

Deep learning driven horizon tracking with confidence

Wenhao Zheng^{a,b}, Rebecca Bell^{b,*}, Lluís Guasch^b, Carlos Cueto^b, Polina Pasalska^b,
Yanhua Wang^{a,b}

^a Resource Geophysics Academy, Imperial College London, London SW7 2AZ, UK

^b Department of Earth Science and Engineering, Imperial College London, London, SW7 2AZ, UK

ARTICLE INFO

Dataset link: <https://dgbeearthsciences.sharefile.com/share/view/sfa1c428d20e84982928fb1a763ab76e>, https://s3.amazonaws.com/open.source.geoscience/open_data/newzealand/Taranaki_Basin/OPUNAKE-3D/OPUNAKE3D-PR34-61-FS-3D.Final_Stack.sgy, https://s3.amazonaws.com/open.source.geoscience/open_data/newzealand/Taranaki_Basin/Keri_3D/Kerry3D.segy

Keywords:

Seismic horizon tracking
Relative geologic time
Prompt-guided interpretation
Tracking confidence

ABSTRACT

Seismic horizon tracking is fundamental to subsurface interpretation, yet it remains challenging in geologically complex settings characterized by faults, local structural uplift, and weak or discontinuous reflectors. Conventional auto-tracking methods rely heavily on seismic amplitudes and are therefore prone to drift, instability, and poor global consistency when reflector continuity is disrupted. We propose a confidence-aware framework jointly constrained by relative geologic time (RGT) and geologically informed prompt points for 2D and 3D seismic horizon interpretation. RGT is predicted from seismic data using an adversarially regularized network designed to retain structurally meaningful local variations and provide consistent stratigraphic guidance. Horizon extraction is then formulated as a prompt-driven local spreading process in which user-defined points supply geological control and overlapping RGT predictions guide stable propagation. Multiple prompts, including well ties and interpreter-defined control points, provide additional anchors for correcting long-range drift and improving within-domain tracking consistency. We further introduce a conditional tracking-confidence assessment by perturbing local prompt positions and measuring the dispersion among the resulting RGT-guided tracking solutions. Low dispersion indicates stable propagation, whereas high dispersion identifies regions that are sensitive to prompt initialization and may require additional geological control. Experiments using synthetic data and field data from the F3 and Opunake surveys demonstrate coherent tracking across structurally deformed, weak-reflection, and fault-dominated regions. The proposed framework integrates automated propagation, sparse expert constraints, and interpretable stability diagnostics into a unified semi-automatic workflow that identifies stable horizon segments and prioritizes prompt-sensitive regions for refinement.

1. Introduction

Seismic horizon interpretation is a fundamental task in seismic data analysis (Yuan et al., 1999; Posamentier et al., 2022; Bond et al., 2012). Horizons define the structural configuration and stratigraphic framework of the subsurface and provide an essential basis for geological modeling, reservoir characterization, and structural and stratigraphic analysis (Waters et al., 2016; Ruppel and Waite, 2020). In hydrocarbon exploration and broader applied geophysical studies, accurate horizon delineation is important for identifying depositional architectures, correlating stratigraphic units, and constraining the spatial distribution of subsurface strata (Jin et al., 2017; Bond, 2015). However, reliable horizon interpretation remains challenging in real seismic datasets (Yan and Wu, 2021; Tschannen et al., 2020). Fault displacement, weak or discontinuous reflections, degraded data quality, and strong stratigraphic deformation can disrupt the lateral continuity of seismic events and obscure the target interface (Hale, 2013; Wrona et al., 2021). Under

these conditions, automatic tracking may jump to an adjacent reflector, terminate prematurely, or produce multiple plausible but inconsistent interpretations (D'Angelo et al., 2024; Wrona et al., 2018; Peters et al., 2019). More robust and geologically consistent strategies are therefore required for horizon interpretation in complex geological settings (Shi et al., 2020; Lin et al., 2024).

Existing automatic horizon-tracking algorithms can be broadly grouped into local propagation methods, global optimization methods, learning-based methods, and RGT-based stratigraphic modeling. Seed-based local trackers propagate an interpreted event by measuring amplitude or waveform similarity, reflector continuity, or other image-based correlation criteria (Marfurt and Alves, 2015). Dynamic time warping extends trace-to-trace matching by allowing nonlinear alignment between seismic events (Jin et al., 2017). These methods are computationally efficient and allow interpreters to control the initialization, but their dependence on local reflection evidence

* Corresponding author.

E-mail address: rebecca.bell@imperial.ac.uk (R. Bell).

<https://doi.org/10.1016/j.acags.2026.100400>

Received 17 June 2026; Received in revised form 17 August 2026; Accepted 20 August 2026

Available online 9 September 2026

2590-1974/© 2026 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

makes them prone to reflector jumping, premature termination, and cumulative drift where the target event is weak, discontinuous, or displaced by faults. Global optimization methods instead formulate horizon extraction as a path- or surface-optimization problem. Representative approaches include least-squares horizons constrained by local slopes and multigrid correlations (Wu and Fomel, 2018) and dynamic-programming-based horizon extraction (Yan and Wu, 2021). Although these formulations improve lateral consistency, their performance still depends on reliable slope estimates, correlations, or cost functions, which may be difficult to obtain in structurally complex and low-signal-to-noise regions.

Learning-based methods reduce the dependence on manually designed similarity measures by learning seismic representations or predicting horizons directly. Multiresolution neural networks have been used to track horizons from limited training images (Peters et al., 2019), while waveform embedding enables automatic horizon picking through unsupervised learning of reflector similarity (Shi et al., 2020). Three-dimensional neural networks have also been developed to extract complete horizon surfaces (Tschannen et al., 2020), and transformer-based architectures have been introduced to capture broader spatial context (Zhao and Zhao, 2025). Recent studies have further incorporated uncertainty encoding and vertical constraints (Liao et al., 2024), constrained seed-generation strategies (Pu et al., 2024), and well-log information for 3D horizon tracking (D'Angelo et al., 2024). These developments improve automation and demonstrate the value of incorporating additional geological information. Nevertheless, they generally address individual stages of the interpretation workflow, and direct horizon predictions may not explicitly preserve global stratigraphic ordering or provide interpreters with sufficient flexibility to refine the results using additional sparse constraints.

Relative geologic time (RGT) provides an alternative representation that explicitly describes the relative temporal ordering of strata. By representing horizons as isolines or isosurfaces of a continuous time field, RGT constrains horizon interpretation through stratigraphic consistency rather than local amplitude continuity. Classical approaches construct RGT or horizon volumes using graph-cut phase unwrapping and interpreted constraints (Wu and Zhong, 2012; Wu and Hale, 2015). Deep-learning-based RGT estimation has subsequently been investigated in both two and three dimensions. Geng et al. (2020) formulated 2D RGT estimation as a supervised regression problem and used an encoder–decoder network with a refinement module to predict RGT directly from seismic sections. Bi et al. (2021) extended this concept to volume-to-volume RGT prediction with simultaneous horizon and fault interpretation. These studies demonstrate the potential of RGT for improving long-range stratigraphic consistency beyond that achievable through local seismic similarity alone. However, although some RGT-based methods incorporate interpreted horizons or other predefined constraints, the integration of geological information is generally limited in scope and often fixed during RGT construction. Consequently, interpreters have limited flexibility to introduce additional sparse constraints, correct local tracking deviations, or progressively adjust the interpretation after the initial horizon has been extracted. A more flexible framework that allows multiple sparse prompts to directly guide and iteratively refine RGT-based horizon propagation therefore remains needed.

In contrast to previous RGT-focused studies that primarily emphasize RGT reconstruction and the direct extraction of RGT isolines or isosurfaces, this study integrates learned RGT prediction with controllable and progressively adjustable horizon propagation. The main contributions are fourfold: (1) an adversarially regularized RGT predictor is developed to reduce oversmoothing and improve the preservation of fault-related discontinuities and structurally important local variations; (2) a prompt-guided local spreading strategy converts horizon extraction into a controlled propagation process and limits cumulative long-range tracking drift; (3) multiple spatially distributed prompts act as geological anchors to improve stratigraphic consistency and

enable targeted correction of locally inconsistent interpretations; and (4) a perturbation-based procedure quantitatively measures the sensitivity of tracked horizons to prompt placement, providing a conditional assessment of propagation stability. The framework is evaluated in both two and three dimensions using pixel-wise and geology-oriented RGT metrics, downstream horizon-tracking errors, computational-cost analysis, gradient-loss ablation, and a controlled comparison with the published deep-RGT method of Geng et al. (2020).

2. RGT prediction for guided horizon tracking

In our proposed spread tracking and constraint workflow, the deep learning model serves as the backbone engine for local feature extraction. Its primary role is to translate seismic data into RGT data. The effectiveness of the proposed spread tracking method relies on the local sharpness and continuity of RGT contours, particularly across faults and stratigraphic discontinuities. However, optimization strategies dominated by pixel-wise objective functions (e.g., ℓ_1 or ℓ_2 norms) are known to favor averaged solutions, often resulting in overly smooth predictions in regions of high structural uncertainty (Fang et al., 2021; Novoselov et al., 2022). Such smoothing can obscure critical discontinuities that are essential for reliable horizon propagation. Because the proposed tracking workflow requires RGT contours that are both laterally continuous and locally sharp across faults and stratigraphic discontinuities, we use an adversarially regularized RGT prediction strategy to reduce smoothing and better preserve structurally meaningful RGT variations (Mosser et al., 2020). Its suitability for horizon tracking is evaluated against a pixel-wise U-Net baseline, a gradient-regularized ablation, and a published deep-RGT predictor in Section 2.4.

2.1. Network structure

The RGT prediction follows a Pix2Pix-style conditional framework (Cira et al., 2021), which provides a practical implementation of adversarially regularized image-to-image (or volume-to-volume) translation. The overall framework consists of a generator–discriminator pair (Fig. 1). To distinguish network tensors from the spatial coordinates used later in the tracking equations, bold symbols are used here: the generator G maps the seismic input tensor $\mathbf{x}$ to a predicted RGT field $G(\mathbf{x})$, while the discriminator D evaluates the local consistency between the seismic input and the corresponding RGT output. The discriminator is used only during training as a regularization mechanism and is not involved during inference or horizon tracking.

The generator is implemented as an encoder–decoder network with residual connections and attention mechanisms (Fig. 2a). While residual blocks ensure stable gradient propagation across scales, the attention mechanisms are specifically designed to emphasize spatial regions with high structural complexity, such as faults and stratigraphic terminations. Furthermore, skip connections are utilized to preserve fine-scale geometric information, which is indispensable for maintaining the continuity of horizon propagation.

To accommodate both 2D and 3D tracking scenarios, the generator is implemented in two configurations. The primary difference lies in the dimensionality of the convolutional operations and the resulting number of trainable parameters. Specifically, convolutional kernels are extended from 3×3 pixels in 2D to $3 \times 3 \times 3$ voxels in 3D, while the overall encoder–decoder topology, residual blocks, and attention mechanisms remain unchanged. The discriminator adopts a PatchGAN structure (Fig. 2b), which operates on overlapping local patches rather than producing a single global realism score.

The generator is trained using a combination of pixel-wise consistency and adversarial regularization. The overall objective function is given by

$$\mathcal{L}(G, D) = \mathbb{E}_{\mathbf{x}, \mathbf{y}} [\log D(\mathbf{x}, \mathbf{y})] + \mathbb{E}_{\mathbf{x}} [\log (1 - D(\mathbf{x}, G(\mathbf{x})))] + \lambda_{L1} \mathbb{E}_{\mathbf{x}, \mathbf{y}} [\|\mathbf{y} - G(\mathbf{x})\|_1], \quad (1)$$

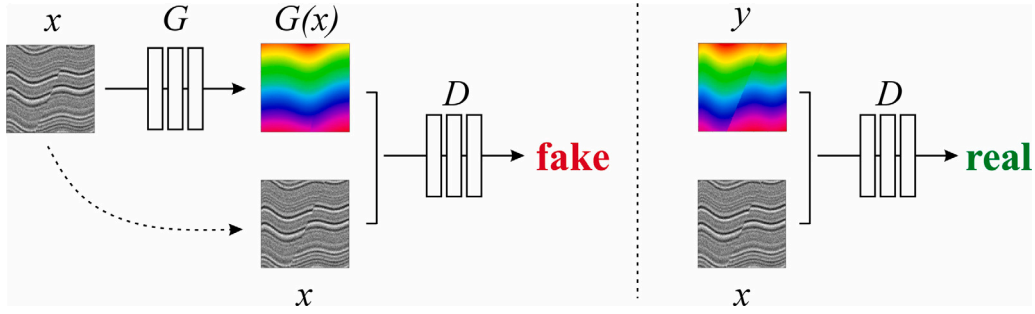


Fig. 1. Schematic illustration of the Pix2Pix-style conditional adversarial network used for RGT prediction. The generator G maps seismic input x to a predicted RGT field $G(x)$. The discriminator D distinguishes between generated RGT patches and reference RGT labels y , conditioned on the same seismic input.

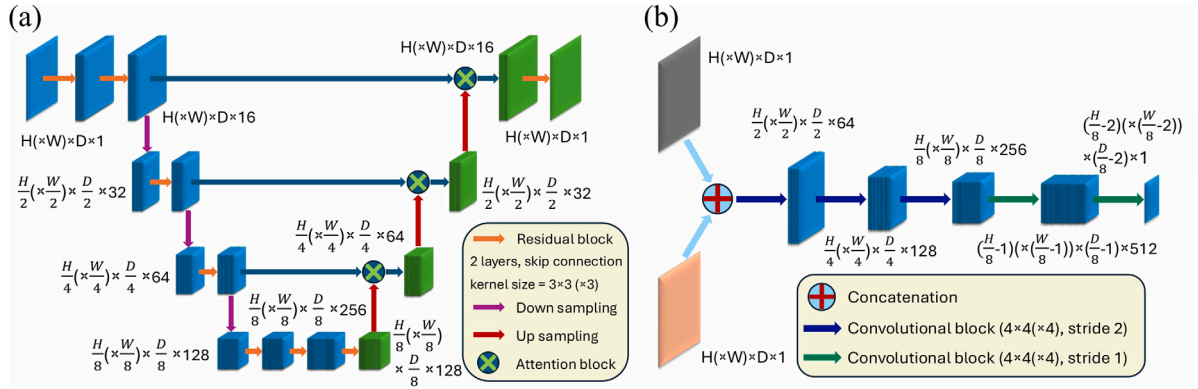


Fig. 2. Network architectures of the RGT predictor. (a) Generator consists of residual blocks, downsampling and upsampling paths, skip connections, and attention mechanisms to emphasize structurally relevant regions. It produces RGT predictions from seismic input in both 2D and 3D configurations. (b) PatchGAN discriminator architecture used during training to evaluate local consistency between seismic input and RGT output. It operates on overlapping patches and serves solely as a regularization mechanism to guide the generator toward structurally plausible RGT representations.

and the adversarial optimization is

$$G^* = \arg \min_G \max_D \mathcal{L}(G, D). \quad (2)$$

Here, x denotes the seismic input tensor, y denotes the reference RGT label tensor, G and D denote the generator and discriminator, respectively, and $G(x)$ is the predicted RGT field. The PatchGAN output $D(x, y)$ is a grid of conditional real-pair scores, with each local score in $(0, 1)$; the adversarial terms denote their average over spatial patches and training samples. The expectation $\mathbb{E}_{x,y}$ is taken over the empirical distribution of paired seismic-RGT training samples, whereas $\mathbb{E}_x$ is taken over its seismic-input marginal. The coefficient λ_{L1} controls the relative contribution of the pixel-wise ℓ_1 regression term and was set to 100 in all experiments. The combined objective balances global consistency and local structural regularization. The pixel-wise ℓ_1 term enforces overall agreement with reference RGT labels and preserves large-scale stratigraphic ordering. However, when used alone it tends to favor spatially averaged solutions that may blur sharp stratigraphic transitions in structurally complex regions. The adversarial regularization counteracts this tendency by penalizing locally implausible RGT patterns, encouraging the preservation of high-frequency structural features such as fault offsets and abrupt stratigraphic terminations. Although Eq. (1) does not explicitly enforce pointwise agreement between predicted and reference RGT gradients, the PatchGAN discriminator supplies patch-scale structural regularization. Whether an additional explicit gradient constraint improves geological fidelity is therefore evaluated empirically through the controlled ablation described in Section 2.4. Based on that comparison, the adversarial-plus- ℓ_1 objective in Eq. (1) is retained for the final framework.

