ACCSCIENCE PUBLISHING

Journal of Seismic Exploration

ARTICLE

Enhancing small-scale target inversion in acoustic full-waveform inversion using an embedded-scattering-point staining algorithm

Jiajun Song[1,2], Yingming Qu[1,2]*, Zhenchun Li[1,2]*, and Chunbo Ma[1,2]

[1]State Key Laboratory of Deep Oil and Gas, China University of Petroleum (East China), Qingdao, Shandong, China

[2]School of Geosciences, China University of Petroleum (East China), Qingdao, Shandong, China

Abstract

Full-waveform inversion (FWI) recovers high-resolution velocity models by iteratively minimizing the misfit between observed and modeled data, but its gradient is computed in the same way as reverse-time migration (RTM) and therefore inherits RTM's sensitivity to illumination. We introduce a staining-enhanced acoustic FWI scheme that isolates the gradient contribution of a marked small-scale target using the embedded-scattering-point staining algorithm: within the marked region, the gradient correlation is computed with a staining wavefield excited only by energy that has interacted with the target, rather than the full wavefield, yielding a locally modified search direction without altering the gradient elsewhere. The gradient and step length are computed via the adjoint-state method and a three-point parabolic line search within a conjugate-gradient-based FWI framework, with the staining correction activated once conventional inversion has reduced the relative misfit below a preset threshold. On a simple three-layer model, staining-enhanced FWI recovers a small square anomaly with markedly higher structural similarity to the true model, as quantified by the structural similarity index, than conventional FWI, and removes a bow-tie-shaped residual-diffraction artifact that otherwise surrounds the anomaly. On a sub-model extracted and modified from the geologically realistic Sigsbee salt-body benchmark, staining-enhanced FWI recovers marked scatterer targets with sharp velocity spikes that conventional FWI fails to resolve, while both methods recover the surrounding structure comparably. We further show that the method tolerates a moderate mislocation of the marked target—up to a capture radius comparable to the target's own spatial extent—beyond which the stained gradient instead focuses spuriously at the marked (incorrect) location, clarifying the method's practical requirements and failure mode. Additional tests with a masked-conventional baseline, a null test without the anomaly, mislocated target regions, and noisy data (signal-to-noise ratio = 5 dB) are used to assess the role of the target mask and the method's robustness.

Keywords: Full-waveform inversion; Staining algorithm; Embedded scattering point; Small-scale target; Adjoint-state method; Gradient enhancement

***Corresponding authors:**
Yingming Qu
(quyingming@upc.edu.cn);
Zhenchun Li
(leonli@upc.edu.cn)

Citation: Song J, Qu Y, Li Z, Ma C. Enhancing small-scale target inversion in acoustic full-waveform inversion using an embedded-scattering-point staining algorithm. *J Seismic Explor.* doi: 10.36922/JSE026350167

Received: August 30, 2026

Revised: October 11, 2026

Accepted: October 12, 2026

Published online: November 16, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Full-waveform inversion (FWI) exploits the full information content of seismic

waveforms—amplitude, phase, and travel time—to recover high-resolution subsurface velocity models.[1] FWI is formulated as a nonlinear least-squares problem, minimized iteratively; each iteration requires the gradient of the misfit function, which is obtained by correlating the forward-propagated wavefield from the actual source with the back-propagated residual wavefield—a computation structurally identical to reverse-time migration (RTM), with the residual data (rather than the raw data) as the migration input. Because of this structural equivalence, any subsurface region that RTM images poorly—due to limited acquisition aperture, complex overburden, unfavorable reflector dip, or simply small physical scale relative to the dominant wavelength—will also contribute a weak, noisy gradient in FWI, and the corresponding part of the model will converge slowly or fail to converge to the correct value.[2] FWI methodology itself has continued to mature rapidly: recent reviews summarize advances spanning acquisition, optimization, and multi-parameter inversion,[3,4] while specific lines of work have targeted the inversion's inherent nonlinearity and cycle-skipping through localized and adaptive misfit formulations,[5] reweighted variational objective,[6] dynamic data matching,[7] improved gradient preconditioning,[8] and deep-learning-based reparameterization,[9,10] and physics-constrained autoencoder architectures with axial self-attention.[11]

A large body of work addresses the non-uniform illumination problem in migration. Ray-based illumination and resolution analyses[12-14] describe amplitude and directional coverage but rely on high-frequency approximations that break down for diffraction- and scattering-dominated wavefields. Wave-equation-based illumination analyses, whether one-way[15,16] or full two-way[17-19] better capture such effects, and the point-spreading-function formalism[15,20,21] unifies illumination and resolution analysis by describing the local imaging operator as a spatially-varying blurring kernel that can, in principle, be deconvolved from the image to correct it. Small-scale and near-surface targets specifically have also begun to be addressed within a reflection-waveform-inversion framework more broadly[22] and, most directly relevant here, within FWI itself: a recent feasibility study shows that Love-shear-horizontal-wave FWI can improve detection of small-scale shallow anomalies relative to conventional seismic exploration methods,[23] underscoring that small-scale-target resolution remains an open problem even in current FWI practice.

