.2645 | 3.2e-15 | 58.32% |
| EMD | 0.1247 | 0.5532 | 78.68% |

The bold values represent the best results.

**Table 7**
Results of the sliding-window hyperparameter sensitivity analysis.

| Window | Training time, min | Metric | DEN | CNL | GR | RT |
|---|---|---|---|---|---|---|
| 48 | 65 | RMSE | **0.0184** | **0.5422** | **0.9309** | **0.6628** |
| | | MAE | **0.0092** | **0.3759** | **0.6691** | **0.4348** |
| 24 | 40 | RMSE | 0.0243 | 0.5926 | 1.0433 | 0.8050 |
| | | MAE | 0.0168 | 0.4364 | 0.7480 | 0.5450 |
| 64 | 120 | RMSE | 0.0183 | 0.5402 | 0.9300 | 0.6615 |
| | | MAE | 0.0090 | 0.3730 | 0.6689 | 0.4295 |

The bold values represent the best results.

**Table 8**
Results of the sliding-stride hyperparameter sensitivity analysis.

| Stride | Metric | DEN | CNL | GR | RT |
|---|---|---|---|---|---|
| 1 | RMSE | **0.0184** | **0.5422** | **0.9309** | **0.6628** |
| | MAE | **0.0092** | **0.3759** | **0.6691** | **0.4348** |
| 2 | RMSE | 0.0190 | 0.5662 | 0.9471 | 0.6708 |
| | MAE | 0.0112 | 0.3987 | 0.6722 | 0.4416 |
| 4 | RMSE | 0.0205 | 0.5974 | 1.0169 | 0.6912 |
| | MAE | 0.0131 | 0.4315 | 0.7254 | 0.4749 |

The bold values represent the best results.

**Table 9**
Analysis of the threshold truncation hyperparameter.

| Threshold | Metric | DEN | CNL | GR | RT |
|---|---|---|---|---|---|
| 0.6 | RMSE | 0.0212 | 0.6200 | 1.0010 | 0.7401 |
| | MAE | 0.0155 | 0.4566 | 0.7120 | 0.5427 |
| 0.75 | RMSE | 0.0185 | 0.5498 | 0.9387 | 0.6712 |
| | MAE | 0.0094 | 0.3811 | 0.6700 | 0.4419 |
| 0.8 | RMSE | **0.0184** | **0.5422** | **0.9309** | **0.6628** |
| | MAE | **0.0092** | **0.3759** | **0.6691** | **0.4348** |
| 0.85 | RMSE | 0.0189 | 0.5515 | 0.9418 | 0.6800 |
| | MAE | 0.0097 | 0.3890 | 0.6707 | 0.4524 |
| 1 | RMSE | 0.0208 | 0.6058 | 0.9989 | 0.7268 |
| | MAE | 0.0136 | 0.4396 | 0.7087 | 0.5330 |

The bold values represent the best results.

cumulative probability threshold $\tau \in \{0.6, 0.75, 0.8, 0.85, 1.0\}$ are reported in Tables 7–9.

The results indicate that when the window length is set to ($w$= 24), the model yields consistently higher errors across all logging curves due to a limited effective receptive field, which prevents it from adequately modeling long-range dependencies across stratigraphic intervals. Expanding the window to ($w$= 48) leads to a pronounced performance improvement, demonstrating that sufficient long-range context is critical for robustly extracting interval-scale lithologic background features and their depth-dependent structures. Further increasing the window to ($w$= 64) provides a larger receptive field; however, the additional accuracy gain over ($w$= 48) exhibits clear diminishing returns. Meanwhile, the training cost rises substantially: the training time increases from approximately 65 min to about 120 min. Balancing accuracy and efficiency, $w$= 48 offers a more favorable trade-off under our experimental setting. The stride sensitivity results are summarized in Table 8. Overall, as the sliding stride ($s$) increases from 1 to 4, the errors on all logging curves increase monotonically, indicating that a higher sample-overlap ratio (i.e., a smaller stride) yields more comprehensive training coverage, thereby improving fitting stability and reconstruction accuracy. Finally, as observed from the results in Table 9, the model performance exhibits a U-shaped trend as the threshold varies. When the threshold is set to 1, the model degenerates into the conventional full-attention mechanism, which corroborates our theoretical analysis of the proposed TCA.