2.2. Synthetic data preparation

To provide structurally controlled training data for RGT estimation, particularly across faults, we adopt an automated workflow to generate synthetic seismic data paired with RGT labels (Geng et al., 2020; Wu et al., 2020). The workflow is designed to explicitly encode stratigraphic deformation processes while maintaining a one-to-one correspondence between seismic reflections and RGT contours.

The generation process starts from a flat reflectivity model constructed from a 1D trace of random reflectivity values (Fig. 3a). Structural complexity is then progressively introduced through folding, planar shearing, and faulting operations (Fig. 3b–d). After structural deformation, the reflectivity model was convolved with a zero-phase Ricker wavelet whose peak frequency was uniformly sampled from $f_p \in [15, 45]$ Hz. Additive zero-mean Gaussian noise was then applied to the convolved seismic data, with the noise standard deviation defined as $\sigma_n = \alpha \text{rms}(s_{nf})$, where s_{nf} denotes the noise-free synthetic seismic image, $\text{rms}(\cdot)$ denotes root-mean-square amplitude, and α was uniformly sampled from $[0.2, 0.6]$. This produced the final synthetic seismic data with variable frequency content and noise levels (Fig. 3e).

In parallel, the initial RGT model is defined by monotonically increasing values in the vertical direction and laterally constant layers (Fig. 3f). This model is then subjected to the identical sequence of folding, shearing, and faulting operations applied to the seismic reflectivity model (Fig. 3g–i). This ensures that RGT contours remain exactly aligned with the deformed seismic structures, providing paired samples for training the RGT predictor and preserving sharp stratigraphic discontinuities, particularly across fault zones. To support both 2D and 3D spread tracking workflows, paired synthetic datasets are

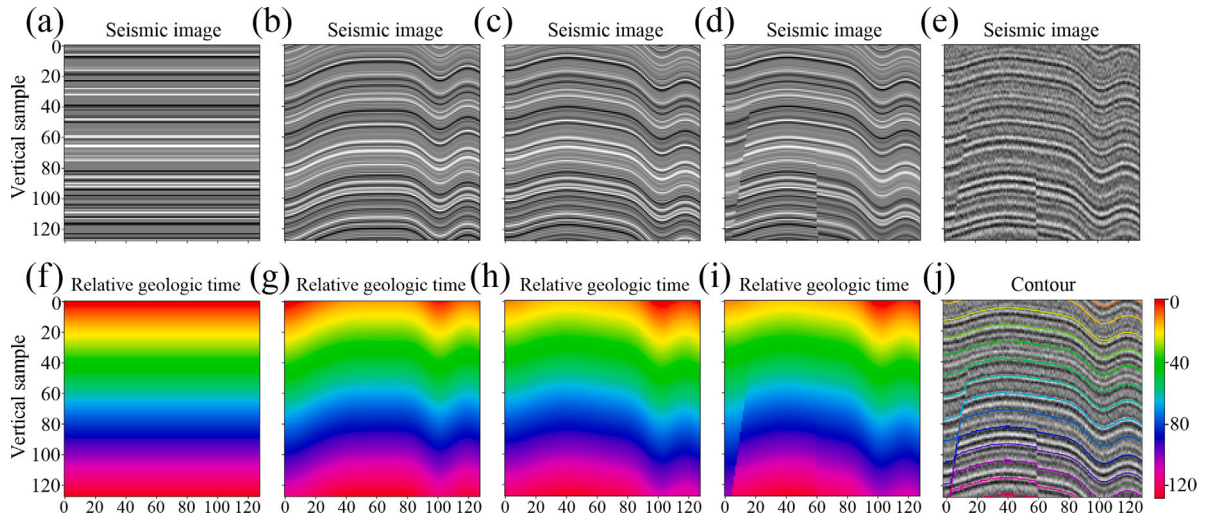


Fig. 3. Synthetic data generation workflow for seismic and RGT pairs. (a) Random reflectivity model and (f) its corresponding initial RGT. (b,g) Folding deformation, (c,h) planar shearing, and (d,i) faulting applied consistently to both seismic reflectivity and RGT models. (e) Final synthetic seismic data obtained by wavelet convolution and noise addition. (j) The resulting seismic–RGT pairs provide structurally consistent training samples with sharp fault-aligned RGT contours.

generated for 2D patches of size 128×128 pixels and 3D volumes of size $128 \times 128 \times 128$ voxels.

2.3. Training and validation strategies

Both 2D and 3D RGT predictors were trained to support the proposed multi-dimensional horizon tracking workflow. The synthetic dataset was divided into training and validation subsets. For the 2D implementation, 9600 paired seismic–RGT images were used for training and 1920 pairs were reserved for validation. For the 3D implementation, 320 paired seismic–RGT volumes were used for training and 64 volumes for validation. To reduce the domain gap between synthetic training data and real seismic data, all seismic and RGT inputs were normalized to the range $[0, 1]$ using min–max scaling.

All 2D models in the comparison were trained and evaluated using the same synthetic data partition and preprocessing. Their reported prediction metrics were calculated from identical held-out samples. For the downstream field comparison, the same seismic sections, manual reference horizons, prompt locations, spread-tracking algorithm, and tracking parameters were used for every RGT predictor. This design holds the tracking stage fixed and isolates the influence of the RGT prediction step to it.

Both 2D and 3D models utilized the Adam optimizer with an initial learning rate of 1×10^{-3} , first- and second-moment coefficients $\beta_1 = 0.5$ and $\beta_2 = 0.999$, numerical-stability constant $\epsilon_{\text{Adam}} = 1 \times 10^{-8}$, and zero weight decay. Under the reported step-decay schedule, the current learning rate was multiplied by 0.1 every 200 epochs for 2D and every 20 epochs for 3D to promote stable convergence.

All experiments were implemented in PyTorch 2.1.0 with CUDA 12.1 and cuDNN 8.9.4 and were conducted using a single NVIDIA RTX A4500 GPU with 20 GB of GPU memory and NVIDIA driver version 553.09. The 2D model was trained for 1000 epochs with a batch size of 16, requiring approximately 18 h in total. The 3D model was trained for 50 epochs with a batch size of 1 and required approximately 15 h. The smaller 3D batch size reflects the substantially higher memory demand of volumetric convolution.

For clarity and reproducibility, the data-generation and optimization parameters are summarized in Table 1. Here, η_0 is the initial learning rate, γ_{lr} is the multiplicative learning-rate decay factor, and E_{lr} is the number of epochs between successive decay operations.

Due to the adversarially regularized training strategy, the loss curves do not exhibit monotonic convergence but instead show

characteristic oscillatory behavior (Fig. 4). These fluctuations reflect the dynamic interaction between the generator and discriminator during optimization. As a result, standard loss values alone are insufficient for selecting an optimal model state. To address this, model checkpoints saved during training were evaluated qualitatively on withheld validation data to assess their ability to delineate sharp stratigraphic boundaries.

Validation results indicate that early-stage models (e.g., at 200 epochs for the 2D case) capture the overall stratigraphic trends but lack the local sharpness required for reliable fault delineation (Fig. 5). As training progresses, the predicted RGT fields gradually exhibit improved continuity and sharper stratigraphic boundaries. At later stages (e.g., 1000 epochs for the 2D model), the RGT predictions accurately follow seismic reflections and preserve fault offsets, providing the level of structural fidelity required by the proposed spread tracking workflow. Based on this validation analysis, the model checkpoints at the 998th epoch for the 2D predictor and the 49th epoch for the 3D predictor were selected as the final inference models used in all subsequent horizon tracking experiments.

2.4. Loss-function ablation and comparison with an existing deep RGT method

Four 2D RGT predictors were evaluated to distinguish pixel-wise reconstruction accuracy, adversarial regularization, explicit gradient fidelity, and downstream tracking performance. U-Net–L2 provides a conventional encoder–decoder baseline trained using a purely pixel-wise ℓ_2 objective. The proposed Pix2Pix predictor combines an encoder–decoder generator, the ℓ_1 consistency term in Eq. (1), and PatchGAN-based adversarial regularization. Pix2Pix-Grad uses the same architecture and training configuration as the proposed Pix2Pix model but additionally incorporates an explicit RGT-gradient fidelity term. The 2D network of Geng et al. (2020) was included as a representative published deep-learning RGT baseline. All four predictors were evaluated on the same synthetic samples and were coupled to the same prompt-guided spread-tracking procedure for the field comparison.

For the gradient-loss ablation, the added term is

$$\mathcal{L}_{\text{grad}} = \mathbb{E}_{\mathbf{x}, \mathbf{y}} [\|\nabla G(\mathbf{x}) - \nabla \mathbf{y}\|_1], \quad (3)$$

and the gradient-regularized objective is

$$\mathcal{L}_{\text{Pix2Pix-Grad}} = \mathcal{L}(G, D) + \lambda_g \mathcal{L}_{\text{grad}}, \quad (4)$$

Table 1

Data-generation, training, and validation parameters used for the 2D and 3D RGT predictors. The same value in both columns indicates a dimension-independent setting.

Parameter	2D predictor	3D predictor
Input size	128 × 128 pixels	128 × 128 × 128 voxels
Convolutional kernel	3 × 3	3 × 3 × 3
Training pairs	9600	320
Validation pairs	1920	64
Seismic and RGT normalization	[0, 1] (min-max scaling)	
Wavelet peak frequency, f_p	$\mathcal{U}(15, 45)$ Hz	
Noise scale, α in $\sigma_n = \alpha \text{rms}(s_{\text{nr}})$	$\mathcal{U}(0.2, 0.6)$	
Element-wise ℓ_1 loss weight, λ_{L_1}	100	
Optimizer	Adam	Adam
Adam coefficients, (β_1, β_2)	(0.5, 0.999)	(0.5, 0.999)
Adam stability constant, ϵ_{Adam}	1×10^{-8}	1×10^{-8}
Weight decay	0	0
Initial learning rate, η_0	1×10^{-3}	1×10^{-3}
Decay factor, γ_{lr}	0.1	0.1
Decay interval, E_{lr}	200 epochs	20 epochs
Training epochs	1000 epochs	50 epochs
Batch size	16	1
Selected checkpoint	epoch 998	epoch 49

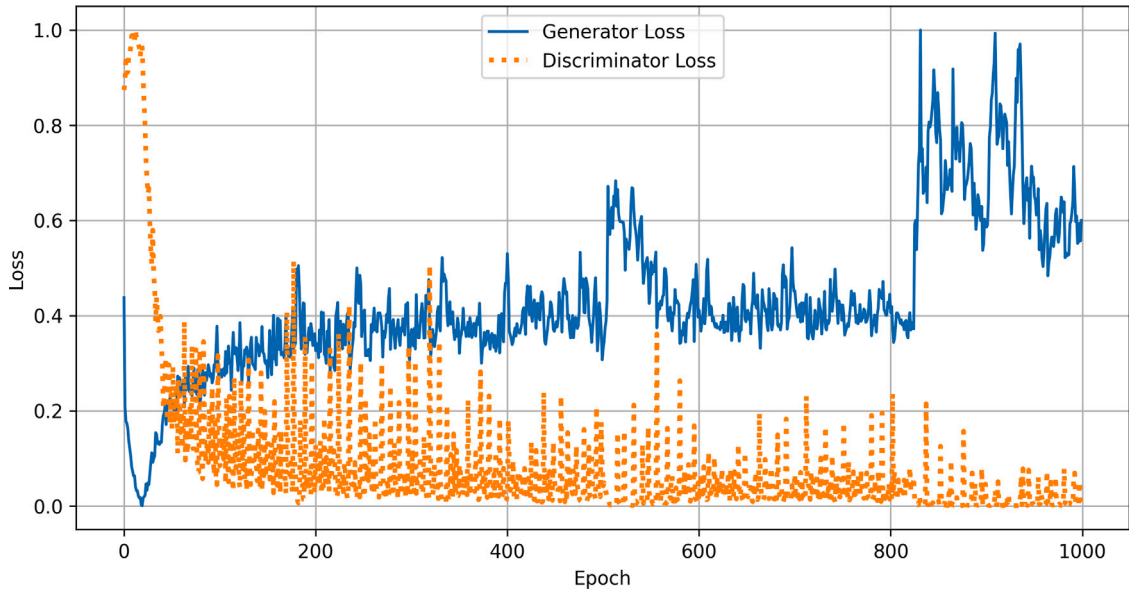


Fig. 4. Training loss curves for the Pix2Pix-style RGT predictor. The generator loss (blue) and discriminator loss (orange) exhibit independent oscillations, reflecting the adversarial optimization process. Despite the absence of monotonic convergence, this dynamic equilibrium guides the generator toward structurally consistent RGT predictions suitable for subsequent horizon tracking.

where ∇ denotes the finite-difference spatial gradient and λ_g controls the contribution of the gradient-fidelity term. We evaluated $\lambda_g \in \{1, 5, 10, 20\}$ on the held-out validation samples. The best validation performance was obtained with $\lambda_g = 1$, which was therefore used for the Pix2Pix-Grad results reported below.

In addition to the image-domain ℓ_1 , SSIM, LPIPS, and GMSD metrics (Xue et al., 2014; Zhang et al., 2018), three geology-oriented measures were calculated. RGT-gradient MAE averages the component-wise absolute errors in the horizontal and vertical RGT derivatives:

$$E_{\nabla R} = \frac{1}{2} \langle |G_x - y_x| + |G_z - y_z| \rangle. \quad (5)$$

The vertical derivative controls local RGT-contour spacing and therefore provides a proxy for apparent stratigraphic-thickness fidelity, whereas the ratio between horizontal and vertical derivatives controls the local isocontour dip. Dip MAE directly compares the reference and predicted dip angles, with

$$\theta(R) = \text{atan2}(-R_x, |R_z|). \quad (6)$$

To evaluate discontinuity preservation without the unstable conversion of RGT jumps into vertical samples, fault-zone jump MAE compares signed, locally detrended adjacent-trace RGT jumps. For an RGT field R , define

$$q_R(x, z) = R(x+1, z) - R(x, z), \quad J_R(x, z) = q_R(x, z) - \text{median}_{\xi \in \mathcal{W}_x} q_R(\xi, z), \quad (7)$$

where $\mathcal{W}_x$ is a nine-trace local horizontal window used to remove the smooth structural trend. The reported error is

$$E_{\text{jump}} = \left\langle |J_{G(x)} - J_y| \right\rangle_{\mathcal{M}_f}, \quad (8)$$

evaluated only at reference fault edges $\mathcal{M}_f$. Because explicit synthetic fault masks were not retained by the validation loader, the reference edges were identified from locally detrended ground-truth RGT jumps using a robust median-absolute-deviation criterion. Fault-zone jump MAE is reported in RGT units and does not divide by the predicted vertical RGT gradient.

Table 2 reveals a clear distinction between pixel-wise RGT reconstruction and geological suitability for horizon propagation. Geng

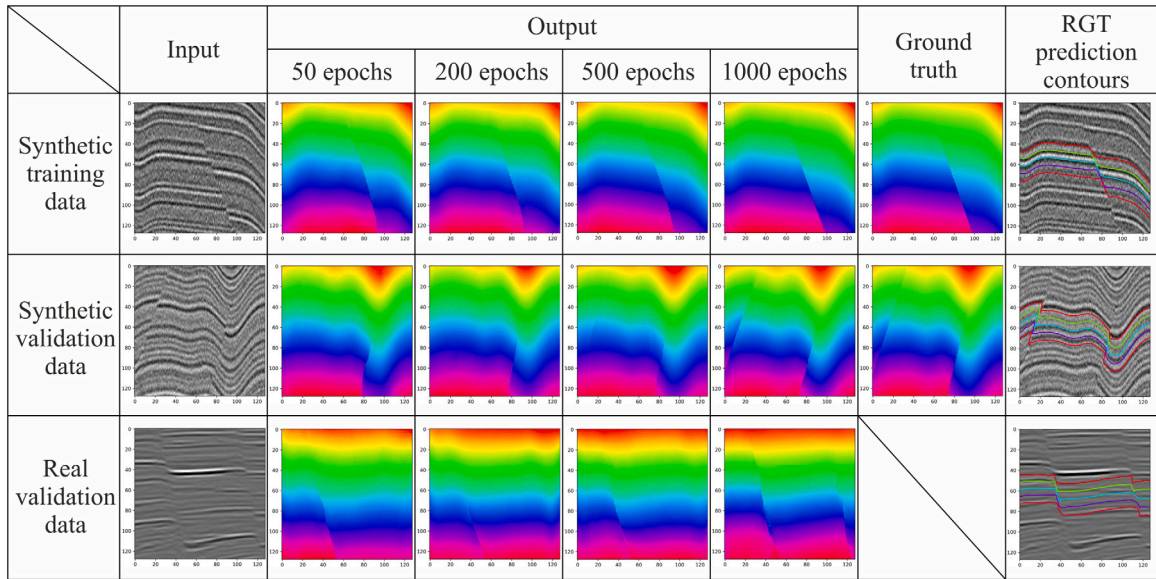


Fig. 5. Validation of the 2D RGT predictor on synthetic and real seismic data at different training stages. Early-stage predictions capture large-scale stratigraphic trends but appear overly smooth, whereas later-stage predictions exhibit sharper RGT contours and improved fault representation, which are critical for stable horizon propagation.