A complementary line of work, the staining algorithm, was introduced by Chen and Jia,[24] inspired by fate-mapping techniques in developmental biology,[25,26] in which a lineage tracer marks a subset of embryonic cells so that their descendants can later be identified. By analogy, the staining algorithm marks a chosen target structure in the velocity model and excites an auxiliary ("stained") wavefield synchronized with the real wavefield but containing only the energy that has interacted with the marked structure. Migrating the corresponding "stained data" therefore images only the marked target at a higher signal-to-noise ratio (SNR) than a conventional image of the same region, because, by construction, it excludes unrelated reflections that would otherwise contaminate the image. Chen and Jia[24] originally implemented staining through a complex-domain-velocity (CDV) formulation, in which the wavefield and velocity are extended into complex space, with the imaginary part of the velocity assigned only within the target region; the real part of the resulting complex wavefield is the conventional field and the imaginary part is the stained field. This is elegant but roughly quadruples the computational cost relative to a standard, real-valued simulation, since every field variable must be propagated in complex arithmetic. Li and Jia[27] generalized the staining formulation to guarantee amplitude preservation between the stained and real wavefields, an essential property for using staining quantitatively rather than only for qualitative target identification, and applied it to target-oriented migration. Chen *et al.*[28] subsequently introduced the embedded-scattering-point (ESP) staining method, in which the target region is instead marked by embedding a small, spatially localized velocity perturbation (scatterers) directly into the real-valued velocity model; taking the difference between the wavefields simulated with the perturbed and unperturbed models isolates the staining wavefield at roughly twice the cost of one conventional simulation—half the cost of the CDV approach—while producing staining data that are kinematically consistent with those of the CDV formulation. Chen *et al.*[28] used the ESP staining wavefield to compute point-spreading functions and broadband acquisition-dip responses for illumination and resolution analysis, and to deconvolve and correct RTM images, including in poorly illuminated subsalt zones. The staining formulation itself has continued to be refined and extended. Li and Jia[29] further generalized the amplitude-preserving staining formulation, Li and Jia[30] extended a high-resolution staining formulation to one-way wave-equation modeling and imaging, and—in work specifically concerned with small-scale near-surface heterogeneities—Chen *et al.*[31] used a staining-based resolution-analysis method to investigate how shallow scattering from small-scale, near-surface bodies degrades seismic imaging, closely paralleling the small-scale-target motivation of the present study.

Building on the observation that a better local migration image implies a better local FWI gradient,

Hu *et al.*[2,32] combined the CDV staining algorithm with acoustic FWI: within a marked target zone the gradient correlation is computed using the stained wavefield, while elsewhere the conventional wavefield is retained, giving a composite gradient that is amplitude-consistent with the unstained gradient but has higher local SNR. Tests on the BP and Marmousi models showed that staining measurably improves the local gradient in poorly illuminated areas (e.g., below the BP model's overhanging salt flank), while in adequately illuminated regions of the Marmousi model, the staining and non-staining gradients were nearly indistinguishable, indicating that the benefit of staining-enhanced FWI is concentrated precisely where illumination, not model complexity, is the limiting factor. This line of work has since seen only limited follow-up: Yu and Jia[33] extended the staining-FWI idea to the frequency domain to improve robustness to low-SNR data, and Liu *et al.*[34] embedded the staining algorithm within least-squares RTM—an iterative, inversion-type imaging framework closely related to FWI—while Tao *et al.*[35] extended CDV staining to viscoacoustic vertical transversely isotropic RTM, and this direction has continued with energy-constrained regional-staining formulations for RTM[36] and their extension to viscoacoustic RTM.[37] However, none of these studies specifically target small-scale structures within FWI gradient computation.

The present study extends this line of work along three directions. First, rather than the CDV formulation used in Hu *et al.*,[2,32] we adopt the ESP staining algorithm of Chen *et al.*,[28] which uses only real-valued simulations and is roughly twice as efficient, making staining computations within FWI more affordable. Second, while prior staining-FWI studies emphasized general poorly-illuminated zones (subsalt shadows, steeply dipping reflectors), we specifically target small-scale structures—bodies whose lateral or vertical extent is comparable to or smaller than the dominant seismic wavelength, or thin/isolated anomalies embedded in an otherwise well-resolved background—where the conventional gradient is weak not necessarily because of acquisition geometry but because the target's own reflectivity contribution is small relative to the surrounding structure and is easily overwritten during stacking-type gradient formation.[38] Third, we embed the staining correction directly within an adjoint-state gradient computation and pair it with a Polak–Ribière-plus conjugate-gradient search direction and a three-point parabolic (quadratic) line search for the step length, and we activate the staining correction only after a preliminary stage of conventional FWI has reduced the large-scale data misfit below a preset threshold—giving a complete, self-consistent, staged acoustic FWI workflow suited to small-target enhancement.[39] We demonstrate the approach on a simple three-layer model with a single embedded small-scale anomaly and on a sub-model extracted and modified from the geologically realistic Sigsbee salt-body benchmark, embedded with small scatterer targets.

The remainder of the paper is organized as follows. Section 2 reviews the acoustic FWI misfit function and its adjoint-state gradient, presents the ESP staining formulation, and describes how the two are combined to obtain a small-scale-target-enhanced gradient. It also details the conjugate-gradient search direction, the three-point parabolic step-length search, and the overall staged workflow. Section 3 presents numerical examples on the simple model and geologically realistic complex Sigsbee-Sub model. Section 4 discusses the results, including the sensitivity of the staining-fused gradient to mislocation of the marked target, and Section 5 provides the conclusion.