Meanwhile, when the threshold is set too low, the model error increases. This phenomenon arises because excessive truncation causes the model to lose critical contextual information from homologous stratigraphic intervals. Based on these quantitative analyses, we consistently adopt $w = 48$, $s = 1$ and $\tau = 0.8$ in all experiments.

### 3.8. Generalization experiments

To evaluate the model's generalization, we jointly train on MH133 together with four neighboring offset wells. The Triassic section of MH13, comprising approximately 2350 samples, serves as the blind-well test set. MH13 is excluded from training; at inference, 40% of its observed data are provided as conditioning information to sample and reconstruct the missing points.

Fig. 11 presents the imputation results for eight key petrophysical logging curves from the MH13 blind well, shown alongside the stratigraphic lithology column for comparison. The continuous missing intervals highlighted by the red dashed boxes are further displayed in enlarged views. As can be observed, our reconstructions exhibit excellent agreement with the ground-truth formation logs. In addition, Table 10 reports the quantitative comparison between Diffusion-IDEA and the baseline (Diffusion-TS) on the blind well. For curves indicative

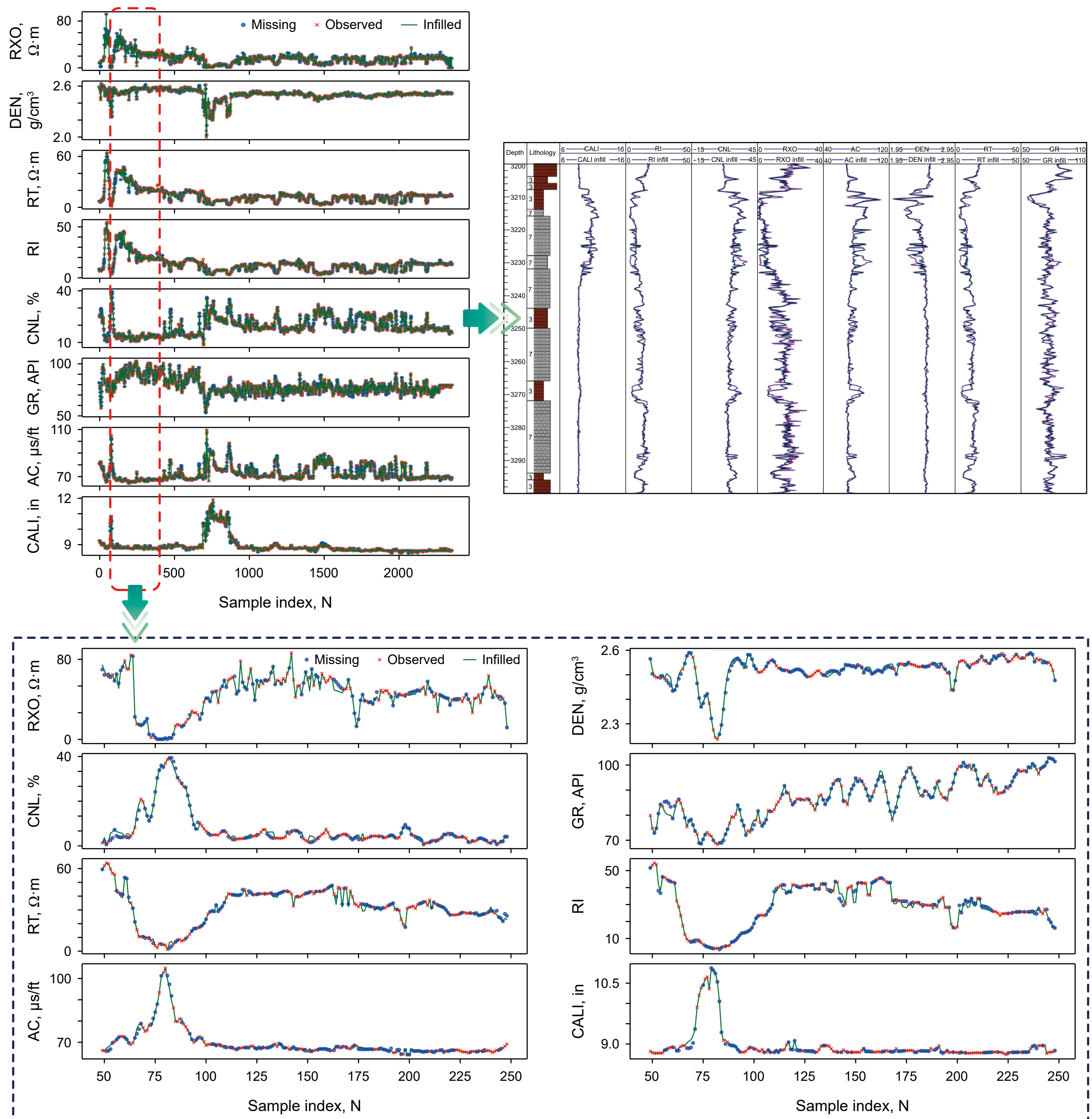


**Fig. 11.** Visualization of blind-well evaluation. (The upper-left panel shows the imputed logging curves with missing and observed samples indicated; the upper-right panel presents the log-interpretation results; the lower panels provide zoomed-in views of the intervals highlighted by the red boxes.)

**Table 10**
Quantitative results for the blind well test (reported in the original physical units).

| Model | Metric | RXO | DEN | CNL | GR | RT | RI | AC | CALI |
|---|---|---|---|---|---|---|---|---|---|
| Diffusion-IDEA | RMSE | **4.3031** | **0.0179** | **1.2508** | **2.9994** | **2.2856** | **1.9711** | **1.5719** | **0.0857** |
| | MAE | **2.6398** | **0.0133** | **0.8931** | **2.2014** | **1.2967** | **1.1952** | **1.0518** | **0.0589** |
| Diffusion-TS | RMSE | 6.1951 | 0.0764 | 2.4802 | 3.6092 | 2.7590 | 3.2034 | 3.4132 | 0.2368 |
| | MAE | 3.7434 | 0.0427 | 1.9757 | 2.5711 | 1.6720 | 2.3136 | 2.3451 | 0.1541 |

The bold values represent the best results.