Table 2

Quantitative comparison of 2D RGT prediction performance on the common synthetic validation samples. Lower values indicate better performance for all metrics except SSIM. RGT-gradient MAE is reported in RGT units per sample, dip MAE in degrees, and fault-zone jump MAE in RGT units. The best mean value in each column is shown in bold.

Model	ℓ_1	LPIPS	SSIM	GMSD	Grad MAE	Dip MAE	Fault MAE
	↓	↓	↑	↓	↓	↓	↓
Pix2Pix	0.0440	0.1048	0.9718	0.0493	0.004232	5.999	0.071662
U-Net-L2	0.0377	0.1517	0.9766	0.0511	0.004385	6.166	0.072598
Pix2Pix-Grad	0.0470	0.1140	0.9707	0.0556	0.005056	6.424	0.073708
Geng et al. (2020)	0.0369	0.1501	0.9780	0.0502	0.004310	6.131	0.072794

et al. (2020) achieves the lowest mean ℓ_1 error (0.0369) and highest mean SSIM (0.9780), while U-Net-L2 also obtains a lower ℓ_1 error than the proposed model. In contrast, the proposed Pix2Pix predictor achieves the lowest LPIPS (0.1048), GMSD (0.0493), RGT-gradient MAE (0.004232), dip MAE (5.999°), and fault-zone jump MAE (0.071662). These results indicate that lower average RGT-value error does not necessarily correspond to better preservation of the local variations required for prompt-guided horizon tracking.

Even with the best-performing tested weight, $\lambda_g = 1$, adding the explicit gradient-fidelity term does not improve the model relative to the original Pix2Pix objective. Pix2Pix-Grad increases RGT-gradient MAE from 0.004232 to 0.005056, dip MAE from 5.999° to 6.424°, and fault-zone jump MAE from 0.071662 to 0.073708. This result is specific to the uniform first-order gradient formulation and tested weight range and should not be interpreted as a general rejection of geological gradient constraints.

For the controlled 2D field benchmark, five representative sections were selected from each of the publicly available Opunake, F3, and SEG Kerry-3D datasets. The reference horizons were manually interpreted, and the same preprocessing and tracking protocol was applied to all three datasets. The downstream implications are illustrated in Fig. 6 using the Opunake, F3, and Kerry field sections. The same prompt positions, tracking algorithm, and tracking parameters were used for all four RGT predictors. The proposed Pix2Pix-driven workflow follows the reference interpretation more consistently across the normal fault in Opunake, the weak and uplifted reflectors in F3, and the multiple

disturbed segments in Kerry. A particularly clear qualitative difference occurs at the Opunake normal fault: the Pix2Pix-Grad-guided horizon rounds the abrupt fault-related offset into a visibly smoother transition (Fig. 6g), whereas the original Pix2Pix-guided result retains a sharper structural break (Fig. 6a). This visible smoothing is consistent with the higher fault-zone jump MAE of Pix2Pix-Grad and indicates that the spatially uniform gradient penalty can attenuate a genuine discontinuity rather than preserve it. The alternative predictors also show larger local departures where smoothing or local RGT errors accumulate along the propagation path.

The quantitative results for these 15 seismic sections were aggregated as follows. Each metric was first calculated for every section, averaged across the five sections within each dataset, and then averaged with equal weight across the three datasets. The values in Table 3 are therefore means of the three dataset-level results rather than statistics obtained by pooling all lateral positions from the three datasets.

The field results reinforce the distinction between RGT reconstruction scores and downstream tracking reliability. Averaged across the three datasets, the proposed model achieves the highest Acc@10 (91%), the lowest median absolute vertical error (5 samples), and the lowest mean maximum absolute vertical error (13 samples). The corresponding values are 82%, 6, and 25 samples for U-Net-L2; 85%, 7, and 19 samples for Pix2Pix-Grad; and 84%, 7, and 18 samples for the Geng et al. (2020) baseline. Relative to Geng et al. (2020), the proposed predictor therefore improves Acc@10 by 7 percentage points,

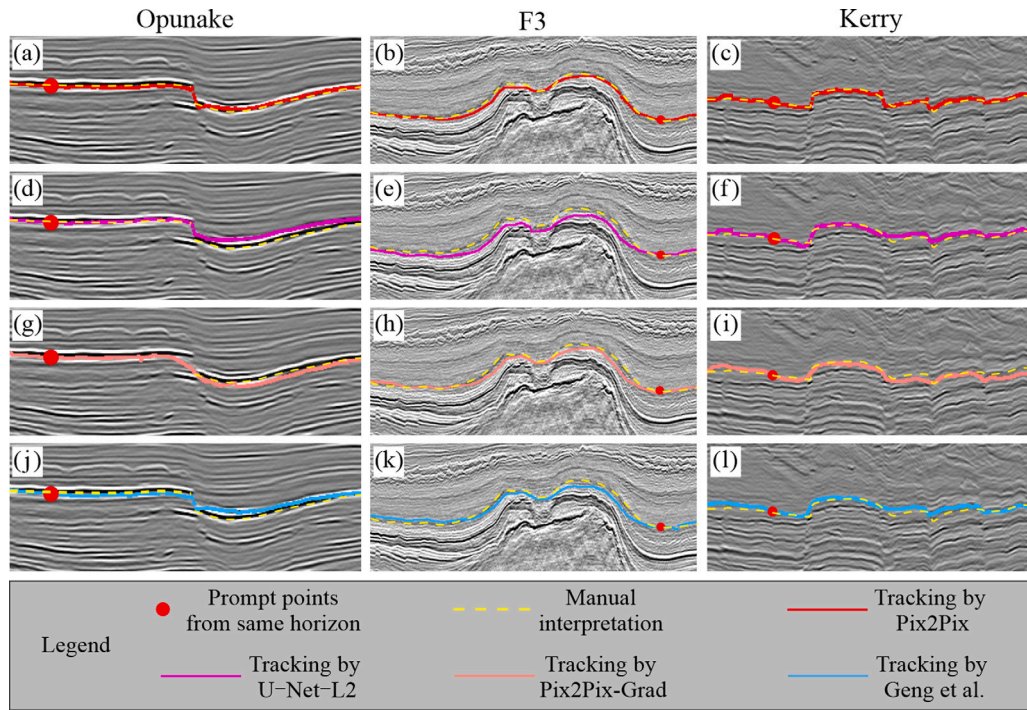


Fig. 6. Comparison of single-prompt horizon tracking guided by four RGT predictors on the Opunake (left column), F3 (middle column), and Kerry (right column) field sections. Rows (a–c), (d–f), (g–i), and (j–l) show results from the proposed Pix2Pix predictor, U-Net-L2, Pix2Pix-Grad, and Geng et al. (2020), respectively. Yellow dashed lines denote manual interpretations, colored solid lines denote tracked horizons, and red circles denote the common prompt locations. The same prompts, spread-tracking algorithm, and tracking parameters were used in every row, so the comparison isolates the influence of the RGT predictor on downstream tracking. The rounded transition across the Opunake normal fault in the Pix2Pix-Grad result (g), compared with the sharper offset retained by Pix2Pix in (a), illustrates the visible smoothing of a structural discontinuity associated with the uniform gradient penalty.

Table 3

Horizon-tracking performance across 15 field sections: five sections from each of the Opunake, F3, and Kerry datasets. Reported values are equal-weight means of the three dataset-level results. Acc@10 is the percentage of evaluated lateral positions whose absolute vertical deviation from the manual interpretation is no greater than 10 samples. Median absolute vertical error is the median pointwise absolute deviation within a section, whereas mean maximum absolute vertical error is obtained by averaging the section-level maximum deviations using the stated dataset-level aggregation.

Model	Acc@10 (%) ↑	Median error (samples) ↓	Mean max. error (samples) ↓
Pix2Pix	91	5	13
U-Net-L2	82	6	25
Pix2Pix-Grad	85	7	19
Geng et al. (2020)	84	7	18

reduces the median absolute error by 2 samples, and reduces the mean maximum error by approximately 28%, despite Geng et al. (2020) achieving a lower mean ℓ_1 error and a higher mean SSIM. The original adversarial-plus- ℓ_1 objective is consequently retained in the final framework.

3. RGT-guided spread tracking framework

In this section, an RGT-guided spread tracking framework is introduced for seismic horizon interpretation. A prompt refers to a user-specified horizon point that acts as a geometric constraint to initiate and guide horizon propagation. Rather than extracting horizons globally in a single step, the proposed approach propagates horizon geometry locally from one or multiple prompts under stratigraphic consistency constraints imposed by RGT predictions.

The key idea behind spread tracking is to treat horizon interpretation as a controlled growth process. RGT serves as a guidance field encoding stratigraphic ordering, while the tracking behavior is governed by propagation rules rather than direct amplitude correlation.

In 2D sections, horizon lines are propagated from a single prompt through directional spreading, while in 3D volumes the same principle is extended to incremental surface growth. Multiple prompts can further be incorporated as redundant stratigraphic constraints to enhance robustness and suppress long-range error accumulation.

3.1. 2D RGT-guided spread tracking with one prompt

The 2D RGT-guided tracking method with a single prompt is designed to recover horizon lines on seismic sections of arbitrary size using only one user-specified high-confidence point, such as well data or other prior constraint. In this framework, the prompt acts as a geometric anchor that initiates local horizon propagation under RGT consistency constraints.

Let $I(x, z)$ denote a 2D seismic section and $R(x, z)$ the corresponding predicted RGT field, where x and z are the lateral and vertical sample coordinates, respectively. Given a user-specified prompt point $\mathbf{p}_0 = (x_0, z_0)$, the target RGT is defined as $r_0 = R(x_0, z_0)$. For any point $\mathbf{p} = (x_p, z_p)$, the shorthand $R(\mathbf{p})$ means $R(x_p, z_p)$. Let $\mathcal{A}(\mathbf{p})$ denote the

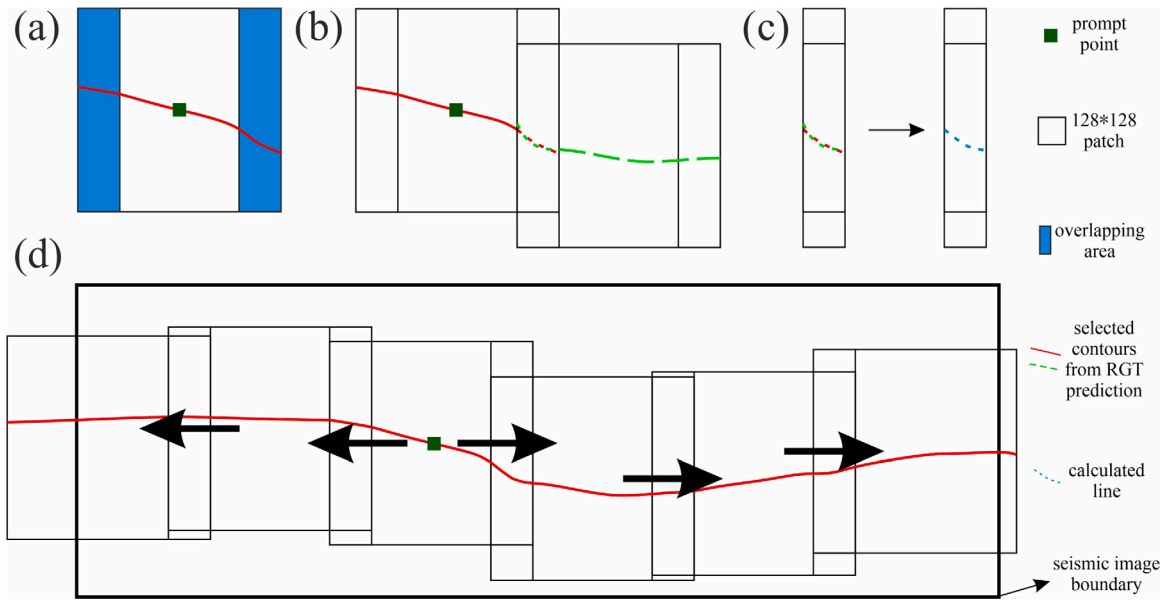


Fig. 7. Illustration of the 2D RGT-guided spread tracking process with a single prompt. (a) An initial horizon segment is extracted in a local patch by selecting the RGT contour consistent with the prompt-defined stratigraphic level. (b) The analysis window propagates laterally with overlap, where a robust reference horizon coordinate is estimated from the overlapping horizon and used to guide tracking in the next patch. (c) Horizon segments from adjacent patches are smoothly fused within the overlapping region using weighted least-squares blending to ensure continuity. (d) Repeating these steps yields a continuous horizon across the entire seismic section, robust to structural complexity and patch-level prediction variability.

set of connected candidate curves within the current patch that pass through the reference point $\mathbf{p}$ and span the available lateral extent of the patch. The initial horizon segment H_0 is defined as the candidate curve that best preserves the prompt-defined RGT value:

$$H_0 = \arg \min_{\Gamma \in \mathcal{A}(\mathbf{p}_0)} \int_{\Gamma} |R(x, z) - r_0|^2 d\ell. \quad (9)$$

Here, Γ is a candidate horizon curve and $d\ell$ is its arc-length element. Numerically, this minimization is implemented as an RGT contour-extraction step. The predicted RGT field is linearly interpolated within the local patch, and the iso-contour corresponding to r_0 is extracted. When several disconnected contour segments are present, the segment passing through, or nearest to, the prompt point is retained as H_0 .

Horizon propagation is achieved through directional spreading based on overlapping local patches. Let P_x^{2D} and P_z^{2D} denote the lateral and vertical dimensions of each 2D analysis patch, respectively. Starting from the initial patch centered at $\mathbf{p}_0$, the analysis window is shifted laterally by a fixed propagation step Δx^{2D} , defined as the horizontal distance between the centers of two adjacent patches. The corresponding lateral overlap width is $O_x^{2D} = P_x^{2D} - \Delta x^{2D}$. The lateral interval shared by two successive patches is denoted by $\Omega_{2D} \subset \mathbb{R}$, and dx denotes the lateral integration element on this interval. Let k index the current patch and let j index the sampled horizon points (x_j, z_j) within the overlap. A scalar reference horizon coordinate for the next patch is estimated by solving the robust optimization problem

$$z_{k+1}^{\text{ref}} = \arg \min_{\zeta \in \mathbb{R}} \sum_{\substack{(x_j, z_j) \in \mathcal{H}_k \\ x_j \in \Omega_{2D}}} \rho(\zeta - z_j), \quad \rho(u) = |u|, \quad (10)$$

where $\mathcal{H}_k$ denotes the horizon segment extracted from the current patch, ζ is the scalar optimization variable, and ρ is the absolute-value penalty. The solution z_{k+1}^{ref} is the median vertical coordinate of the overlapping horizon points, providing robustness against local outliers and small-scale prediction variability. The propagated reference point is therefore $\mathbf{p}_{k+1} = (x_k + \Delta x^{2D}, z_{k+1}^{\text{ref}})$, where x_k is the lateral coordinate of the current patch center.

Given the propagated reference point, the horizon segment in the new patch is extracted by enforcing RGT consistency:

$$H_{k+1} = \arg \min_{\Gamma \in \mathcal{A}(\mathbf{p}_{k+1})} \int_{\Gamma} |R(x, z) - R(\mathbf{p}_{k+1})|^2 d\ell, \quad (11)$$

thereby constraining each propagation step to remain within a narrow stratigraphic band defined by the RGT field (Fig. 7b). This local constraint prevents abrupt stratigraphic jumps during tracking.

Because successive patches overlap, two horizon estimates may coexist within Ω_{2D} . To ensure continuity and suppress seam artifacts, the final horizon in the overlapping region is obtained by solving a locally weighted least-squares problem:

$$\hat{z}^{2D}(\cdot) = \arg \min_{z: \Omega_{2D} \rightarrow \mathbb{R}} \sum_{i \in \{k, k+1\}} \int_{\Omega_{2D}} \omega_i^{2D}(x) (z(x) - z_i^{2D}(x))^2 dx, \quad (12)$$

which admits the closed-form solution

$$\hat{z}^{2D}(x) = \frac{\omega_k^{2D}(x) z_k^{2D}(x) + \omega_{k+1}^{2D}(x) z_{k+1}^{2D}(x)}{\omega_k^{2D}(x) + \omega_{k+1}^{2D}(x)}, \quad \omega_i^{2D}(x) = \exp \left[-\frac{(x - c_i^{2D})^2}{2\sigma_{2D}^2} \right], \quad (13)$$

where $z_k^{2D}(x)$ and $z_{k+1}^{2D}(x)$ denote the horizon-coordinate functions contributed by two adjacent patches, c_i^{2D} is the lateral coordinate of the center of patch i , and σ_{2D} is the standard deviation that controls the spatial width of the 2D Gaussian blending function (Fig. 7c). This fusion strategy balances continuity between neighboring patches while reducing edge effects from patch-level RGT predictions.