2. Method and analysis

2.1. Acoustic full-waveform inversion and embedded-scattering-point staining formulation

Following the generalized least-squares formulation[1], the data misfit function is given as **Equation 1**:

$$E(\boldsymbol{v}) = \frac{1}{2}(L(\boldsymbol{v}) - \boldsymbol{d}_{obs})^T \left(L(\boldsymbol{v}) - \boldsymbol{d}_{obs}\right) \tag{1}$$

where L is the (nonlinear) acoustic modeling operator, $\boldsymbol{v}$ is the velocity model, and $\boldsymbol{d}_{obs}$ is the observed data. Throughout this paper, the bold symbol **v** denotes the discretized velocity model (a vector), whereas v or $v(\mathrm{x})$ denotes the scalar velocity at position x. The model is updated iteratively as in **Equation 2**:

$$\boldsymbol{v}^{(k+1)} = \boldsymbol{v}^{(k)} - \alpha_k \boldsymbol{d}_k \tag{2}$$

where k is the iteration index, α_k is the step length, and $\boldsymbol{d}_k$ is a search direction derived from the gradient $\nabla_v E^{(k)}$ via a nonlinear conjugate-gradient recursion.

The gradient itself is obtained via the adjoint-state method applied to the constant-density acoustic wave equation governing the forward wavefield p (**Equation 3**):

$$\frac{1}{v^2(\boldsymbol{x})}\frac{\partial^2 p(t,\boldsymbol{x};\boldsymbol{x}_s)}{\partial t^2} - \Delta p(t,\boldsymbol{x};\boldsymbol{x}_s) = f(t,\boldsymbol{x}_s) \tag{3}$$

with p the pressure wavefield excited by the source f at $\boldsymbol{x}_s$. The adjoint wavefield λ is obtained by injecting the data residual at the receiver locations and propagating it backward in time with the same wave operator, and the FWI gradient is the zero-lag correlation of the second time derivative of the forward wavefield with λ, summed over shots (**Equation 4**):

$$g(\boldsymbol{v}) = \frac{\delta E(\boldsymbol{v})}{\delta \boldsymbol{v}} = \frac{2}{v^3} \sum_{x_s} \int_0^T \frac{\partial^2 p}{\partial t^2} \lambda dt. \quad (4)$$

Equation 4 is structurally identical to an RTM image formed by cross-correlating source and residual-data wavefields; consequently, any region imaged poorly by RTM—including small-scale bodies whose reflectivity is weak relative to surrounding structure—produces a correspondingly weak, low-SNR contribution to $g(\boldsymbol{v})$.

Following common practice, we approximated the diagonal of the Hessian of the misfit function by the source-side illumination—the accumulated forward-wavefield energy at each point (**Equation 5**):

$$I(\boldsymbol{x}) = \sum_{x_s} \int_0^T p(t, \boldsymbol{x}; \boldsymbol{x}_s)^2 dt \quad (5)$$

and used $I(\boldsymbol{x})$ to precondition the raw gradient before it enters the search-direction recursion (**Equation 6**):

$$\hat{g}(\boldsymbol{x}) = \frac{g(\boldsymbol{v}, \boldsymbol{x})}{I(\boldsymbol{x}) + \varepsilon} \quad (6)$$

with ε as a small stabilization constant. This illumination-normalized gradient $\hat{g}$ (rather than the raw g) is what is actually carried forward into the conjugate-gradient recursion and, for brevity, we continue to write it simply as $g(\boldsymbol{v})$ below except where the distinction matters.

To isolate the wavefield energy associated with a chosen target structure, we marked the target using the ESP staining method.[28] A small velocity perturbation is embedded only within the target region (**Equation 7**):

$$\delta v(\boldsymbol{x}) = \begin{cases} \eta v(\boldsymbol{x}), & \boldsymbol{x} \in \Omega_{target} \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

where Ω_{target} denotes the marked (small-scale target) region and η is a small perturbation coefficient (in practice, $|\eta| \leq 0.1$). Substituting $v + \delta v$ for the velocity in **Equation 3**, the squared-slowness coefficient can be expanded to first order in δv (**Equation 8**),

$$\frac{1}{(v+\delta v)^2} = \frac{1}{v^2}\left(1 + \frac{\delta v}{v}\right)^{-2} \approx \frac{1}{v^2}\left(1 - \frac{2\delta v}{v}\right) + O\left(\frac{\delta v^2}{v^2}\right) \quad (8)$$

Writing the total field as $p + \delta p$ and substituting **Equation 8** into the perturbed wave equation gives the governing equation for $p + \delta p$; subtracting the unperturbed **Equation 3** and discarding the $O(\delta v^2)$ term from **Equation 8**—justified because η, and hence δv, is small— isolates the equation governing the staining wavefield δp excited purely by the embedded target (**Equation 9**):

$$\frac{1}{v^2} \frac{\partial^2 \delta p}{\partial t^2} - \Delta \delta p = \frac{2\eta}{v^2} \frac{\partial^2 p}{\partial t^2} \quad (9)$$

Equation 9 has the same left-hand-side operator as **Equation 3**, so δp propagates according to the same physics as p; the right-hand side acts as a secondary source that is switched on only where the real wavefield p interacts with the marked region Ω_{target}. Consequently δp is the first-order (Born) response of the wavefield to the embedded perturbation and contains only energy that has interacted with Ω_{target}. It shares the kinematics of p; however, because its secondary source involves the second time derivative of the incident field and the scattered energy propagates further through the model, its waveform, phase, and amplitude are not identical to those of the corresponding part of the real wavefield. It should therefore be distinguished from the amplitude-preserving staining wavefields of the generalized staining formulations, which are designed to be a true subset of the real wavefield. This makes δp suitable for forming a locally enhanced, high-SNR correlation image or gradient in Ω_{target} that is kinematically consistent with the conventional wavefield; its amplitude relative to the conventional gradient is not preserved and is instead controlled by the normalization and weighting described in Section 2.2.