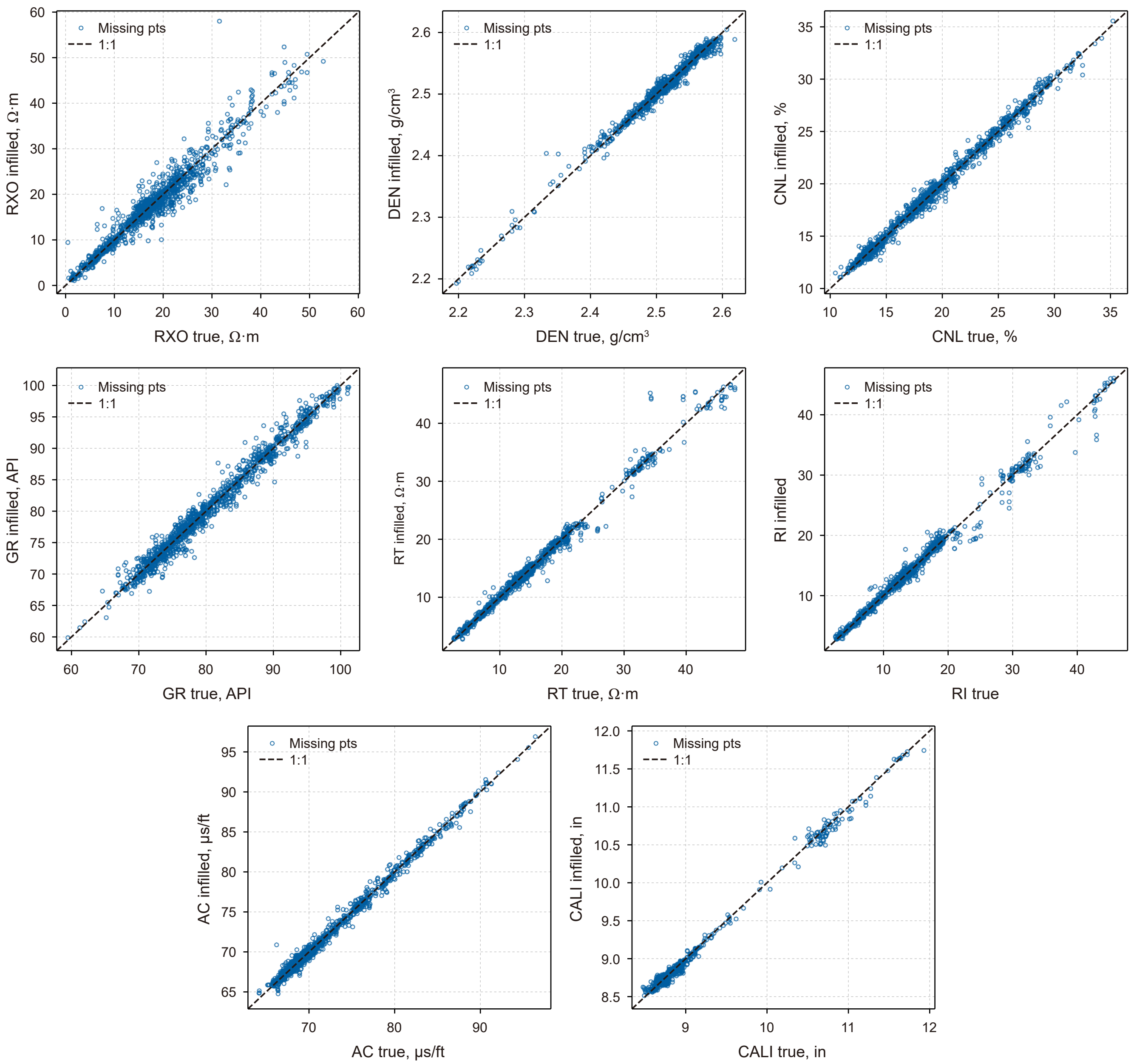

**Fig. 12.** MH13 blind-well test: scatter plots of infilled versus true values at missing points for each log curve (Diffusion-IDEA).

of porosity and lithology, the proposed model achieves high physical fidelity; for example, for DEN, the RMSE reaches 0.0179. For resistivity logs characterizing fluid properties (RT, RXO, and RI), the model effectively handles the large dynamic range. Compared with the MAE of 1.6720 achieved by Diffusion-TS on deep laterolog resistivity (RT), our model reduces it to 1.2967. Furthermore, Fig. 12 provides scatter plots for Diffusion-IDEA, offering additional evidence of the reconstruction accuracy.

In the above blind-well tests, although our model still maintained excellent overall imputation accuracy, some curves (e.g., GR, whose RMSE increased from 0.9309 in the main experiment to 2.9994) still exhibited substantial performance fluctuations. Therefore, we introduced multidimensional statistical metrics to

**Table 11**
Quantitative results of the data shift between the training set and the blind well.

| | |
|---|---|
| Wasserstein distance | 9.9637 |
| Jensen–Shannon divergence (JSD) | 0.1368 |
| Training-set volatility (Std) | 12.2484 |
| Blind-well volatility (Std) | 7.5555 |
| Volatility shift ratio | −38.31% |

quantitatively analyze the data distribution shift between the training set and the blind well.

Using the GR curve, which exhibited the most pronounced fluctuation, as an example, the quantitative results (Table 11) reveal intense lateral facies variation within the study area. First, the Wasserstein distance between the training set and the blind-well data approaches 10 API, clearly indicating a substantial depositional baseline shift in the region where the blind well is located. Second, the Jensen–Shannon (JS) divergence further demonstrates a significant probabilistic mismatch in the volumetric proportions of lithofacies components between the two datasets. Meanwhile, statistical analysis shows that the fluctuation level of the blind well (Std = 7.55) decreases markedly by 38.31% relative to that of the training set (Std = 12.24). This can be attributed to the fact that the model mainly learned prior features of frequently alternating high-frequency thin interbeds during training, whereas the blind well encountered more massive strata. Notably, even under an input-end baseline drift as high as 9.96 API, the proposed model still controls the final absolute reconstruction error to 2.9994 API. This result strongly evidences the robustness of the model under extreme geological variations: rather than mechanically propagating the bias, the model adaptively absorbs the macroscopic shift through mechanisms such as adaptive filtering and interpretable decomposition, thereby avoiding catastrophic degradation in reconstruction performance.