The propagation and overlap-based refinement steps are repeated iteratively until the boundary of the seismic section is reached, yielding a continuous 2D horizon line across the entire image (Fig. 7d). In our implementation, the patch dimensions were set to $P_x^{2D} = P_z^{2D} = 128$ pixels and the lateral propagation step was set to $\Delta x^{2D} = 100$ pixels. The corresponding overlap width was therefore $O_x^{2D} = P_x^{2D} - \Delta x^{2D} = 28$ pixels. The Gaussian blending width was fixed at one quarter of the lateral patch size, $\sigma_{2D} = P_x^{2D}/4 = 32$ pixels (see Table 4).

Table 4

Symbols and implementation values for single-prompt 2D spread tracking. A pixel denotes one sample on the lateral or vertical image grid.

Symbol or operation	Value	Meaning
P_x^{2D}	128 pixels	Lateral patch size
P_z^{2D}	128 pixels	Vertical patch size
Δx^{2D}	100 pixels	Patch-center propagation step
O_x^{2D}	28 pixels	Lateral overlap width
Ω_{2D}	width 28 pixels	Shared lateral interval
$\rho(u)$	$ u $	Robust median penalty
RGT contour interpolation	linear	Subsample contour extraction
$\omega_i^{2D}(x)$	dimensionless	Local 2D Gaussian blending weight
σ_{2D}	32 pixels	Gaussian width (one quarter of P_x^{2D})
dx	one lateral-coordinate element	Integration measure, not a tuning parameter
$d\ell$	one curve arc-length element	Integration measure, not a tuning parameter

By restricting RGT prediction and horizon selection to local regions, the proposed spread tracking strategy minimizes unnecessary inference and explicitly limits error accumulation, leading to stable horizon delineation across faults and structurally complex regions.

3.2. 3D RGT-guided tracking method with one prompt

The 3D RGT-guided tracking method generalizes the 2D spread tracking strategy to volumetric seismic data for tracking continuous horizon surfaces from a single user-provided prompt point. Although both 2D and 3D tracking rely on the same RGT-guided principle, the 3D implementation differs fundamentally in the nature of the tracked object and the topology of propagation. Rather than extracting a global isosurface, the horizon surface is constructed progressively through controlled spatial spreading using overlapping local volumes, thereby explicitly limiting error accumulation in structurally complex settings.

Let $I(x, y, z)$ denote a 3D seismic volume and $R(x, y, z)$ the corresponding predicted RGT volume, where x , y , and z are the inline, crossline, and vertical sample coordinates, respectively. Given a prompt point $\mathbf{p}_0 = (x_0, y_0, z_0)$, the target horizon is defined by the associated RGT value $r_0 = R(x_0, y_0, z_0)$. In contrast to the 2D case, where the tracking result is a horizon line, the 3D tracking result is a horizon surface $S \subset \mathbb{R}^3$ that is incrementally constructed from local RGT-consistent surface patches extracted within overlapping analysis cubes.

Let P_x^{3D} , P_y^{3D} , and P_z^{3D} denote the inline, crossline, and vertical dimensions of a local 3D analysis cube, respectively. In our implementation, $P_x^{3D} = P_y^{3D} = P_z^{3D} = 128$ voxels. The tracking process is initialized by extracting a $128 \times 128 \times 128$ -voxel seismic cube centered at the prompt point, from which the initial RGT-consistent surface patch is obtained (Fig. 8a). This surface patch serves as the initial seed surface for subsequent spatial propagation.

Surface propagation proceeds through spatial expansion over a 2D lateral neighborhood while maintaining volumetric overlap between adjacent cubes. The inline and crossline propagation steps are denoted by Δx^{3D} and Δy^{3D} , respectively. The corresponding overlap widths are defined as $O_x^{3D} = P_x^{3D} - \Delta x^{3D}$ and $O_y^{3D} = P_y^{3D} - \Delta y^{3D}$. In the implementation reported here, $\Delta x^{3D} = \Delta y^{3D} = 100$ voxels, giving $O_x^{3D} = O_y^{3D} = 28$ voxels. This volumetric overlap introduces a stronger topological constraint than the strip-like overlap used in 2D and plays a critical role in stabilizing surface growth. As illustrated in the top-view schematic (Fig. 8b), overlap configurations may involve either pairwise overlap along a single propagation direction or multi-way overlap at junctions where inline and cross-line spreading intersect. Incremental surface growth is first performed along the inline direction through successive overlapping cubes (Fig. 8c), followed by propagation along the cross-line direction to achieve multi-directional surface expansion (Fig. 8d). Rather than treating these cases separately, both scenarios are handled within a unified surface fusion framework.

Let $\Omega_{3D} \subset \mathbb{R}^2$ denote the lateral area shared by the contributing cubes and let $dA = dx dy$ denote its area element. Let $\{z_k^{3D}(x, y)\}_{k=1}^K$ denote the local surface-coordinate functions contributed by K

overlapping cubes within Ω_{3D} , where $K = 2$ corresponds to pairwise overlap and $K = 4$ corresponds to quadruple overlap. The fused surface is obtained by solving a locally weighted least-squares problem,

$$\hat{z}^{3D}(\cdot, \cdot) = \arg \min_{z: \Omega_{3D} \rightarrow \mathbb{R}} \sum_{k=1}^K \int_{\Omega_{3D}} \omega_k^{3D}(x, y) (z(x, y) - z_k^{3D}(x, y))^2 dA, \quad (14)$$

which admits the closed-form solution

$$\hat{z}^{3D}(x, y) = \frac{\sum_{k=1}^K \omega_k^{3D}(x, y) z_k^{3D}(x, y)}{\sum_{k=1}^K \omega_k^{3D}(x, y)}. \quad (15)$$

The weighting function $\omega_k^{3D}(x, y)$ is defined as a Gaussian distance-based weight:

$$\omega_k^{3D}(x, y) = \exp \left[-\frac{\left(x - c_{k,x}^{3D} \right)^2 + \left(y - c_{k,y}^{3D} \right)^2}{2\sigma_{3D}^2} \right], \quad (16)$$

where $(c_{k,x}^{3D}, c_{k,y}^{3D})$ denotes the lateral center of the k th cube and σ_{3D} is the standard deviation controlling the spatial width of the 3D Gaussian blending zone. This weighting assigns higher confidence to surface estimates near cube centers while smoothly attenuating contributions toward cube boundaries. This formulation naturally generalizes pairwise overlap in 2D tracking to multi-way overlap in 3D and leverages the increased geometric redundancy to suppress seam artifacts and surface warping induced by local prediction inconsistencies.

The Gaussian blending width was fixed at one quarter of the lateral cube size, $\sigma_{3D} = \min(P_x^{3D}, P_y^{3D})/4 = 32$ voxels. Table 5 summarizes the symbols and numerical values used in the 3D implementation.

The propagation and overlap-based fusion steps are repeated iteratively until the boundary of the seismic volume is reached, yielding a continuous 3D horizon surface derived from a single prompt point (Fig. 8e). By emphasizing incremental surface growth and volumetric overlap control rather than global isosurface extraction, the proposed method maintains robustness for large-scale 3D seismic interpretation in structurally complex geological settings.

3.3. Multi-prompt constrained RGT-guided tracking

Single-prompt RGT-guided tracking serves as the fundamental engine for local horizon growth; however, it lacks the redundant constraints necessary to ensure long-range structural integrity. In the presence of complex structural variations, a single stratigraphic anchor may not sufficiently prevent the gradual drift of horizon geometry in areas far from the initialization point. To further enhance robustness, the proposed framework naturally extends to a multi-prompt constrained tracking strategy by incorporating multiple independent stratigraphic anchor points, such as well ties or other high-confidence prior interpretations.

Let M denote the number of user-provided prompt points associated with the same geological horizon. For each prompt, an independent RGT-guided spread tracking process is performed, producing a horizon

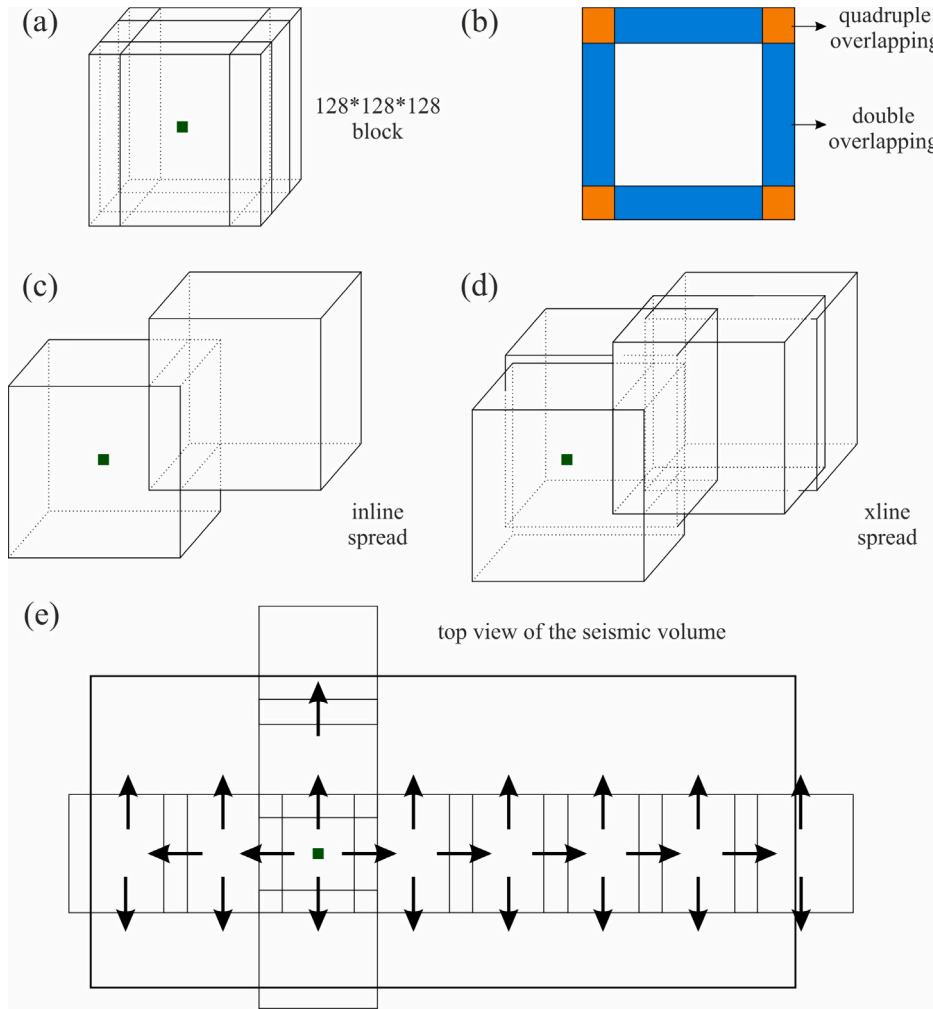


Fig. 8. Schematic illustration of the 3D RGT-guided spread tracking framework with a single prompt. (a) Initialization of tracking by extracting an RGT-consistent surface patch from a local $128 \times 128 \times 128$ seismic cube centered at the prompt point. (b) Schematic of volumetric overlap configurations, where pairwise (double) overlap occurs along a single propagation direction and quadruple overlap arises at junctions of inline and cross-line spreading. (c) Incremental surface propagation along the inline direction using overlapping local cubes. (d) Subsequent surface expansion along the cross-line direction to achieve multi-directional growth of the horizon surface. (e) Top-view schematic of the seismic volume showing the overall spreading strategy, where controlled volumetric overlap and weighted surface fusion ensure continuity during propagation.

Table 5

Symbols and implementation values for single-prompt 3D spread tracking. A voxel denotes one sample on each axis of the seismic volume.

Symbol or operation	Value	Meaning
$p_x^{3D}, p_y^{3D}, p_z^{3D}$	128 voxels each	Inline, crossline, and vertical cube sizes
$\Delta x^{3D}, \Delta y^{3D}$	100 voxels each	Cube-center propagation steps
O_x^{3D}, O_y^{3D}	28 voxels each	Inline and crossline overlap widths
K	2 or 4	Number of overlapping cube estimates
Ω_{3D}	pairwise or four-way overlap	Shared lateral fusion area
$\alpha_k^{3D}(x, y)$	dimensionless	Local 3D Gaussian blending weight
σ_{3D}	32 voxels	Gaussian width (one quarter of the lateral cube size)
$dA = dx dy$	lateral area element	Integration measure, not a tuning parameter

line in 2D or a horizon surface in 3D. These independently propagated results represent multiple valid local solutions constrained by different stratigraphic anchors (Fig. 9a).

Rather than treating these solutions separately, the multi-prompt strategy enforces spatial consistency by integrating them through distance-based weighting. Specifically, the influence of each prompt is

assumed to decay with increasing spatial distance, reflecting the fact that tracking reliability generally decreases and cumulative deviation may increase as propagation moves away from its originating constraint. Because the spatial domains differ between 2D horizon lines and 3D horizon surfaces, their weighting functions are defined separately.

For 2D tracking, let $\mathcal{P}^{2D} = \{\mathbf{p}_i^{2D}\}_{i=1}^M$, with $\mathbf{p}_i^{2D} = (x_i, z_i)$, denote the prompt set, and let $T_i^{2D}(x)$ denote the horizon-coordinate function propagated from the i th prompt. The lateral distance from location x to the i th prompt is $d_i^{2D}(x) = |x - x_i|$. For 3D tracking, let $\mathcal{P}^{3D} = \{\mathbf{p}_i^{3D}\}_{i=1}^M$, with $\mathbf{p}_i^{3D} = (x_i, y_i, z_i)$, denote the prompt set, and let $T_i^{3D}(x, y)$ denote the horizon surface propagated from the i th prompt. The corresponding planar distance is $d_i^{3D}(x, y) = \sqrt{(x - x_i)^2 + (y - y_i)^2}$. Lowercase d is used for distance so that it cannot be confused with the discriminator D in Section 2.1. The normalized contribution weights are

$$W_i^{2D}(x) = \begin{cases} 1, & d_i^{2D}(x) = 0, \\ 0, & d_i^{2D}(x) > 0 \text{ and } \min_{1 \leq j \leq M} d_j^{2D}(x) = 0, \\ \frac{[d_i^{2D}(x)]^{-1}}{\sum_{j=1}^M [d_j^{2D}(x)]^{-1}}, & \min_{1 \leq j \leq M} d_j^{2D}(x) > 0, \end{cases} \quad (17a)$$

$$W_i^{3D}(x, y) = \begin{cases} 1, & d_i^{3D}(x, y) = 0, \\ 0, & d_i^{3D}(x, y) > 0 \text{ and } \min_{1 \leq j \leq M} d_j^{3D}(x, y) = 0, \\ \frac{[d_i^{3D}(x, y)]^{-1}}{\sum_{j=1}^M [d_j^{3D}(x, y)]^{-1}}, & \min_{1 \leq j \leq M} d_j^{3D}(x, y) > 0. \end{cases} \quad (17b)$$

The inverse-distance exponent is fixed to one in both dimensions. Prompt coordinates are assumed to be distinct in the lateral domain; duplicate prompts, if present, are merged before weighting. The first two branches explicitly remove the inverse-distance singularity at a prompt location. The dimension-labeled symbols ω_i^{2D} and ω_k^{3D} in Sections 3.1–3.2 denote local patch- or cube-blending weights, whereas uppercase W_i^{2D} and W_i^{3D} here denote global weights assigned to independently propagated prompt solutions. The final multi-prompt constrained estimates are

$$T_{\text{integrated}}^{2D}(x) = \sum_{i=1}^M W_i^{2D}(x) T_i^{2D}(x), \quad (18a)$$

$$T_{\text{integrated}}^{3D}(x, y) = \sum_{i=1}^M W_i^{3D}(x, y) T_i^{3D}(x, y). \quad (18b)$$

These dimension-specific formulations can be interpreted as extensions of the weighted least-squares fusion used in overlap-based tracking, from local patch or cube overlap to spatially distributed prompt constraints (Fig. 9b). The current formulation uses lateral or planar distance as a practical proxy for prompt influence and does not explicitly account for structural barriers between the prompt and the evaluated location. Accordingly, the current multi-prompt formulation should be applied only when all prompts have been independently verified to represent the same target horizon and lie within a structural domain with reliable stratigraphic connectivity. Euclidean proximity alone should not be used to establish horizon correlation across major faults. Where interpreted fault traces or surfaces separate prompt locations, tracking and fusion should be performed independently within each fault block, or interpreter-defined fault masks should be used to exclude cross-fault contributions. Without such screening, spatially close prompts located in different structural blocks may contribute to the same integrated estimate despite lacking valid stratigraphic connectivity, potentially producing a geologically inappropriate blended horizon.