In practice, δp is obtained from two real-valued forward simulations—one with the background model v and one with the perturbed model $v + \delta v$—and their difference, at roughly twice the cost of a single conventional simulation, i.e., about half the cost of the CDV formulation[24] that requires propagating all field variables in complex arithmetic throughout the model.

2.2. Staining-enhanced gradient and optimization

Small-scale target bodies—thin beds, isolated lenses, or local anomalies comparable to or smaller than the dominant wavelength—typically contribute weak, easily-masked reflectivity to both the forward wavefield and the resulting migration/gradient image, even when overall acquisition illumination is adequate, simply because their reflection energy is small relative to the surrounding background structure. We propose to enhance the FWI gradient specifically over such targets by replacing the correlation field used in **Equation 4** with the ESP staining wavefield inside Ω_{target}, while retaining the conventional wavefield elsewhere (**Equations 10** and **11**):

$$q(\boldsymbol{x}, t) = \begin{cases} \delta p(\boldsymbol{x}, t), & \boldsymbol{x} \in \Omega_{target} \\ p(\boldsymbol{x}, t), & \boldsymbol{x} \notin \Omega_{target} \end{cases} \quad (10)$$

$$g_{stain}(\boldsymbol{v}) = \frac{2}{v^3} \sum_{x_s} \int_0^T \frac{\partial^2 q}{\partial t^2} \lambda dt \quad (11)$$

The residual wavefield λ is unchanged from the conventional adjoint computation, since the data residual itself does not depend on staining; only the source-side

correlation field is switched between p and δp according to spatial location. Because δp contains, by construction, only the energy relevant to Ω_{target} and excludes unrelated background reflections that would otherwise dilute the local correlation, g_{stain} has higher local SNR than $g(\boldsymbol{v})$ inside Ω_{target}; outside Ω_{target}, $g_{stain} = g(\boldsymbol{v})$, so the modification is strictly local and does not perturb the inversion elsewhere.

Equation 11 does not define the gradient of the misfit function in **Equation 1**. Physically, because the staining wavefield δp is the first-order (Born) response of the forward wavefield to the embedded perturbation of **Equation 7**, the local correlation in **Equation 11** measures how strongly energy scattered from the marked region is present in the current back-propagated data residual; it is, in effect, a residual-weighted image of the marked target formed only with target-scattered energy. Mathematically, differentiating the adjoint-state gradient of **Equation 4** along the perturbation δv shows that the term $\frac{2}{v^3}\sum\int\frac{\partial^2\delta p}{\partial t^2}\lambda dt$ is one component of the full Hessian–vector product Hδv, namely the component that arises from perturbing the source-side wavefield while the data residual is held fixed (Appendix A1). Three properties follow. First, the staining correlation carries second-order (curvature) information about E along the target perturbation rather than first-order descent information. Second, it is linear in the perturbation coefficient and therefore changes sign with the polarity of the embedded perturbation; this dependence is removed by defining the local term per unit perturbation (Appendix A2). Third, it vanishes when the adjoint wavefield, and hence the data residual, vanishes (Appendix A1); when the local term is frozen, this property is lost, and stationarity is then ensured by the descent safeguard (Appendix A3). For these reasons, the staining-fused quantity used below (**Equation 14**) is treated as a staining-modified search direction rather than a gradient, and descent of the original objective is enforced explicitly by checking each search direction against the true adjoint-state gradient.

As with other target-oriented imaging and inversion schemes, this formulation requires the target region Ω to be specified a priori. Because the staining wavefield of **Equation 9** is generated by treating the marked region itself as a genuine scattering perturbation, the resulting stained gradient tends to focus correlation energy toward the marked location when the marked region does not overlap the true target—a direct consequence of the underlying Born/RTM imaging principle (data generated by a scatterer focuses back onto that scatterer) rather than a limitation specific to this implementation. In practice, Ω can be obtained from conventional illumination or resolution analysis (identifying weakly illuminated or poorly resolved zones, as reviewed in Section 1), from geological or well-log priors, or from the output of a preliminary stage of conventional FWI/RTM. The sensitivity of the stained gradient to errors in the assumed target location is examined directly in Section 3.

Because Ω_{target} is small in this application (by definition—a small-scale body), the embedded-scattering-point density in **Equation 7** should be chosen fine enough to adequately sample the target's spatial extent while keeping η small enough that the linearization leading to **Equation 9** remains valid; this is a practical trade-off between staining-wavefield accuracy and computational/model-perturbation cost.