In addition, we further conducted a few-shot fine-tuning experiment to verify whether enhanced data diversity could mitigate the transfer error in the blind well. Specifically, only 10% of the logging intervals from the blind well were randomly sampled and incorporated into the original training set as a supplement to spatial phase-transition diversity, after which the model was lightly fine-tuned. The remaining 90% of the completely unseen intervals in the blind well were then used for re-evaluation. As shown in Table 12, after introducing a small amount of diversity data from the target work area, the model's adaptability to transferred stratigraphic conditions was effectively activated. The RMSE of the GR curve decreased from 2.9994 to 2.5821, while the other seven logging curves also exhibited consistent error convergence. Therefore, we conclude that the performance fluctuations of the proposed model in the purely blind-well setting do not stem from limitations in the representational capacity of the model architecture; rather, they reflect the objective geological challenge posed by severe spatial distribution shifts in subsurface strata.

**Table 13**
Comparative analysis of computational complexity and efficiency.

| Window | Computational efficiency | Diffusion-IDEA | Diffusion-TS | Transformer |
|---|---|---|---|---|
| 24 | Training time, min | 40 | 58 | 45 |
| | Inference time, min | 1.5 | 1.8 | 1.5 |
| 48 | Training time, min | 65 | 100 | 84 |
| | Inference time, min | 5 | 8 | 7.5 |
| 64 | Training time, min | 120 | 210 | 175 |
| | Inference time, min | 12 | 20 | 18 |

### 3.9. Computational cost and efficiency analysis

While high-fidelity well-log reconstruction remains the paramount objective, computational efficiency in practical geophysical interpretation is equally critical for field deployment. As previously discussed, our proposed Diffusion-IDEA incorporates a selective state space architecture coupled with a sparsification strategy to mitigate computational complexity. To rigorously quantify this trade-off, we conducted a comprehensive evaluation of the computational overhead for Diffusion-IDEA relative to representative baseline models, specifically Diffusion-TS and the Transformer.

The experimental results are tabulated in Table 13, from which several salient observations can be drawn. First, Diffusion-IDEA consistently outperforms representative baselines (Diffusion-TS and the Transformer) in both training duration and inference efficiency across all window scales. Second, these findings empirically validate the theoretical complexity advantages of our proposed architecture. As the window size expands, the training times for the Transformer and Diffusion-TS undergo a sharp nonlinear escalation due to the quadratic complexity inherent in the attention mechanism (e.g., the training time for Diffusion-TS surges from 58 to 210 min). In contrast, Diffusion-IDEA exhibits near-linear and graceful scaling. Furthermore, during the inference phase, despite the iterative denoising nature of diffusion models, Diffusion-IDEA requires only 5 min at a window size of 48, notably surpassing the non-diffusion Transformer. Collectively, these results indicate that Diffusion-IDEA strikes an optimal trade-off between high-fidelity reconstruction quality and the computational demands of practical industrial deployment.

## 4. Conclusion

In this paper, we propose Diffusion-IDEA, a novel framework for well log curve reconstruction, aimed at rectifying the current behavioral bias in the field whereby the reconstruction task is treated as a prediction task. We inject observed information into an iterative sampling process to progressively recover missing well log points under the current conditioning constraints. Specifically, we introduce three principal enhancements to the training pipeline: (i) integrating Dual Mamba to meet long-window training requirements; (ii) adopting a threshold-limited cross-attention mechanism (TCA) so that the attention fusion is more consistent

**Table 12**
Few-shot fine-tuning experiment.

| Model | Metric | RXO | DEN | CNL | GR | RT | RI | AC | CALI |
|---|---|---|---|---|---|---|---|---|---|
| Diffusion-IDEA | RMSE | 4.3031 | 0.0179 | 1.2508 | 2.9994 | 2.2856 | 1.9711 | 1.5719 | 0.0857 |
| | MAE | 2.6398 | 0.0133 | 0.8931 | 2.2014 | 1.2967 | 1.1952 | 1.0518 | 0.0589 |
| Few-shot | RMSE | 4.0120 | 0.0157 | 1.2236 | 2.5821 | 2.1974 | 1.8682 | 1.5424 | 0.0814 |
| | MAE | 2.2984 | 0.0110 | 0.8609 | 2.1034 | 1.1844 | 1.1148 | 1.0311 | 0.0552 |

with the diffusion model's noise schedule (high noise in early steps, low noise in later steps); and (iii) combining adaptive filtering with an interpretable decomposition framework (DEA) to achieve effective denoising and interpretable signal decomposition, thereby enhancing the robustness and interpretability of reconstruction.