By integrating multiple independently propagated solutions, the proposed multi-prompt constrained tracking strategy introduces redundant stratigraphic control that actively suppresses long-range drift and stabilizes horizon geometry in complex geological settings. Importantly,

this strategy operates entirely at the tracking stage and does not require retraining or modifying the RGT prediction network. As a result, robustness is improved while preserving the flexibility and efficiency of the single-prompt framework.

3.4. Confidence assessment based on prompt-induced propagation stability

Horizon tracking in structurally complex settings can be sensitive to prompt initialization. Even under RGT-based stratigraphic constraints, small perturbations of the prompt location may produce different but geologically plausible horizon geometries. We use this sensitivity to define a conditional, stability-based confidence measure. Specifically, if locally perturbed prompts lead to similar tracking results, horizon propagation is considered stable and assigned high confidence. Conversely, large dispersion among the tracked results indicates increased sensitivity to prompt initialization and therefore lower conditional tracking confidence. Throughout this assessment, the trained RGT predictor and the corresponding predicted RGT field are held fixed. The proposed measure therefore characterizes prompt-induced propagation stability rather than the complete predictive uncertainty of the RGT-guided tracking framework. For compactness, the equations below are written for a 3D surface. The 2D definition is obtained by omitting the crossline coordinate y , replacing $T_i(x, y)$ with the horizon-coordinate function $T_i(x)$, and evaluating the resulting map as $C(x)$ rather than $C(x, y)$.

Let $\mathbf{p}_0 = (x_0, y_0, z_0)$ denote a reference prompt point on the target horizon, obtained from manual interpretation, well data, or other high-confidence prior information (Fig. 10a). Starting from $\mathbf{p}_0$, an RGT-guided spread tracking is performed to obtain a reference horizon surface $T_0(x, y)$, where $T_0(x, y)$ denotes the tracked vertical coordinate (time or depth) at location (x, y) .

To define “each side” unambiguously, let $\mathcal{Q}$ be the ordered lateral transect through $\mathbf{p}_0$ used for the local propagation test and let s be its sample coordinate, with $s = s_0$ at $\mathbf{p}_0$. In 2D, $\mathcal{Q}$ is the section trace in the x direction; in 3D, it is the active inline or crossline transect through the reference point. Let $n_{\text{side}} = 10$ be the number of perturbed samples selected on either side. The perturbation set is

$$\mathcal{P}_{\text{pert}}(\mathbf{p}_0) = \{\mathbf{p}(s_0 \pm m\Delta s) : m = 1, \dots, n_{\text{side}}\}, \quad (19)$$

where $\mathbf{p}(s)$ is the point on the reference horizon at transect coordinate s . Thus, the 10 nearest sampled horizon points on either side of $\mathbf{p}_0$ are used, giving $N_{\text{pert}} = 2n_{\text{side}} = 20$ perturbed prompts and a maximum perturbation radius $r_{\text{pert}} = n_{\text{side}}\Delta s = 10\Delta s$ (Fig. 10b). The sampling increment was set to $\Delta s = 1$ lateral grid sample. Enumerating the members of $\mathcal{P}_{\text{pert}}(\mathbf{p}_0)$ as $\{\mathbf{p}_i^{\text{pert}}\}_{i=1}^{N_{\text{pert}}}$, each point is used independently to generate a tracked horizon realization $T_i^{\text{pert}}(x, y)$ while the RGT field remains fixed (Fig. 10c and Table 6).

To evaluate the conditional propagation stability associated with $\mathbf{p}_0$, we examine the response of the tracked horizons at the reference lateral location (x_0, y_0) . Let $h_0 = T_0(x_0, y_0)$ denote the vertical coordinate (time or depth) of the reference tracking result, and let $h_i = T_i^{\text{pert}}(x_0, y_0)$, $i = 1, \dots, N_{\text{pert}}$, denote the vertical coordinate obtained from the perturbed-prompt tracking realizations. The local prompt-induced vertical-coordinate dispersion at $\mathbf{p}_0$ is defined operationally as the dispersion of these $N_{\text{pert}} + 1$ tracked vertical coordinate values and is quantified using the standard deviation

$$C(\mathbf{p}_0) = \sqrt{\frac{1}{N_{\text{pert}} + 1} \sum_{i=0}^{N_{\text{pert}}} (h_i - \bar{h})^2}, \quad (20)$$

with

$$\bar{h} = \frac{1}{N_{\text{pert}} + 1} \sum_{i=0}^{N_{\text{pert}}} h_i. \quad (21)$$

Here, $i = 0$ denotes the reference realization, $\bar{h}$ is the mean of all $N_{\text{eval}} = N_{\text{pert}} + 1 = 21$ evaluated vertical coordinates, and $C(\mathbf{p}_0)$ is their

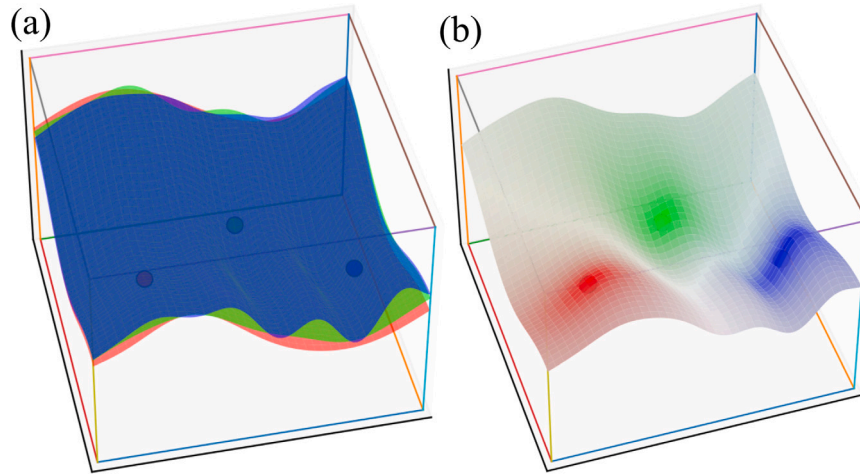


Fig. 9. Multi-prompt constrained RGT-guided spread horizon tracking. (a) Independent 3D RGT-guided horizons propagated from multiple prompt points. Each colored surface corresponds to a single-prompt constraint, illustrating that unconstrained long-range propagation can lead to divergent horizon geometries. (b) Multi-prompt constrained integrated horizon obtained via distance-weighted fusion. RGB color blending encodes the relative contribution of each prompt, with intensity indicating local weight dominance. The final horizon is computed as a spatially weighted combination of individual surfaces, producing a smooth integrated geometry under the stated same-horizon and structural-connectivity assumptions.

Table 6

Parameters and definitions used in the prompt-perturbation stability assessment.

Symbol or operation	Value	Meaning
N_{pert}	20	Number of perturbed prompts
n_{side}	10	Symmetric samples on either side along Q
N_{eval}	21	Reference plus perturbed realizations
Δs	1 lateral grid sample	Perturbation sampling increment
r_{pert}	$10\Delta s$	Maximum one-sided perturbation radius
$\mathcal{P}_{\text{pert}}(\mathbf{p}_0)$	$\pm 10\Delta s$ along Q	Perturbed-prompt set around $\mathbf{p}_0$
Q in 2D	section trace	Ordered perturbation direction
Q in 3D	active inline or crossline transect	Ordered perturbation direction
RGT prediction	fixed	No network or RGT perturbation
C	same unit as h	Dispersion; lower means more stable

population standard deviation. Consequently, C has the same physical unit as the horizon coordinate h (samples, time, or depth) and is not a probability.

Larger values of $C(\mathbf{p}_0)$ indicate stronger sensitivity of the tracked horizon to local prompt perturbations and therefore lower propagation stability. Smaller values indicate that neighboring prompt locations produce similar horizon geometries, corresponding to stable propagation and higher conditional tracking confidence. Accordingly, $C(\mathbf{p}_0)$ is used as an inverse proxy for prompt-induced propagation confidence: low $C(\mathbf{p}_0)$ corresponds to high-confidence tracking, whereas high $C(\mathbf{p}_0)$ identifies locations where the tracking result is sensitive to prompt initialization. Applying this procedure to reference points along the target horizon yields a spatially varying prompt-sensitivity map $C(x, y)$. This map is conditional on the fixed RGT prediction and should be interpreted as a propagation-stability diagnostic rather than as a calibrated probability of geological correctness or a complete estimate of predictive uncertainty.

4. Results

Seismic horizon tracking in field data is inherently challenged by geological complexity, such as pronounced stratigraphic variations, discontinuous reflectors, and fault-related deformations. These factors

introduce irregular amplitude patterns and lateral inconsistencies that often lead to tracking drift or ambiguous interpretations in traditional global methods. Following the three-dataset 2D predictor comparison in Section 2.4, the detailed framework evaluation in this section focuses on the F3 and Opunake datasets and uses the selected adversarial-plus- ℓ_1 Pix2Pix predictor throughout. Both the F3 and Opunake volumes are provided in the time domain with a temporal sampling interval of 4 ms; accordingly, their vertical coordinates and errors are reported in samples or milliseconds of two-way time (TWT). Our evaluation prioritizes the framework's ability to maintain stratigraphic consistency through a controlled growth process. The analysis follows a systematic workflow: initial propagation using single prompts in 2D and 3D, followed by a conditional propagation-stability diagnostic that provides a conditional, stability-based proxy for tracking confidence, and finally multi-prompt integration to incorporate high-confidence geological constraints. This design is specifically intended to demonstrate how the spread tracking strategy mitigates long-range error accumulation and adapts to varying degrees of geological ambiguity.

4.1. Application on field data with stratigraphic variations

The F3 dataset, acquired from the North Sea, contains complex sedimentary and structural features, including stratigraphic traps,

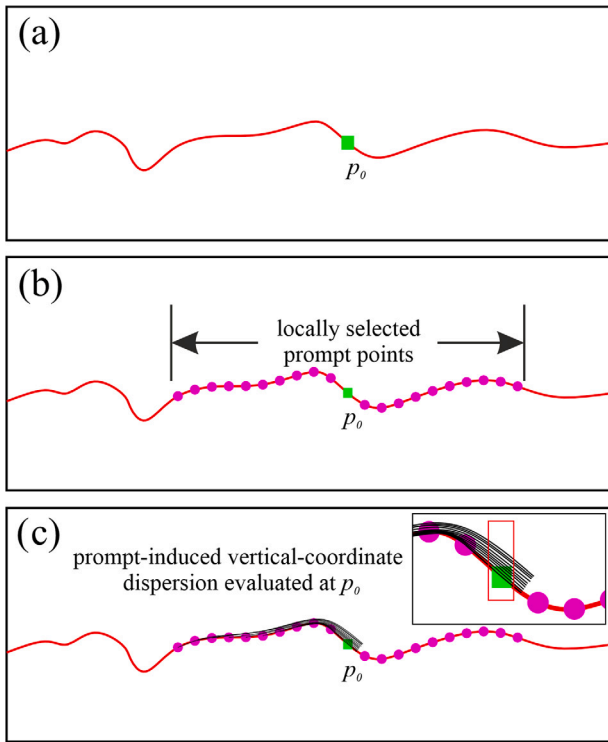


Fig. 10. Schematic illustration of the proposed conditional confidence assessment based on prompt-induced propagation stability. (a) Selection of a reference prompt point p_0 (green square) on a high-confidence reference horizon (red line). (b) Generation of multiple locally perturbed prompt points (magenta dots) within a fixed neighborhood along the reference horizon. (c) Independent propagation from each perturbed prompt under the same fixed RGT prediction, producing multiple horizon realizations used to quantify prompt-induced vertical-coordinate dispersion and conditional propagation confidence at p_0 .

sediment bodies, and folds, and provides manually interpreted horizons for Ref. (Kushwaha et al., 2021). The F3 volume is provided in the time domain with a temporal sampling interval of 4 ms and an inline and crossline bin size of 25 m $\times$ 25 m. For accuracy evaluation, we used a vertical tolerance of 10 samples (40 ms TWT).

Two target horizons at shallow (~ 400 ms) and deep (~ 1400 ms) levels were selected to test the framework under different degrees of geological complexity. The shallow horizon is relatively flat and characterized by clear reflections, although local discontinuities occur in areas affected by shallow data loss. In contrast, the deeper horizon is affected by local structural relief and lies within a lithological transition zone between carbonate-cemented streaks and sand-mudstone sequences (Sanda et al., 2020). The lack of continuous, high-amplitude reflectors at this depth makes conventional amplitude-based tracking unreliable, providing a suitable test for the proposed RGT-guided spread tracking framework (Fig. 11a).

We first evaluate single-prompt RGT-guided spread horizon tracking on 2D crossline sections, where the prompt points are selected from manually interpreted horizons. The tracked horizons closely follow the manually interpreted stratigraphic trends for both shallow and deep targets, yet they reveal distinct propagation behaviors that correlate with local structural complexity.

For the shallow horizon, the tracking result remains nearly coincident with the manual interpretation along the same stratigraphic level, with vertical differences within 10 samples (40 ms) at all lateral locations and within 5 samples (20 ms) at approximately 96% of

the locations, reflecting a laterally simple stratigraphic setting (Fig. 11a). In contrast, for the deeper horizon, high-amplitude seismic reflectors are absent and manual interpretations exhibit fluctuations. The spread tracking framework maintains lateral continuity and accurately captures large-scale dip changes associated with the uplift structure. Vertical differences remain within 10 samples at 91% of locations and within 5 samples at 78% of locations. Compared with the manual interpretation, the tracked result is smoother and follows the regional stratigraphic trend, indicating that horizon propagation is governed by structural context rather than being trapped by local amplitude inconsistencies. This robustness is further evidenced by the predicted RGT fields, which transition from laterally uniform patterns to complex dip variations within the central uplift, effectively capturing thickness changes and structural influences (Fig. 11b).

To further quantify the spatial variability of tracking behavior, prompt-induced time-dispersion maps derived from local prompt-point selection are analyzed (Fig. 12). This assessment reveals that regions associated with structural uplift and local data loss exhibit notably elevated dispersion values, typically ranging from 10 to 20 ms, with localized peaks exceeding 20 ms. In contrast, the majority of laterally uniform regions display consistently low dispersion values, generally below 10 ms, indicating low sensitivity to prompt initialization and stable, high-confidence propagation. These spatial patterns are closely correlated with variations in stratigraphic curvature, thickness changes, and reflector continuity.

3D RGT-guided spread tracking is further applied to the target horizons to evaluate the framework's performance in volumetric seismic data (Fig. 13a). The shallow horizon surface exhibits a laterally smooth geometry with minimal undulation, reflecting the depositional stability of the shallow strata and the absence of significant post-depositional deformation. In contrast, the deeper horizon surface clearly delineates a pronounced uplift structure, including well-defined grooves and gullies, indicating substantial stratigraphic and structural complexity. The tracked 3D horizon surface is geometrically consistent with the overlying carbonate strata, following regional stratigraphic trends. The fine contour-like bands visible on the rendered 3D horizons (Figures 13a, 15a, and 17a–c) mainly arise from representing the tracked surfaces on the finite inline-crossline-time grid and from shaded surface rendering. Although local RGT interpolation and overlap-weighted fusion are applied during tracking, small grid-scale variations in horizon time are accentuated in oblique 3D views, producing a terrace-like appearance. No additional display smoothing was applied because it could obscure genuine small-scale relief and fault offsets.

The corresponding prompt-induced time-dispersion maps reveal distinct differences between the two targets (Fig. 13b,c). The shallow horizon maintains uniformly high conditional propagation stability, with time dispersion below 10 ms over more than 80% of the area, indicating low prompt sensitivity, stable propagation, and high conditional tracking confidence. In contrast, the deeper horizon displays low conditional propagation stability (10–20 ms time dispersion, with localized peaks > 20 ms) concentrated around the uplift and structurally disturbed zones. These elevated values represent increased dispersion among horizon solutions generated from neighboring prompts, indicating lower conditional tracking confidence and reduced propagation stability in structurally deformed regions.