In the implementation used for the numerical examples below, the hard 0/1 switch of **Equation 10** is replaced by a smoothly tapered spatial weight $m(\boldsymbol{x}) \in [0,1]$ centered on Ω_{target}, obtained by convolving the indicator function of Ω_{target} with an isotropic Gaussian kernel of width σ_m (**Equations 12** and **13**):

$$m(\boldsymbol{x}) = \left(G_{\sigma_m} * \Omega_{target}\right)(\boldsymbol{x}) \tag{12}$$

$$G_{\sigma_m}(\boldsymbol{x}) = \frac{1}{2\pi\sigma_m^2} exp\left(-\frac{|\boldsymbol{x}|^2}{2\sigma_m^2}\right) \tag{13}$$

so that the staining-fused gradient actually used in the inversion is **Equation 14**:

$$\hat{g}_{\text{stain}}(\boldsymbol{x}) = \left[1-m(\boldsymbol{x})\right]g(\boldsymbol{v})(\boldsymbol{x}) + m(\boldsymbol{x})\hat{g}_{\delta p}(\boldsymbol{x}) \tag{14}$$

where $\hat{\boldsymbol{g}}_{\boldsymbol{\delta p}}$ is the staining-based local gradient, normalized by its own peak amplitude, $\hat{\boldsymbol{g}}_{\boldsymbol{\delta p}}(\boldsymbol{x}) = \boldsymbol{g}_{\boldsymbol{\delta p}}(\boldsymbol{x}) / \boldsymbol{max}_x\left|\boldsymbol{g}_{\boldsymbol{\delta p}}(\boldsymbol{x})\right|$, with $\boldsymbol{g}_{\boldsymbol{\delta p}}$ computed from **Equation 11** restricted to $\boldsymbol{\Omega}_{\boldsymbol{target}}$. Two further practical simplifications were adopted for computational efficiency: (i) $\hat{\boldsymbol{g}}_{\boldsymbol{\delta p}}$ was computed once, at the iteration at which staining is activated, from the current velocity model and the current data residual, and was kept fixed thereafter, whereas the conventional gradient, the data residual and the adjoint wavefield are recomputed at every iteration, and (ii) the fused gradient of **Equation 14** was only activated once the conventional inversion had already reduced the relative data misfit below a threshold (in our tests, 0.65 of the initial misfit), so that the early, large-scale iterations proceed as conventional FWI and staining is reserved for refining the small-scale target once the background velocity is reasonably converged. Both simplifications trade some theoretical generality (a fully consistent implementation would re-stain and re-correlate every iteration) for practicality; the effect of the first is quantified in Section 3.1, and its consequence for the stationary points of the iteration is analyzed in Appendix A3.

Because the staining wavefield, and hence the local correlation in **Equation 11**, is linear in the perturbation coefficient η to first order (Appendix A2), the local correlation is defined per unit perturbation, i.e., it is divided by η before normalization. The normalized local gradient is then independent of both the magnitude and the polarity of the embedded perturbation within the range where the linearization of **Equation 8** holds, so that η affects the result only through the accuracy of that linearization; for the positive η used in this study, this is identical to normalizing the raw correlation. The normalization by the maximum absolute value preserves the sign of the local correlation, which is set by the data residual through the adjoint wavefield rather than by the polarity of the perturbation, and the relative scale between the local term and the illumination-normalized conventional gradient is controlled by the weight in **Equation 14** rather than by the raw staining amplitude.

Given the (possibly staining-fused) gradient, the search direction $\boldsymbol{d}_k$ used in the model update of **Equation 2** is obtained via a Polak–Ribière-plus nonlinear conjugate-gradient recursion (**Equations 15** and **16**):

$$\beta_k = max\left(0, \frac{g_k^T\left(g_k - g_{k-1}\right)}{g_{k-1}^T g_{k-1}}\right) \tag{15}$$

$$\boldsymbol{d}_k = g_k + \beta_k \boldsymbol{d}_{k-1} \tag{16}$$

with β_k forced to zero (automatic restart to steepest descent) whenever it would otherwise be negative, whenever $\boldsymbol{d}_k$ fails a descent-direction check ($g_k^T \boldsymbol{d}_k \le 0$) evaluated with the true adjoint-state gradient of **Equation 4** rather than with the staining-fused vector, at the first iteration, and—importantly—at the iteration where the staining-fused gradient is switched on for the first time, since the abrupt change in gradient character at that point makes the previous conjugate direction a poor basis for extrapolation. If the staining-fused direction fails this check, it is replaced by the illumination-preconditioned conventional gradient of **Equation 6**, which is always a descent direction because its inner product with the true gradient is positive. As shown in Appendix A3, with this safeguard, every accepted update decreases E, and the iteration can become stationary only where the true gradient vanishes, even though the fused vector in **Equation 14** does not vanish there when the local term is frozen. In the three-layer example, the staining-fused direction passed the check at every iteration (its cosine with the true gradient ranged from 0.63 to 0.97), so the fallback was never triggered.

Given $\boldsymbol{d}_k$, the step length α_k is estimated using a three-point parabolic (quadratic) line search rather than a full line search, which would require many additional forward simulations per iteration. Three trial step lengths $\alpha_1 < \alpha_2 < \alpha_3$ are chosen, and the misfit function is evaluated at each, giving E_1, E_2, E_3; fitting a parabola through the three points and solving for its minimum gives the estimated optimal step length (**Equation 17**):

$$\alpha_k^* = \alpha_2 - \frac{1}{2}\frac{(\alpha_2-\alpha_1)^2\left(E_2-E_3\right)-(\alpha_2-\alpha_3)^2\left(E_2-E_1\right)}{(\alpha_2-\alpha_1)\left(E_2-E_3\right)-(\alpha_2-\alpha_3)\left(E_2-E_1\right)} \tag{17}$$