Relative to conventional local reconstructions that operate on a single curve at a time, this study introduces a new paradigm for the joint reconstruction of the full suite of conventional well logs: the model simultaneously models multiple log responses-natural gamma ray, density, neutron, acoustic, resistivity, and caliper-within a unified latent space. This design substantially reduces the risk of petrophysical mismatch induced by per-curve independent completion. Comprehensive experiments show that conditionally constrained diffusion reconstruction methods consistently outperform prediction-oriented baselines (Diffusion-IDEA reduces average RMSE by 45.8% and average MAE by 52.4% relative to Dual Mamba; Diffusion-TS reduces average RMSE and average MAE by 38.3% and 44.2% relative to a Transformer baseline). These results empirically validate our earlier proposition that curve reconstruction is better formulated as a conditional reconstruction problem rather than a conventional prediction problem. Nevertheless, limitations remain: at present, we cannot perform fully blind infilling of an individual log solely from the remaining observed logs, a constraint tied to our training mechanism. In future work, we will address this by refining the fusion strategy for conditioning information and by exploring a mask-based diffusion training paradigm, with the goal of further improving model performance. However, while the current framework performs excellently in localized infilling, in scenarios of extreme data loss (i.e., total blind filling), the reconstruction essentially degenerates into a highly irregular inverse problem due to the complete absence of local constraints. Future research will be dedicated to advancing this framework to achieve global reconstruction and extrapolation of well-log sequences.

## 5. Discussion

In the above work, we adopted a random masking strategy to simulate missing well-logging intervals and used the diffusion process to learn the data distribution of observable intervals. This training mechanism improved the model's reconstruction accuracy for local missing regions to a certain extent, but it still has clear limitations objectively. First, this strategy emphasizes imputation by using local context, and it still cannot perform fully blind completion for single or multiple well-logging curves. In practical oilfield engineering deployment, this means that when extreme operating conditions arise, such as complete failure of logging instruments or severe obstruction, causing one curve (such as sonic or density) to be missing over the entire well section, the current model cannot directly synthesize a complete substitute curve from scratch. Therefore, the practical application scenarios of the present model remain mainly limited to local assimilation tasks, including well-log data quality control, repair of local bad-hole intervals or logging skipping sections, and multi-interval splicing. Second, the model still has limited ability to capture global topological information that involves multi-level stratigraphic transitions and extreme geological discontinuity zones.

To address the above limitations, in future research we will further adjust the masking-based training strategy and explore multi-granularity structured masking schemes. Specifically, during training we will extend the structured parallel masking mechanism across different scales and granularities, for example, segment-level masks based on key structural blocks and global masks over complete single-curve sequences. In this way, we aim to strengthen the model's learning and reconstruction capabilities for stratigraphic interfaces and complex lithologic zones, and to realize multiple reconstruction schemes, including fully blind completion of a single well-logging curve and joint completion of partially missing multivariable logging regions within a well. At the same time, we will further attempt to introduce stratigraphic boundary labels as soft guidance and encode geologically critical points interpreted by experts, such as depositional unconformities and lithofacies discontinuity interfaces, as functional placeholder features. During the reverse sampling evolution of the diffusion model, these labels carrying physical meaning will act as local deterministic anchors and impose stratigraphic trend constraints on the generation trajectory in latent space. We expect this strategy to effectively alleviate the over-smoothing effect that readily occurs in long-interval blind completion tasks, without compromising the model's generalization ability, and to improve the model's structural interpretation stability when it faces large-scale continuous missing regions.

## CRediT authorship contribution statement

**Cheng Feng:** Validation, Supervision, Resources, Project administration, Funding acquisition. **Guo-Zeng Zhang:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Zi-Yan Feng:** Supervision, Investigation, Formal analysis, Data curation. **Jie Yu:** Investigation, Funding acquisition, Conceptualization. **Meng-Ru Hu:** Visualization, Validation. **Ke-Gang Ling:** Supervision, Resources.