The F3 tests used a 951 $\times$ 463 crossline section in 2D and the complete 651 $\times$ 951 $\times$ 463 volume in 3D. With local inputs of 128 $\times$ 128 pixels or 128³ voxels and a propagation step of 100 samples, one horizon required 10 patches in 2D and 70 cubes (7 $\times$ 10) in 3D. On an NVIDIA RTX A4500 GPU, the measured end-to-end runtimes were 0.9 s and 14.7 s, respectively, including RGT inference, propagation, and fusion (Table 7). In 3D, RGT inference accounted for approximately 10.8 s (73% of the total), while propagation and fusion required approximately 3.0 s and 0.9 s. The estimated peak VRAM was 1.2 GB in 2D and 5.5 GB in 3D.

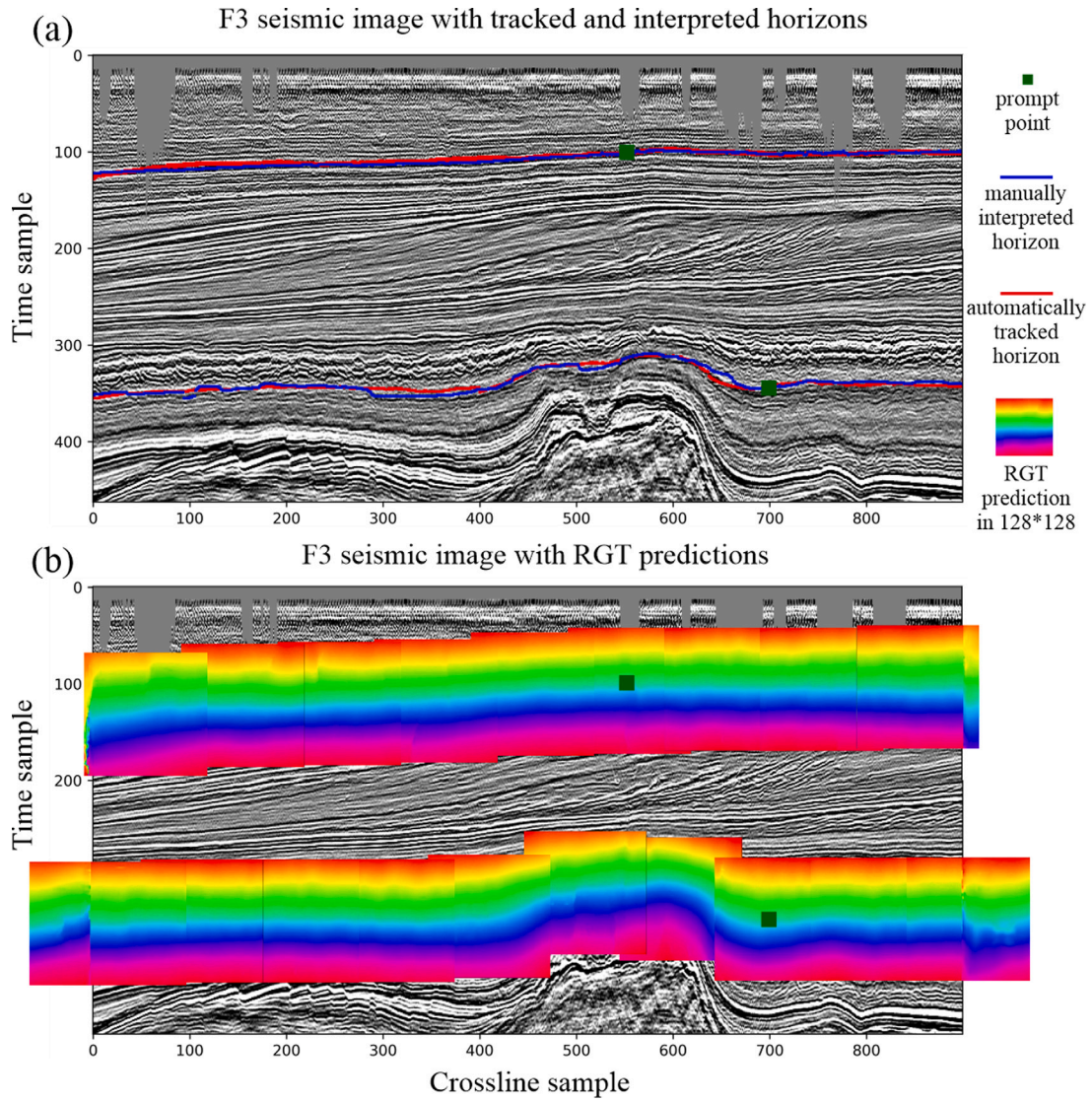


Fig. 11. Single-prompt RGT-guided horizon tracking results on a representative 2D crossline section from the F3 dataset. (a) Comparison between automatically tracked horizons (red) and manually interpreted horizons (blue) for a shallow and a deep target horizon. Green squares indicate the locations of the prompt points used to initiate horizon propagation. The tracked horizons closely follow the overall stratigraphic trends of the reference interpretations, including gradual dip changes and uplift-related deformation. (b) Corresponding RGT predictions generated in overlapping 128×128 windows along the section. The tracked horizons are located near the central RGT contours within each predicted RGT patch.

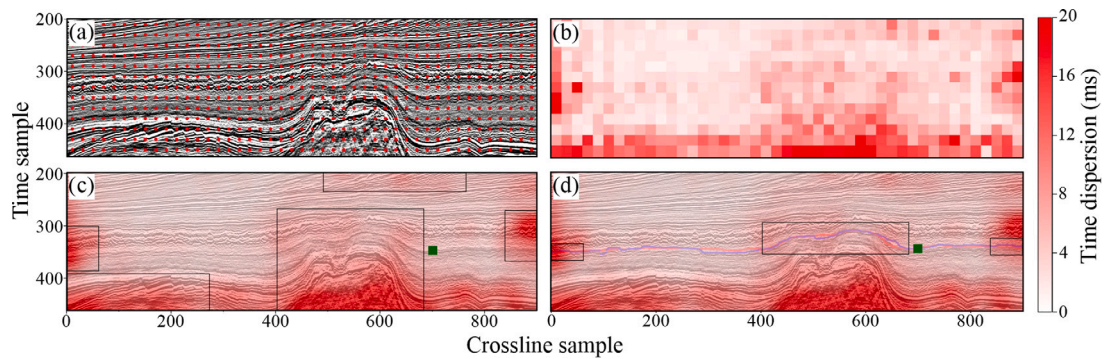


Fig. 12. Conditional confidence assessment based on prompt-induced propagation stability in the F3 dataset. (a) Representative 2D crossline section showing the grid of prompt points used for local prompt-sensitivity evaluation. (b) Prompt-sensitivity map where the color scale represents the standard deviation of the tracked horizon times generated from locally perturbed prompts, thereby mapping prompt-induced time dispersion. (c) Smoothed prompt-induced time-dispersion map overlaid on the seismic section, highlighting spatially coherent zones of elevated prompt sensitivity and reduced propagation stability. The boxed regions mark areas with consistently high dispersion values. (d) Final tracked horizons with boxed high prompt-induced time dispersion zones, indicating critical areas that require additional interpretive attention or multi-well constraints.

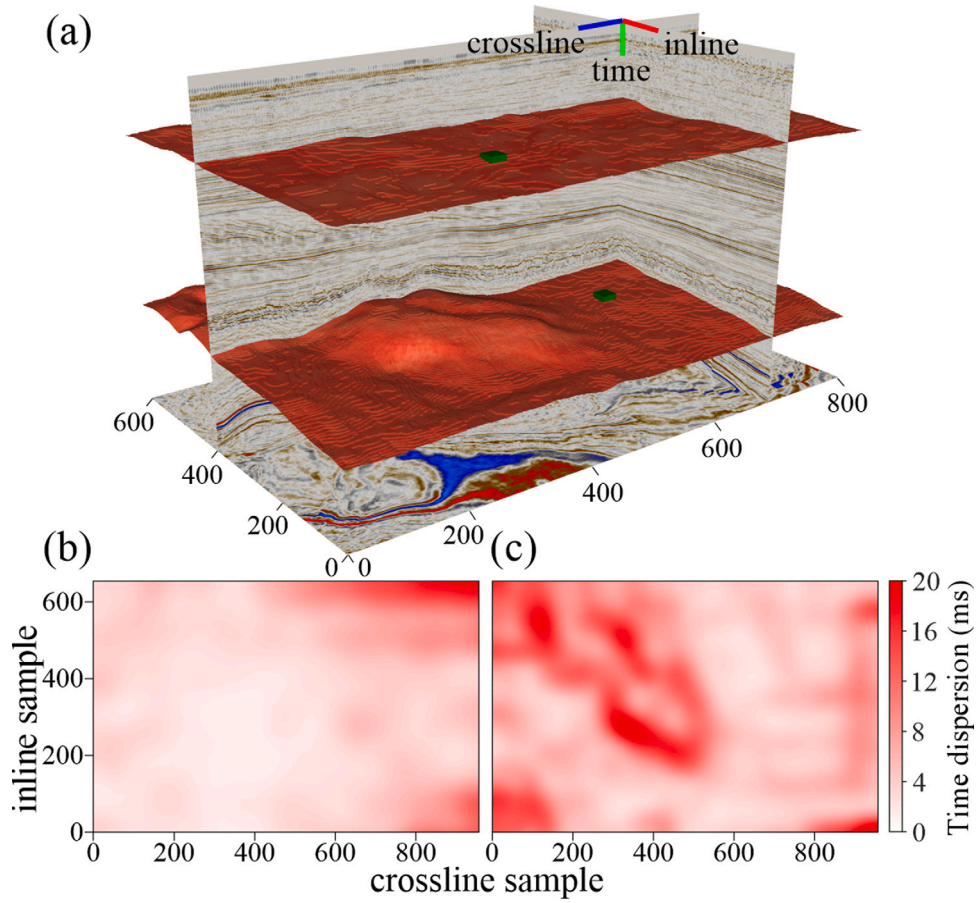


Fig. 13. 3D RGT-guided horizon tracking and conditional propagation-confidence assessment in the F3 dataset. (a) 3D visualization of single-prompt RGT-guided horizon tracking results for two target horizons. Red surfaces represent the tracked horizons, and green cubes denote the prompt points used to initiate horizon propagation. (b–c) prompt-induced time-dispersion maps for the (b) shallow at approximately 400 ms and (c) deep at approximately 1400 ms horizons based on the standard deviation of tracking results.

Table 7

Computational efficiency of single-prompt RGT-guided horizon tracking on the F3 data. Module-wise runtimes and peak VRAM are approximate estimates. Confidence-generation time is the conservative serial cost per evaluation location for 21 complete tracking realizations.

Computational item	2D workflow	3D workflow
Input size	951 × 463 samples	651 × 951 × 463 samples
Local input size	128 × 128 pixels	128 × 128 × 128 voxels
Number of local inputs	10 patches	70 cubes (7 × 10)
RGT inference	~0.5 s	~10.8 s
Horizon propagation	~0.3 s	~3.0 s
Overlap-based fusion	~0.1 s	~0.9 s
Single-prompt total	0.9 s	14.7 s
Confidence generation per location	18.9 s	308.7 s
Peak VRAM during inference	~1.2 GB	~5.5 GB

For a single-horizon 3D tracking task, the number of lateral cubes can be estimated as

$$N_{\text{cube}} = \left(\left\lceil \frac{N_x - P_x^{3D}}{\Delta x^{3D}} \right\rceil + 1 \right) \left(\left\lceil \frac{N_y - P_y^{3D}}{\Delta y^{3D}} \right\rceil + 1 \right),$$

where $N_x \times N_y$ is the lateral survey size. For the F3 volume, this gives $7 \times 10 = 70$ cubes. Under sequential processing, runtime scales approximately linearly with N_{cube} , whereas peak GPU memory is mainly controlled by one 128^3 cube. Because only one complete volume was benchmarked, this relation is a theoretical scaling estimate rather than an empirical multi-size comparison.

For n confidence-evaluation locations, each requiring $N_{\text{pert}} + 1 = 21$ realizations, the conservative serial upper-bound runtimes without cached RGT reuse are written separately as

$$T_{\text{conf}}^{2D} = 21nt_{2D}, \quad T_{\text{conf}}^{3D} = 21nt_{3D}.$$

With $t_{2D} = 0.9$ s and $t_{3D} = 14.7$ s, these expressions correspond to 18.9 s per evaluation location in 2D and 308.7 s (5.15 min) per evaluation location in 3D, or 18.9n s and 308.7n s for a complete map. These values represent conservative serial upper bounds obtained by repeating the complete workflow without caching the fixed RGT prediction. Because the realizations are independent and share the same

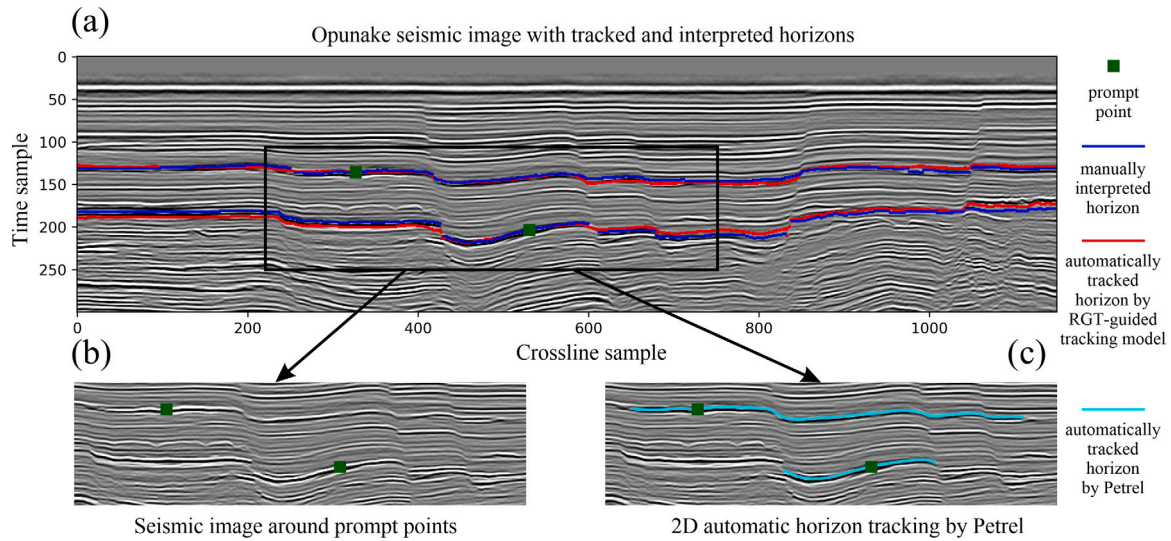


Fig. 14. Comparison of 2D RGT-guided spread tracking results and conventional amplitude-based method in the fault-dominated Opunake dataset. (a) Tracked horizons (red lines) initiated from user-defined prompt points (green squares) versus manual interpretation labels (blue lines). (b) Zoomed-in seismic image highlighting the complex normal faulting around the prompt points. (c) Results from Petrel's conventional 2D auto-tracking, illustrating termination at structural discontinuities.

RGT field, cached inference and parallel execution can substantially reduce the practical runtime.

4.2. Application on field data dominated by faults

Faults introduce significant stratigraphic displacements that challenge the continuity of automated tracking. To evaluate the spread tracking framework in such settings, we utilized Opunake seismic data from the Taranaki Basin, New Zealand. This region is characterized by extensional stress that has formed multiple normal faults with vertical displacements ranging from 50 to 100 m (Ma et al., 2023). The seismic data are provided in the time domain with a temporal sampling interval of 4 ms and a horizontal trace interval of 12.5 m in both inline and crossline directions. Two target horizons at approximately 500 and 700 ms TWT were manually interpreted as reference labels (Fig. 14a). The time volume contains a central fault with an approximately 80 ms TWT offset, flanked by three smaller faults with offsets of approximately 50–70 ms TWT (Fig. 14b).

From the single-prompt spread tracking results, the shallow horizon demonstrates a high degree of correlation with manual labels, with 97% of tracked points using the same 10-sample threshold. For the deeper horizon, the framework maintains an accuracy of 86%. Although the deep results exhibit localized fluctuations of 40 ms, they consistently honor the overall stratigraphic trend. These minor deviations likely stem from increased fault offsets at depth, which introduce greater complexity into the local RGT field. To further validate the framework, we conducted a comparative analysis with Petrel's conventional 2D auto-tracking (Fig. 14c). It terminates at fault planes and lateral discontinuities. In contrast, the RGT-guided spread tracking framework demonstrates superior continuity and consistency, effectively bridging fault-related gaps by leveraging the broader stratigraphic context provided by the RGT field.