In practice, the three trial points are taken as $\alpha_1 = 0$ (the current model, $E_1 = E(\boldsymbol{v}^{(k)})$), $\alpha_2 = c_1 s_{k-1}$, and $\alpha_3 = c_2 s_{k-1}$, where s_{k-1} is the step length accepted at the previous iteration and $c_1 < c_2$ are fixed multipliers (we use $c_1 = 1.0$, $c_2 = 2.5$). If the parabola in **Equation 17** is concave-down, ill-conditioned, or yields a non-positive step, a fallback rule is used: if the misfit is still monotonically decreasing across the three trial points, the step is extrapolated further outward; otherwise the trial point with the lowest misfit among {0, α_2, α_3} is accepted directly. The scale s_{k-1} carried into the next iteration is updated from the accepted step with a floor to avoid the search collapsing to vanishing trial steps. A candidate step is accepted only if it decreases the misfit evaluated on the box-constrained model of **Equation 18**, i.e., $E(\mathrm{v_k} - \alpha\, \mathrm{d_k}) < E(\mathrm{v_k})$; otherwise, the smaller trial step is halved up to three times, and the inversion is terminated if no decrease is obtained. In our implementation, the parabolic estimate is used only if it does not exceed $4\alpha_3$, and outward extrapolation uses the step $2\alpha_3$.

Finally, the model update of **Equation 2** is applied with box constraints that keep the velocity within a physically admissible range and, where appropriate, exclude a known, undeformed shallow layer (e.g., a water column) from the update (**Equation 18**):

$$v^{(k+1)}(\boldsymbol{x}) = \Pi_{[v_{min}, v_{max}]}\left[v^{(k)}(\boldsymbol{x}) - \alpha_k^* M(\boldsymbol{x}) d_k(\boldsymbol{x})\right] \tag{18}$$

where $\Pi_{[v_{min}, v_{max}]}[\cdot]$ denotes element-wise clipping to $[v_{min}, v_{max}]$ and $M(\boldsymbol{x}) \in \{0, 1\}$ is a fixed mask equal to zero over the excluded shallow layer and one elsewhere.

2.3. Overall workflow

The proposed staining-enhanced acoustic FWI proceeds as a two-stage process, summarized in Figure 1. In the first stage, conventional FWI reduces the large-scale background misfit using the standard adjoint-state gradient. Once the relative data misfit falls below a preset threshold, the inversion transitions permanently to the second stage, in which the fused gradient of Section 2.2 is used for the remainder of the run. This one-time switch also triggers a conjugate-gradient restart, as noted in Section 2.2.

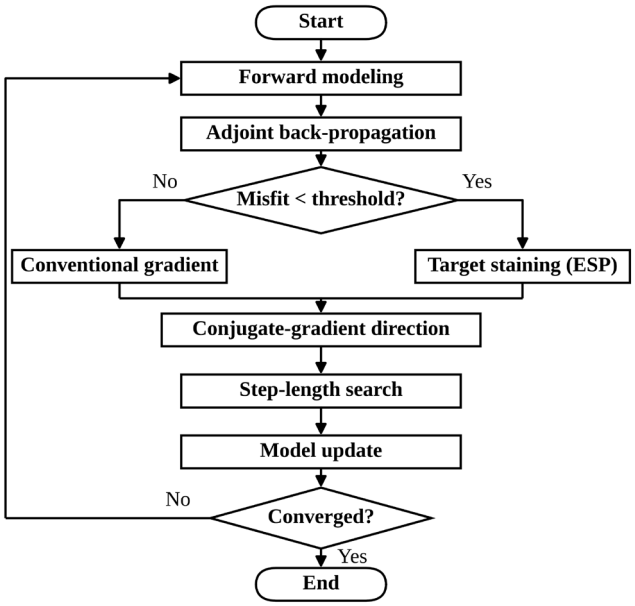


Figure 1. Workflow of the staining-enhanced acoustic full-waveform inversion (FWI) scheme. The inversion begins as conventional FWI; once the relative data misfit falls below the preset threshold (0.65 in our tests), the gradient is permanently switched to the embedded-scattering-point (ESP)-staining-fused form for the remainder of the run.

The two-stage strategy is deliberate and motivated by the distinct spectral sensitivities of conventional and staining-enhanced gradients. Conventional FWI is most effective at recovering long-wavelength, large-scale velocity structures in early iterations, when the data residual is dominated by phase and travel-time discrepancies associated with the background model. Activating staining too early—when the background model is still far from the true solution—is counterproductive for two reasons. First, the staining wavefield is derived from the current model; if the background velocity is substantially in error, the wavefield kinematics are incorrect, and the stained gradient may focus spuriously or amplify mislocated energy. Second, early iterations require global updates across the entire model; a target-localized gradient would under-update the background and could trap the inversion in a local minimum with a correct target but an incorrect surrounding structure. By deferring staining activation until the background has converged, we ensure that the staining wavefield propagates through a kinematically reasonable overburden and that the local enhancement operates on a residual misfit dominated by fine-scale structure rather than large-scale phase errors.

The activation threshold of 0.65 for the relative misfit is determined empirically from the behavior observed in our numerical tests. In the simple three-layer model, the relative misfit drops rapidly during the first few iterations as the layer boundaries and background velocities converge (the threshold is reached at iteration 5), after which the decrease slows markedly, indicating that the large-scale structure has been largely recovered and that further conventional iterations yield diminishing returns. In the more complex Sigsbee sub-model, the convergence is slower, but a similar plateau emerges once the salt body and surrounding sediments are approximately in place. A threshold of 0.65 captures this transition point in the three-layer model, and a threshold of 0.45 was used for the more slowly converging Sigsbee sub-model (Section 3.2); in both cases the threshold is low enough to ensure that the background is reasonably well resolved, yet high enough to leave sufficient iterations for the staining-enhanced stage to refine the small-scale targets. We do not claim 0.65 as a universal constant; for models with stronger nonlinearity or poorer initial models, a lower threshold may be prudent, whereas for well-constrained near-surface problems, staining could conceivably be activated earlier. An ablation study over the activation threshold would help quantify this trade-off and is left for future work.