## Data availability

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

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We gratefully acknowledge the financial support from the National Natural Science Foundation of China (No. 42364007), the Key Research and Development Project (Nos. 2024B01016, 2024B01016-1, and 2024B01016-3), the Tianshan Talent Program (No. 2024TSYCCX0069) of Xinjiang Uygur Autonomous Region.

## References

Al-Fakih, A., Koeshidayatullah, A., Mukerji, T., et al., 2025. Well log data generation and imputation using sequence based generative adversarial networks. Sci. Rep. 15 (1), 11000. https://doi.org/10.1038/s41598-025-95709-0.

Chen, X., Lv, F., Zhang, J., et al., 2025. Artificial seismic waves generation for complex matching conditions based on diffusion model. Soil Dynam. Earthq. Eng. 192, 109290. https://doi.org/10.1016/j.soildyn.2025.109290.

Chen, Z., Zhang, Y., Li, J., et al., 2025. Large AI models for well log curve reconstruction. Petrol. Explor. Dev. 52 (3), 744–756 (in Chinese).

Cheng, C., Gao, Y., Chen, Y., et al., 2022. Reconstruction method of old well logging curves based on BI-LSTM model-taking Feixianguan formation in East Sichuan as an example. Coatings 12 (2), 113. https://doi.org/10.3390/coatings12020113.

Dong, S.Q., Sun, Y.M., Xu, T., et al., 2023a. How to improve machine learning models for lithofacies identification by practical and novel ensemble strategy and principles. Pet. Sci. 20 (2), 733–752. https://doi.org/10.1016/j.petsci.2022.09.006.

Dong, S.Q., Zhong, Z.H., Cui, X.H., et al., 2023b. A deep kernel method for lithofacies identification using conventional well logs. Pet. Sci. 20 (3), 1411–1428. https://doi.org/10.1016/j.petsci.2022.11.027.