3D RGT-guided spread tracking demonstrates robust performance in a fault-dominated dataset, revealing a coherent, fault-controlled stratigraphic geometry (Fig. 15a). The extracted horizon surfaces delineate multiple normal faults, clearly capturing their locations, dip directions, and depth-dependent variations in offset while maintaining lateral coherence across fault zones. At both shallow and deep surfaces, two prominent faults with significant displacement are identified in the central part of the study area. The right-hand fault exhibits a larger offset at depth that progressively decreases toward shallower

levels, whereas the left-hand fault displays a pronounced lateral transition from faulted to unfaulted geometry at the deeper level. The prompt-induced time-dispersion map for the deeper horizon exhibits pronounced spatial variability. Elevated dispersion values are sharply concentrated along fault traces and within structurally disturbed regions. In contrast, regions away from fault zones display consistently lower values, indicating reduced sensitivity to prompt initialization, higher geometric stability, and higher conditional tracking confidence (Fig. 15b).

4.3. Results for multi-prompt tracking

Horizon tracking based on a single prompt inherently propagates outward, making the results increasingly susceptible to local stratigraphic ambiguity and distance-related error accumulation. To mitigate this, our framework incorporates multiple prompt points to suppress long-range drift.

In the fault-dominated Opunake dataset, the 2D horizon tracked from a single central prompt accurately captures local trends near the prompt but shows deviations of up to 15 samples, equivalent to approximately 60 ms, at distances greater than 500 m from the prompt (Fig. 16a). By introducing additional prompts, the tracking is locally corrected in their respective vicinities (Fig. 16b). With three spatially distributed prompts, the final horizon geometry emerges from a distance-dependent weighted fusion, passing precisely through all prompts while minimizing deviations across the entire section (Fig. 16c).

We further evaluate 3D multi-prompt spread tracking on the F3 dataset using horizon constraints derived from real well ties as stratigraphic prompts. The target horizon is the deeper horizon at approximately 1400 ms, which is significantly deformed by uplift structures. In the single-prompt configuration using an isolated well (Fig. 17a), the tracked horizon honors the reference interpretation near the well but exhibits progressive deviations with distance, locally exceeding 20 samples in structurally complex uplift zones (Fig. 17d). Quantitative analysis reveals that only 68% of the surface remains within a 10-sample deviation (Fig. 17g).

Adding a second prompt point from another well expands the constraint coverage and substantially reduces large deviations in previously unconstrained areas (Fig. 17b,e), increasing the accuracy to approximately 82% (Fig. 17h). To further refine the result, a third

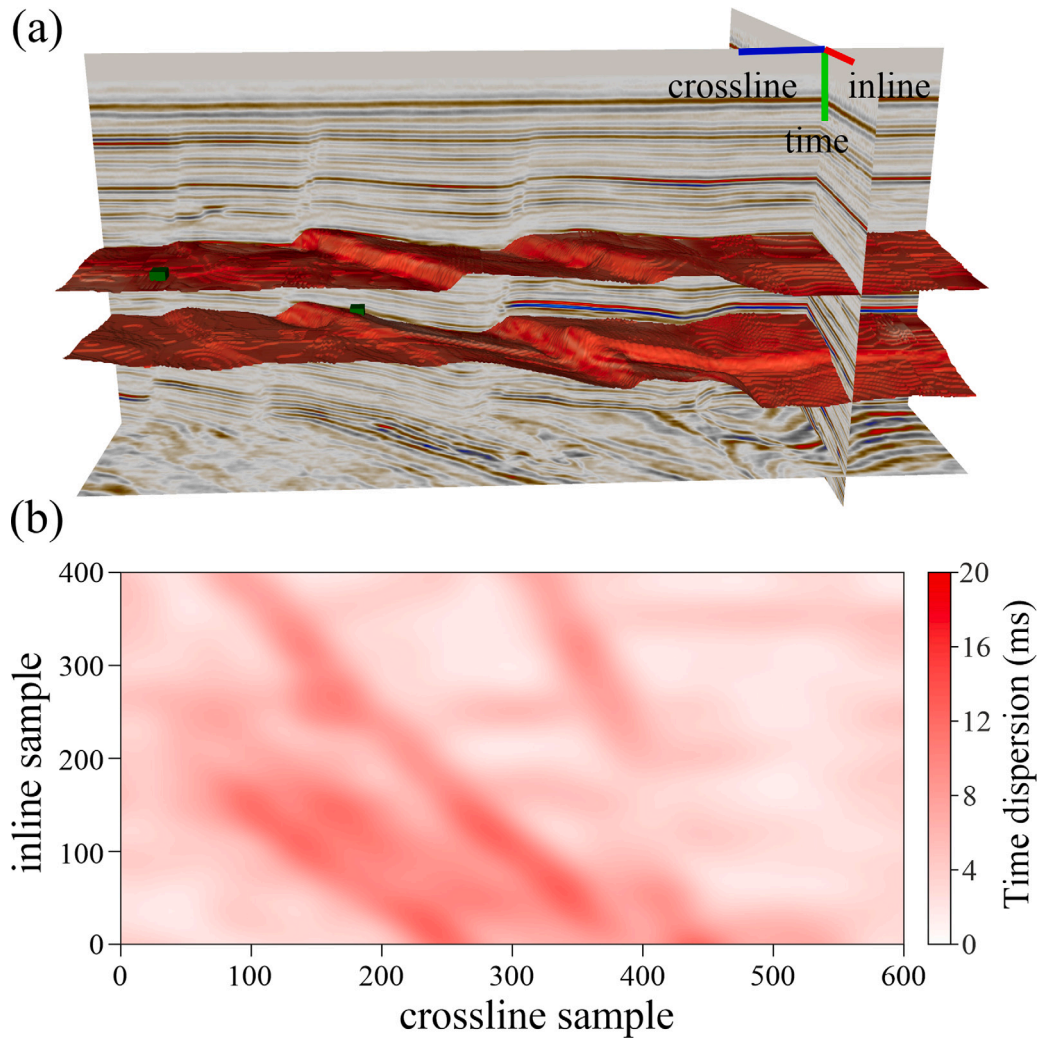


Fig. 15. 3D horizon tracking and conditional propagation-confidence assessment in the fault-dominated Opunake dataset. (a) 3D horizon surfaces capturing complex fault-controlled geometry. The model delineates fault locations and dip directions while maintaining lateral coherence across fault zones. (b) prompt-induced time-dispersion map for the deeper horizon. High prompt-induced time dispersion strips are sharply concentrated along fault traces, indicating lower conditional confidence and increased sensitivity of horizon propagation to prompt initialization, whereas areas away from faults show more stable, higher-confidence tracking.

prompt was strategically placed in a region of high prompt sensitivity identified by the conditional propagation-stability diagnostic, serving as a targeted corrective constraint (Fig. 17c,f). With this additional anchor, the residual misfit is suppressed, and the overall accuracy improves to 89% (Fig. 17i). These results demonstrate that the framework benefits not only from high-resolution well data but also from human-guided prompts.

5. Discussion

5.1. Methodological implications and practical utility

The main advantage of the proposed framework lies not only in improved horizon-tracking accuracy, but also in its use of a geologically meaningful constraint for automatic interpretation. By using the predicted RGT field as a stratigraphic guide, tracking is controlled by stratigraphic consistency rather than only by local amplitude continuity. This is important in structurally complex areas, where seismic reflections may be discontinuous or ambiguous (Lin et al., 2024). The field examples indicate that RGT-guided propagation can better maintain plausible horizon geometries across faulted, uplifted, and weak-reflection zones.

The methodological contribution lies in integrating RGT prediction with controllable local propagation rather than in using RGT alone. Previous RGT-focused methods have primarily reconstructed a global time field and directly extracted RGT isolines or isosurfaces (Geng et al., 2020; Bi et al., 2021). Here, RGT instead serves as an intermediate stratigraphic coordinate for prompt-guided spreading within overlapping patches or volumes. Overlap-based fusion and positioning the tracked horizon near the center of each RGT block reduce boundary effects, while multi-prompt correction limits long-range drift. This retains the interpreter control of seed-based tracking while reducing dependence on local amplitude similarity.

The controlled model comparison demonstrates that RGT-prediction quality and downstream tracking quality are related but not equivalent. The method of Geng et al. (2020) and U-Net-L2 achieve competitive or superior pixel-wise metrics, yet produce lower Acc@10 values and larger maximum deviations under the common tracking protocol. Conversely, the proposed predictor achieves the lowest mean LPIPS, GMSD, RGT-gradient MAE, dip MAE, and fault-zone jump MAE, together with the best field tracking performance among the evaluated 2D predictors. This suggests that local structural preservation and contour usability are more relevant to the final interpretation than average RGT-value error alone.

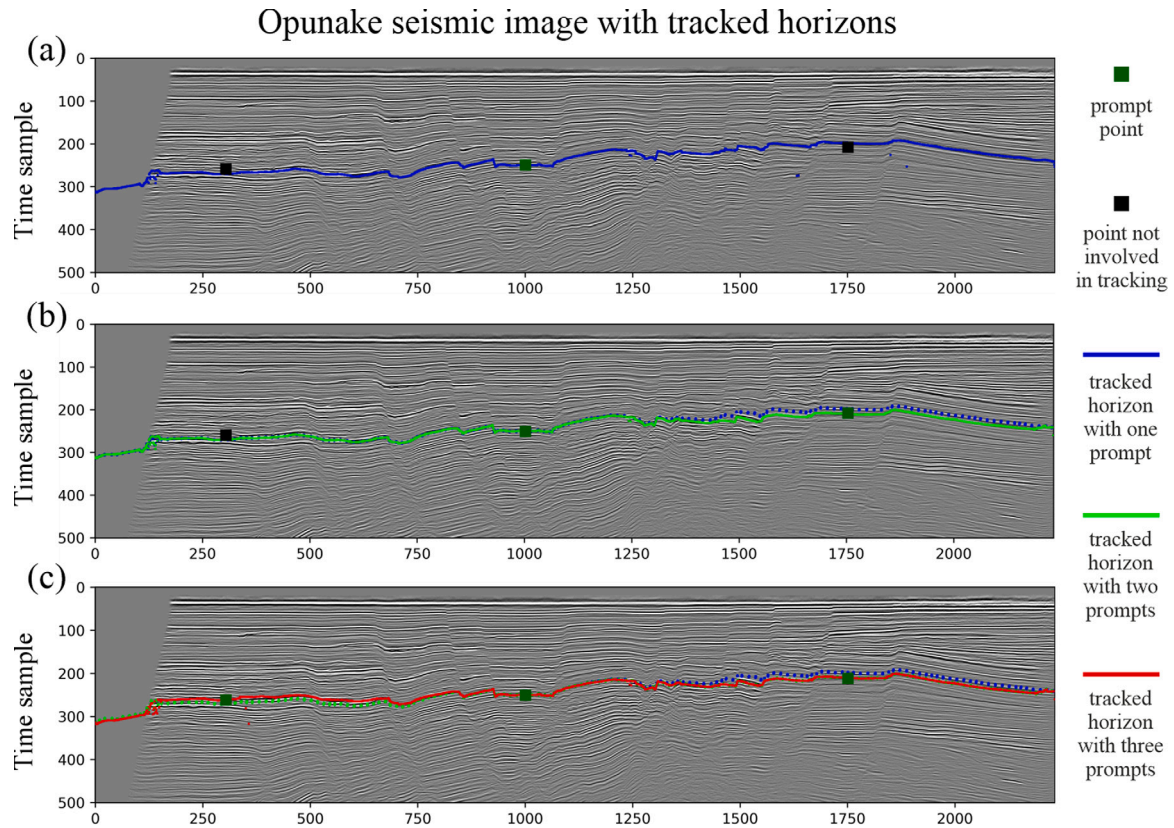


Fig. 16. 2D multi-prompt RGT-guided horizon tracking in the Opunake dataset. (a) Single-prompt tracking result initiated from the central prompt point. Black squares indicate interpreted horizon control points. (b) Two-prompt tracking result using the central and right control points, showing local correction relative to the single-prompt result. (c) Three-prompt integrated tracking result obtained by distance-weighted fusion, reducing deviations near both lateral control points and improving consistency relative to the single-prompt result.

The gradient-loss ablation shows that explicit derivative matching does not automatically improve geological fidelity. One plausible explanation is that a uniform gradient penalty treats smooth stratigraphic variation, fault discontinuities, and high-frequency fluctuations in the same manner, potentially competing with patch-scale adversarial regularization. Consistent with this interpretation, Pix2Pix-Grad rounds the abrupt normal-fault offset in Opunake (Fig. 6g), whereas the original Pix2Pix predictor retains a sharper break (Fig. 6a). Even at the best-performing tested weight ($\lambda_g = 1$), the added constraint slightly degrades the RGT-gradient, dip, fault-zone jump, and horizon-tracking metrics. The adversarial-plus- ℓ_1 objective is therefore retained as the best-performing balance under the present experimental configuration.

The prompt-based design improves practical controllability by using spatially distributed anchors to correct solutions propagated from an isolated seed. In the F3 test, increasing the number of prompts from one to three improved the proportion of the tracked 3D horizon within a 10-sample error range from 68% to 89%. This demonstrates the potential to combine automated propagation with sparse well ties or high-confidence interpreter input. The strategy therefore complements constrained seed selection and well-guided tracking (Pu et al., 2024; D'Angelo et al., 2024), while allowing additional prompts to correct locally inconsistent horizon segments.

The map should be interpreted as a prompt-sensitivity, or conditional propagation-stability, map. It measures the dispersion produced by perturbing prompt positions while holding the RGT field fixed. High values indicate sensitivity to prompt initialization, whereas low values indicate similar horizon geometries from neighboring prompts. The measure is therefore an inverse proxy for conditional tracking

confidence, not a probability of geological correctness. Unlike methods that encode predictive uncertainty in the learned model (Liao et al., 2024), it isolates propagation-stage sensitivity and is complementary to RGT predictive uncertainty. The map consequently supports expert-led quality control: after independent geological validation, low-sensitivity regions may require less intensive review, whereas prompt-sensitive regions require additional control or refinement.

The end-to-end runtimes of 0.9 s for 2D and 14.7 s for 3D tracking on the F3 data indicate interactive or near-interactive potential on the tested hardware. The stability assessment is more expensive because each location requires 21 tracking realizations and its serial cost grows approximately linearly with the number of evaluated locations. It is therefore better suited in its current form to targeted quality control than dense full-volume mapping. Because these realizations are independent and share a fixed RGT field, cached inference, parallel execution, and adaptive sampling provide direct routes to acceleration.

5.2. Limitations, confidence scope, and future perspectives

Although the proposed RGT- and prompt-constrained framework improves tracking stability and consistency in the tested examples, its reliability remains dependent on the quality of the predicted RGT field. Local RGT errors, including structural distortion, blurred fault boundaries, or incorrect stratigraphic ordering, may propagate into the tracked horizon, particularly in deeply deformed, fault-dense, weak-reflection, or poorly imaged regions. The method should therefore be regarded as a semi-automatic interpretation framework that supports geological quality control rather than as a fully automatic solution for all complex settings. Moreover, a stable result may still be incorrect if

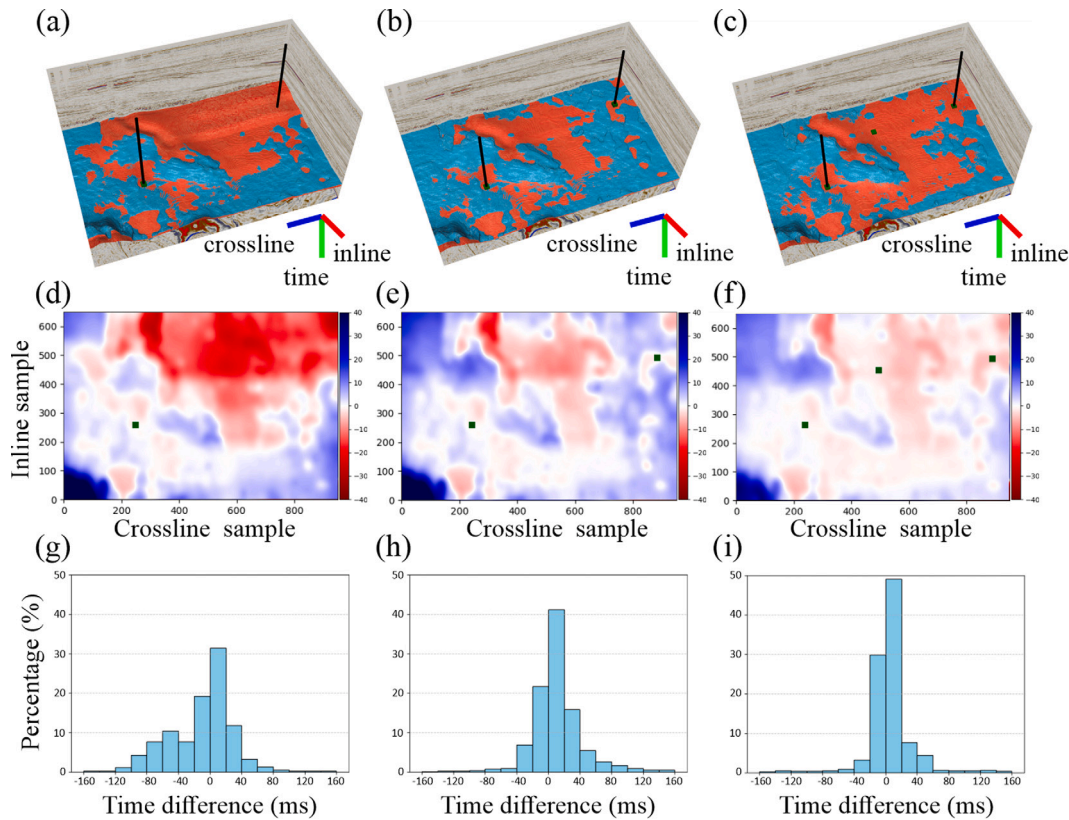


Fig. 17. 3D multi-prompt horizon tracking with well-based and targeted corrective prompts in the F3 dataset. (a–c) Tracked 3D horizon surfaces (red) using different prompt configurations compared with manual interpretation (blue): (a) a single well-based prompt, (b) two well-based prompts, and (c) two well-based prompts integrated with an additional targeted corrective prompt. Black vertical lines indicate well trajectories, and green squares denote the prompt point locations. (d–f) Corresponding error maps illustrating time differences between the tracked horizons and the reference manual interpretation. (g–i) Statistical histograms showing the distribution of time differences across the entire tracked surface for each configuration.

the underlying RGT prediction is systematically biased, whereas high prompt sensitivity may reflect genuine geological ambiguity rather than tracking failure.