A practical consideration concerns the refresh frequency of the staining gradient. In the implementation used for the numerical examples below, the staining-based local gradient is computed once, at the activation iteration, from the current velocity model and data residual, rather than re-deriving it at every FWI iteration. This approximation is justified because, once the background model has converged—which is precisely when staining

is activated—the spatial character of the wavefield near the target does not change enough in later iterations to materially alter the staining pattern. Recomputing the staining wavefield at every iteration would roughly double the per-iteration cost (one additional forward simulation with the perturbed model), and a direct comparison on the three-layer model shows that refreshing the local gradient every 10 iterations changes the recovered target only marginally (Section 3.1). However, if staining were activated from the first iteration—when the model evolves rapidly—this approximation would no longer hold, and the staining wavefield would need periodic or per-step refreshing.

The overall per-iteration cost of the staining-enhanced stage is therefore comparable to conventional FWI, plus a one-time overhead of two additional forward simulations for the ESP staining pass. Compared with the CDV staining formulation, which quadruples the computational cost by requiring complex arithmetic throughout, the ESP approach reduces staining overhead by about a factor of two. This cost efficiency is particularly important in three-dimensional applications, where even doubling the wavefield simulation cost is substantial, and it makes periodic refreshing of the staining gradient affordable for realistically sized problems.

Figure 1 illustrates the complete workflow. Aside from the one-time staining activation and the associated conjugate-gradient restart, the per-iteration procedure—forward modeling, adjoint back-propagation, gradient formation, search-direction and step-length estimation, and model update—follows directly from the equations above and is not repeated here.

3. Results

Two test cases are reported: (i) a simple layered model containing a single small localized velocity anomaly, used to demonstrate—in a controlled setting and with quantitative image-similarity metrics—that staining-enhanced FWI recovers a small-scale target more faithfully than conventional FWI; and (ii) a sub-model extracted and modified from the Sigsbee salt model, a complex geologically realistic test in which the same small-scale-target enhancement is evaluated within a much larger, structurally complicated background.

3.1. Three-layer model

A simple three-layer velocity model of 1 km × 1 km, containing a single isolated square anomaly embedded in the middle layer (Figure 2, true model), was employed as a controlled test for small-scale-target recovery. Both conventional FWI and staining-enhanced FWI commence from the same initial model (Figure 2) and identical acquisition geometry. The forward wavefields and shot records generated with and without staining are first compared (Figures 3 and 4), followed by an examination of the resulting gradients (Figure 5), the inverted velocity models (Figure 6), the convergence curves (Figure 7), and single-trace velocity profiles extracted through the anomaly (Figures 8 and 9).

The forward and adjoint simulations employ an acoustic staggered-grid finite-difference scheme on a 101 × 101 grid with 10 m spacing. Absorptive boundary conditions were applied to eliminate spurious reflections caused by artificial boundary truncation. A Ricker wavelet with a dominant frequency of 20 Hz serves as the source, sampled at 1 ms intervals over 601 time steps (0.6 s total record length). The acquisition geometry consists of 19 evenly distributed surface shots, with a fixed spread of 101 receivers covering the full model width. Direct-arrival energy was muted from both observed and modeled shot gathers prior to residual formation. Both inversions start from the same initial model and acquisition configuration and run for 50 iterations. The uppermost 100 m was excluded from the gradient update, and the staining correction was activated once the relative data misfit fell below 0.65 of its initial value.

Figure 6 shows the inverted velocity models. With the frozen local gradient (Figure 6b), staining-enhanced FWI reconstructs the small square anomaly compactly with well-defined edges, whereas conventional FWI (Figure 6a) recovers only a weak, smeared anomaly surrounded by an hourglass-shaped residual-diffraction artifact. The masked-conventional baseline (Figure 6c) uses the same mask, activation criterion, normalization, and weighting as the proposed method but replaces δp with p in the local term. It does not reproduce the improvement and instead drives the velocity at the target center below the background, indicating that the conventional gradient inside such a small target is dominated by side-lobe and cross-talk contributions. The improvement therefore originates from the staining wavefield rather than from re-emphasizing the update inside the target region.

Table 1 summarizes the comparison quantitatively. Staining-enhanced FWI yields the smallest velocity error and the highest local structural similarity index measure around the target, while the four models agree away from the target, confirming that the staining correction is local. Because the upper bound of the box constraint in **Equation 18** was 4,000 m/s, equal to the anomaly velocity, the agreement of the staining-enhanced results with the true value inside Ω partly reflects this bound.