Fan, P., Deng, R., Qiu, J., et al., 2021. Well logging curve reconstruction based on kernel ridge regression. Arabian J. Geosci. 14 (16), 1559. https://doi.org/10.1007/s12517-021-07792-y.
Han, M., Lin, B., Zhang, R., et al., 2024. A missing well-logs imputation method based on conditional denoising diffusion probabilistic models. SPE J. 29 (5), 2165–2180. https://doi.org/10.2118/219452-PA.
Haritha, D., Satyavani, N., Ramesh, A., 2025. Generation of missing well log data with deep learning: CNN-Bi-LSTM approach. J. Appl. Geophys. 233, 105628. https://doi.org/10.1016/j.jappgeo.2025.105628.
Huang, J., He, T., Zhu, W., et al., 2025. State of health estimation method for lithium-ion batteries based on multi-feature fusion and Swin Transformer model. Ionics 1–24. https://doi.org/10.1007/s11581-025-06657-8.
Lee, S., Hong, J., Liu, L., et al., 2024. Ts-fastformer: fast transformer for time-series forecasting. ACM Trans. Intell. Syst. Technol. 15 (2), 1–20. https://doi.org/10.1145/3630637.
Li, F., Liu, H., Yang, X., et al., 2023. Reconstruction of acoustic well log curves based on U-Net neural network. J. Ocean Univ. China 53 (8), 86–92. https://doi.org/10.16441/j.cnki.hdxb.20220239, 103, (in Chinese).
Li, H., Chen, M., Zhang, X., et al., 2025. Density well log curve reconstruction method based on Bayesian-optimized CNN-LSTM. Well Logging Technol. 49 (2), 198–208. https://doi.org/10.16489/j.issn.1004-1338.2025.02.007 (in Chinese).
Li, Z., Jiang, S., 2025. Reconstructing and interpreting analysis of sonic log curves based on machine learning and SHAP algorithm. Geol. Sci. Technol. Bull. 44 (1), 321–331. https://doi.org/10.19509/j.cnki.dzkq.tb20230504 (in Chinese).
Liu, M., Han, J., Cao, Z., et al., 2024a. A study on well log curve reconstruction based on modal decomposition and LSTM attention representation. Acta Electron. Sin. 52 (4), 1399–1410 (in Chinese).
Liu, S., Wang, Y., Ding, F., et al., 2024b. Joint iterative state and parameter estimation for bilinear systems with autoregressive noises via the data filtering. ISA Trans. 147, 337–349. https://doi.org/10.1016/j.isatra.2024.01.035.
Mei, Y., Chen, M., Shen, Y., et al., 2025. Well log curve reconstruction method based on bidirectional LSTM and multi-head attention mechanism. Oil Drill. Production Technol. 47 (3), 277–288. https://doi.org/10.13639/j.odpt.202503025, 328. (in Chinese).
Patro, B.N., Agneeswaran, V.S., 2025. Mamba-360: survey of state space models as transformer alternative for long sequence modelling: methods, applications, and challenges. Eng. Appl. Artif. Intell. 159, 111279. https://doi.org/10.1016/j.engappai.2025.111279.
Qin, J., Pei, D., Guo, Q., et al., 2024. Intersection-union dual-stream cross-attention Lova-SwinUnet for skin cancer hair segmentation and image repair. Comput. Biol. Med. 180, 108931. https://doi.org/10.1016/j.compbiomed.2024.108931.