Uncertainty or ambiguity associated with RGT prediction itself is not represented in the present prompt-sensitivity map. A more complete uncertainty framework could incorporate probabilistic RGT prediction through deep ensembles, dropout-based Bayesian approximations, heteroscedastic prediction, or other uncertainty-aware learning strategies. Because RGT predictive uncertainty and prompt-induced horizon dispersion are initially defined in different domains, RGT uncertainty should first be propagated through the tracking procedure into the final horizon-coordinate domain (time or depth) before the two components are calibrated and combined. Developing and validating such an integrated framework would require modified probabilistic RGT predictors, retraining, and dedicated calibration experiments and is therefore beyond the scope of the present study.

The gradient-loss ablation is also limited to a spatially uniform first-order derivative penalty evaluated over a restricted set of global weights ($\lambda_g \in \{1, 5, 10, 20\}$). Although $\lambda_g = 1$ performed best among the tested settings, these experiments do not exclude potential benefits from more geology-aware constraints. Future work could distinguish continuous bedding regions from fault boundaries, apply orientation-dependent or multiscale gradient terms, or constrain contour spacing in the local stratigraphic-normal direction. Fault-mask-aware objectives may further preserve discontinuities without imposing the same regularization behavior within continuous strata and across structural breaks.

The effectiveness of the multi-prompt strategy also depends on the availability and formulation of geological constraints. In practical

applications, well ties and high-confidence interpreted points may be spatially sparse and may not adequately sample complex stratigraphic relationships. Moreover, the current multi-prompt formulation uses planar distance as a practical proxy for prompt influence and does not explicitly account for geological connectivity. Although spatially distributed prompts provide useful local control over different segments of the same correlated horizon, geometric proximity does not necessarily imply stratigraphic continuity. In strongly faulted settings, a prompt in a neighboring structural block may influence the integrated estimate across a major discontinuity. Future extensions could therefore modulate distance-based weights using interpreted fault surfaces, fault-probability volumes, RGT discontinuities, or structure-constrained geodesic distances to reduce inappropriate cross-fault interaction.

A further limitation concerns the geological applicability of the spreading formulation. The current method assumes that the target horizon remains locally traceable within the predicted RGT field and that successive propagation windows belong to the same stratigraphic domain. This assumption may not hold near major unconformities, stratigraphic terminations, or salt–sediment interfaces. At an erosional unconformity, the target stratigraphic unit may have been partly or completely removed, and continuity-based propagation could extend the horizon into an interval where the corresponding surface no longer exists. Similarly, at onlap, downlap, toplap, or pinch-out terminations, the method may continue the horizon beyond its true geological endpoint because no explicit termination criterion is currently imposed. Salt-related settings present an additional challenge because sedimentary RGT continuity is not defined within salt, while poor illumination and disrupted reflections around salt boundaries may produce unreliable RGT estimates. These conditions could cause the tracked horizon

to enter a salt body, bridge across a boundary, or follow a spurious contour. Such geological settings were not explicitly evaluated in the present experiments and remain important targets for future validation.

The empirical validation also has a limited scope. The field evaluation uses examples from the F3, Opunake, and Kerry datasets and treats the available manual interpretations as reference horizons. These interpretations provide necessary operational benchmarks but do not quantify interpreter variability or the non-uniqueness of horizon placement in poorly imaged regions. The reported model comparisons also provide point estimates rather than variability across repeated training runs. In addition, the controlled published-baseline comparison is restricted to 2D RGT prediction; the 3D method of Bi et al. (2021) addresses a different joint horizon–fault formulation and was therefore not included in the common downstream tracking protocol. Consequently, the reported improvements should be interpreted as evidence within the tested datasets and configurations rather than as a complete measure of cross-survey generalization or superiority over all 3D tracking methods.

Future developments should prioritize three related directions. First, interpreted unconformity surfaces, stratigraphic termination maps, fault surfaces, and salt-body geometries could be incorporated as propagation barriers or validity masks. Termination-aware criteria could stop spreading when reflector convergence, loss of seismic support, or RGT discontinuity indicates that the target horizon is no longer present. Second, the synthetic training dataset could be expanded to include erosional truncation, depositional terminations, salt deformation, acquisition footprints, and imaging artifacts, together with field-data fine-tuning or domain adaptation to reduce the synthetic-to-real gap. Third, broader validation across surveys, repeated model training, multiple interpreter references, and controlled 3D benchmarking would provide a more complete assessment of robustness, generalization, and practical scalability.

6. Conclusion

This study presents a semi-automatic framework for 2D and 3D seismic horizon tracking that integrates adversarially regularized RGT prediction, prompt-guided local spreading, multi-prompt constraints, and perturbation-based assessment of propagation stability. By using RGT as an intermediate stratigraphic coordinate and propagating horizons through overlapping patches or volumes, the framework reduces its dependence on local seismic-amplitude continuity while limiting boundary artifacts and long-range drift. Controlled comparisons showed that pixel-wise RGT accuracy alone does not determine downstream tracking reliability. The selected Pix2Pix predictor achieved an Acc@10 of 91%, with a mean maximum absolute vertical error of 13 samples, under the common field-tracking protocol. The gradient-loss ablation further showed that a uniform first-order gradient penalty did not improve geological fidelity and tended to smooth abrupt fault offsets.

Applications to the F3 and Opunake datasets demonstrated coherent tracking across weak-reflection, uplifted, and fault-dominated regions. Multi-prompt constraints further reduced long-range deviation, increasing the proportion of the tracked 3D F3 surface within 10 samples of the reference interpretation from 68% with one prompt to 89% with three prompts. The perturbation-derived stability maps identified structurally disturbed regions where additional geological control may be required, while tracking times of approximately 0.9 s in 2D and 14.7 s in 3D indicate interactive or near-interactive potential. Importantly, this measure quantifies prompt-induced propagation stability under a fixed RGT prediction rather than complete predictive uncertainty. Overall, the framework provides a practical balance between automated propagation, sparse interpreter control, computational efficiency, and interpretable quality assessment.

CRedit authorship contribution statement

Wenhao Zheng: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Rebecca Bell:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Lluís Guasch:** Writing – review & editing, Supervision, Methodology. **Carlos Cueto:** Writing – review & editing, Supervision, Methodology. **Polina Pasalska:** Writing – review & editing, Data curation. **Yanghua Wang:** Writing – review & editing, Supervision, Funding acquisition.

Funding statement

This work was supported by the Resource Geophysics Academy (RGA) at Imperial College London. The funding source had no role in the study design, data analysis, interpretation of results, or preparation of the manuscript.

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.

Data and code availability statement

The F3, Opunake, and Kerry-3D seismic datasets used in this study are publicly available at <https://dgbearthsciences.sharefile.com/share/view/sfa1c428d20e84982928fb1aa763ab76e>, https://s3.amazonaws.com/open.source.geoscience/open_data/newzealand/Taranaiki_Basin/OPUNAKE-3D/OPUNAKE3D-PR3461-FS.3D.Final_Stack.segy, and https://s3.amazonaws.com/open.source.geoscience/open_data/newzealand/Taranaiki_Basin/Keri_3D/Kerry3D.segy, respectively. The processed data, trained models, and source code generated in this study are available from the corresponding author upon reasonable request.

References

- Bi, Z., Wu, X., Geng, Z., Li, H., 2021. Deep relative geologic time: A deep learning method for simultaneously interpreting 3-D seismic horizons and faults. *J. Geophys. Res.: Solid Earth* 126 (9), e2021JB021882.
- Bond, C.E., 2015. Uncertainty in structural interpretation: Lessons to be learnt. *J. Struct. Geol.* 74, 185–200.
- Bond, C.E., Lunn, R.J., Shipton, Z.K., Lunn, A.D., 2012. What makes an expert effective at interpreting seismic images? *Geology* 40 (1), 75–78.
- Cira, C.-I., Manso-Callejo, M.-Á., Alcarria, R., Fernández Pareja, T., Bordel Sánchez, B., Serradilla, F., 2021. Generative learning for postprocessing semantic segmentation predictions: A lightweight conditional generative adversarial network based on Pix2pix to improve the extraction of road surface areas. *Land* 10 (1), 79.
- D'Angelo, J., Zhao, Z., Zhang, Y., Ashok, P., Chen, D., van Oort, E., 2024. Real-time automated geosteering interpretation combining log interpretation and 3D horizon tracking. *Geosciences* 14 (3), 71.
- Fang, W., Fu, L., Zhang, M., Li, Z., 2021. Seismic data interpolation based on U-net with texture loss. *Geophysics* 86 (1), V41–V54.
- Geng, Z., Wu, X., Shi, Y., Fomel, S., 2020. Deep learning for relative geologic time and seismic horizons. *Geophysics* 85 (4), WA87–WA100.
- Hale, D., 2013. Methods to compute fault images, extract fault surfaces, and estimate fault throws from 3D seismic images. *Geophysics* 78 (2), O33–O43.
- Jin, S., Chen, S.Q., Wei, J.X., Li, X.Y., 2017. Automatic seismic event tracking using a dynamic time warping algorithm. *J. Geophys. Eng.* 14 (5), 1138–1149.
- Kushwaha, P.K., Maurya, S.P., Rai, P., Singh, N.P., 2021. Estimation of subsurface rock properties from seismic inversion and geo-statistical methods over F3-block, Netherland. *Explor. Geophys.* 52 (3), 258–272.
- Liao, Z., Zhu, P., Zhang, H., Li, Z., Li, Z., Ali, M., 2024. A deep learning-based seismic horizon tracking method with uncertainty encoding and vertical constraint. *IEEE Trans. Geosci. Remote Sens.* 62, 1–13.
- Lin, L., Zhong, Z., Li, C., Gorman, A., Wei, H., Kuang, Y., Wen, S., Cai, Z., Hao, F., 2024. Machine learning for subsurface geological feature identification from seismic data: Methods, datasets, challenges, and opportunities. *Earth-Sci. Rev.* 257, 104887.
- Ma, X., Yao, G., Zhang, F., Wu, D., 2023. 3-D seismic fault detection using recurrent convolutional neural networks with compound loss. *IEEE Trans. Geosci. Remote Sens.* 61, 1–15.

- Marfurt, K.J., Alves, T.M., 2015. Pitfalls and limitations in seismic attribute interpretation of tectonic features. *Interpretation* 3 (1), SB5–SB15.
- Mosser, L., Dubrule, O., Blunt, M.J., 2020. Stochastic seismic waveform inversion using generative adversarial networks as a geological prior. *Math. Geosci.* 52 (1), 53–79.
- Novoselov, A., Balázs, P., Bokelmann, G., 2022. SEDENOSS: Separating and denoising seismic signals with dual-path recurrent neural network architecture. *J. Geophys. Res.: Solid Earth* 127 (3), e2021JB023183.
- Peters, B., Granek, J., Haber, E., 2019. Multiresolution neural networks for tracking seismic horizons from few training images. *Interpretation* 7 (3), SE201–SE213.
- Posamentier, H.W., Paumard, V., Lang, S.C., 2022. Principles of seismic stratigraphy and seismic geomorphology I: Extracting geologic insights from seismic data. *Earth-Sci. Rev.* 228, 103963.
- Pu, Y., Zhang, B., Wan, Z., Yu, W., Cao, D., 2024. Generating more constrained seeds for seismic horizon extraction. *Geophysics* 89 (6), IM77–IM90.
- Ruppel, C.D., Waite, W.F., 2020. Timescales and processes of methane hydrate formation and breakdown, with application to geologic systems. *J. Geophys. Res.: Solid Earth* 125 (8), e2018JB016459.
- Sanda, O., Mabrouk, D., Tabod, T.C., Marcel, J., Essi, J.M.A., Ngos III, S., 2020. The integrated approach to seismic attributes of lithological characterization of reservoirs: case of the F3 block, North Sea-Dutch sector. *Open J. Earthq. Res.* 9 (3), 273–288.
- Shi, Y., Wu, X., Fomel, S., 2020. Waveform embedding: Automatic horizon picking with unsupervised deep learning. *Geophysics* 85 (4), WA67–WA76.
- Tschannen, V., Delescluse, M., Ettrich, N., Keuper, J., 2020. Extracting horizon surfaces from 3D seismic data using deep learning. *Geophysics* 85 (3), N17–N26.
- Waters, C.N., Zalasiewicz, J., Summerhayes, C., Barnosky, A.D., Poirier, C., Gałuszka, A., Cearreta, A., Edgeworth, M., Ellis, E.C., Ellis, M., Jeandel, C., Leinfelder, R., McNeill, J.R., Richter, D.d., Steffen, W., Syvitski, J., Vidas, D., Wagreich, M., Williams, M., An, Z., Grinevald, J., Odada, E., Oreskes, N., Wolfe, A.P., 2016. The Anthropocene is functionally and stratigraphically distinct from the Holocene. *Science* 351 (6269), aad2622.
- Wrona, T., Pan, I., Bell, R.E., Gawthorpe, R.L., Fossen, H., Brune, S., 2021. 3D seismic interpretation with deep learning: A brief introduction. *Lead. Edge* 40 (7), 524–532.
- Wrona, T., Pan, I., Gawthorpe, R.L., Fossen, H., 2018. Seismic facies analysis using machine learning. *Geophysics* 83 (5), O83–O95.
- Wu, X., Fomel, S., 2018. Least-squares horizons with local slopes and multigrid correlations. *Geophysics* 83 (4), IM29–IM40.
- Wu, X., Geng, Z., Shi, Y., Pham, N., Fomel, S., Caumon, G., 2020. Building realistic structure models to train convolutional neural networks for seismic structural interpretation. *Geophysics* 85 (4), WA27–WA39.
- Wu, X., Hale, D., 2015. Horizon volumes with interpreted constraints. *Geophysics* 80 (2), IM21–IM33.
- Wu, X., Zhong, G., 2012. Generating a relative geologic time volume by 3D graph-cut phase unwrapping method with horizon and unconformity constraints. *Geophysics* 77 (4), O21–O34.
- Xue, W., Zhang, L., Mou, X., Bovik, A.C., 2014. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Trans. Image Process.* 23 (2), 684–695.
- Yan, S., Wu, X., 2021. Seismic horizon extraction with dynamic programming. *Geophysics* 86 (2), IM51–IM62.
- Yuan, T., Spence, G.D., Hyndman, R.D., Minshull, T.A., Singh, S.C., 1999. Seismic velocity studies of a gas hydrate bottom-simulating reflector on the northern Cascadia continental margin: Amplitude modeling and full waveform inversion. *J. Geophys. Res.: Solid Earth* 104 (B1), 1179–1191.
- Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O., 2018. The unreasonable effectiveness of deep features as a perceptual metric. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 586–595.
- Zhao, Z., Zhao, J.-g., 2025. Seismic horizon tracking based on the TransUnet model. *Geophysics* 90 (2), IM1–IM13.