Ren, Q., Zhang, H., Azevedo, L., et al., 2023. Reconstruction of missing well-logs using facies-informed discrete wavelet transform and time series regression. SPE J. 28 (6), 2946–2963. https://doi.org/10.2118/217425-PA.
Shan, L., Liu, Y., Tang, M., et al., 2021. CNN-BiLSTM hybrid neural networks with attention mechanism for well log prediction. J. Petrol. Sci. Eng. 205, 108838. https://doi.org/10.1016/j.petrol.2021.108838.
Shang, F., Lu, Y., Cao, M., 2022. Well log curve reconstruction method based on an improved LSTM neural network. Comput. Technol. Dev. 32 (6), 198–202 (in Chinese).
Su, L., Zuo, X., Li, R., et al., 2025. A systematic review for transformer-based long-term series forecasting. Artif. Intell. Rev. 58 (3), 80. https://doi.org/10.1007/s10462-024-11044-2.
Tong, G., Ge, Z., Peng, D., 2024. RSMformer: an efficient multiscale transformer-based framework for long sequence time-series forecasting. Appl. Intell. 54 (2), 1275–1296. https://doi.org/10.1007/s10489-023-05250-8.
Vaswani, A., Shazeer, N., Parmar, N., et al., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.
Wang, H., Zhao, B., Liu, J., et al., 2024. Density well log curve reconstruction method based on multi-head attention mechanism TCN-BiGRU. Geophys. Prospect. 1–15 (in Chinese).
Wang, Z., Kong, F., Feng, S., et al., 2025. Is mamba effective for time series forecasting? Neurocomputing 619, 129178. https://doi.org/10.1016/j.neucom.2024.129178.
Wu, Y., Meng, X., Hu, H., et al., 2025. Affirm: interactive mamba with adaptive fourier filters for long-term time series forecasting. Proc. AAAI Conf. Artif. Intell. 39 (20), 21599–21607. https://doi.org/10.1609/aaai.v39i20.35463.
Yuan, X., Qiao, Y., 2024. Diffusion-TS: interpretable diffusion for general time series generation. arXiv preprint arXiv:2403.01742. https://doi.org/10.48550/arXiv.2403.01742.
Zhang, G., Chen, T., Chen, J., et al., 2025. Geophysical logging curve reconstruction based on diff-mamba modeling. Math. Geosci. 1–23. https://doi.org/10.1007/s11004-025-10224-2.
Zhang, H., Wu, W., Song, X., 2024a. Well logs reconstruction based on deep learning technology. IEEE Geosci. Remote Sens. Lett. 21, 1–5. https://doi.org/10.1109/LGRS.2024.3354210.
Zhang, H., Yang, X., Chen, Z., et al., 2022. Well log data reconstruction based on enhanced bidirectional long short-term memory neural networks. Prog. Geophys. 37 (3), 1214–1222. https://doi.org/10.6038/pg2022FF0194 (in Chinese).
Zhang, L., Dang, H., Liu, Q., et al., 2024b. Well log curve reconstruction using LSTM recurrent neural networks with geological stratification constraints. Sci. Technol. Eng. 24 (19), 8045–8051 (in Chinese).
Zhou, W., Zhao, H., Li, X., et al., 2025. Missing well logs reconstruction based on cascaded bidirectional long short-term memory network. Expert Syst. Appl. 259, 125270. https://doi.org/10.1016/j.eswa.2024.125270.