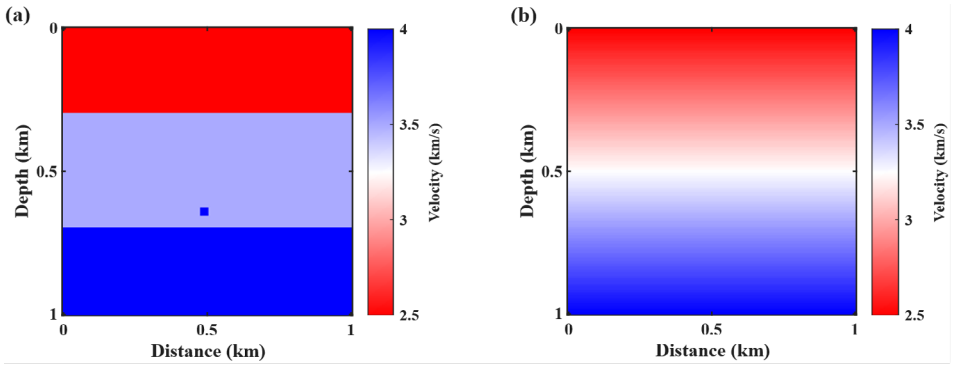


Figure 2. Simple three-layer test model. (a) True velocity model, containing a single small square anomaly embedded in the middle layer. (b) Initial (smoothed) velocity model used to start the full-waveform inversion iterations.

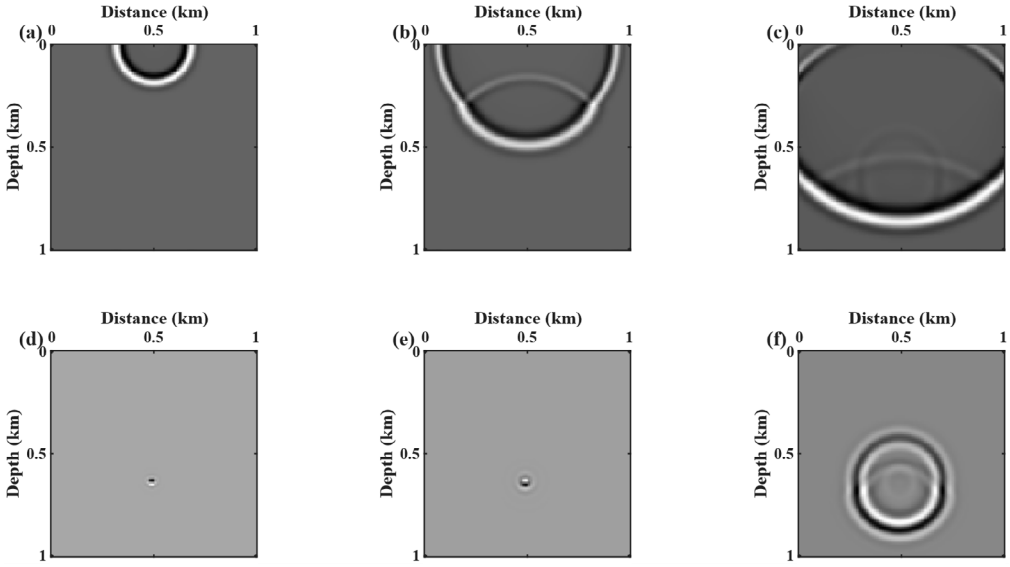


Figure 3. Comparison of the conventional and embedded-scattering-point-staining wavefield snapshots at three representative propagation times, showing that the staining wavefield is excited only after energy interacts with the marked target and is much weaker than the conventional wavefield

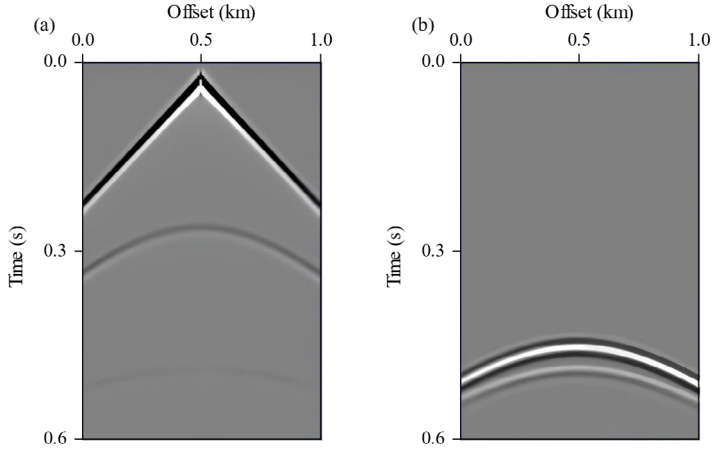


Figure 4. Comparison of conventional and embedded-scattering-point-staining shot gathers for the simple model, showing that the staining record isolates the target's diffraction response while excluding unrelated reflections

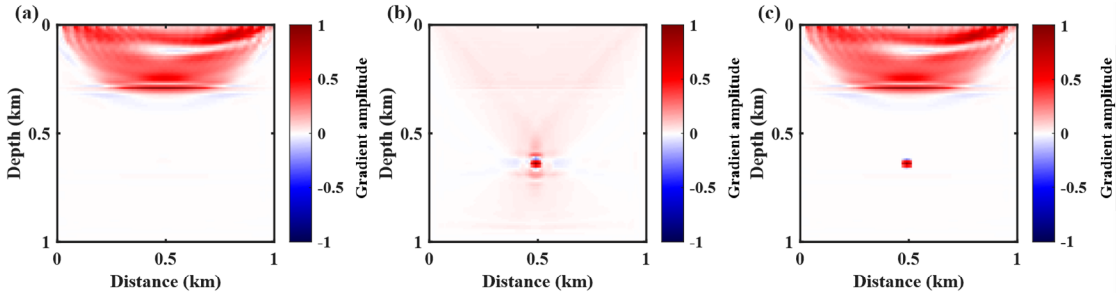


Figure 5. Comparison of the conventional and staining-fused gradients for the simple model. (a) Conventional gradient. (b) Staining-based local gradient. (c) Fused staining gradient.

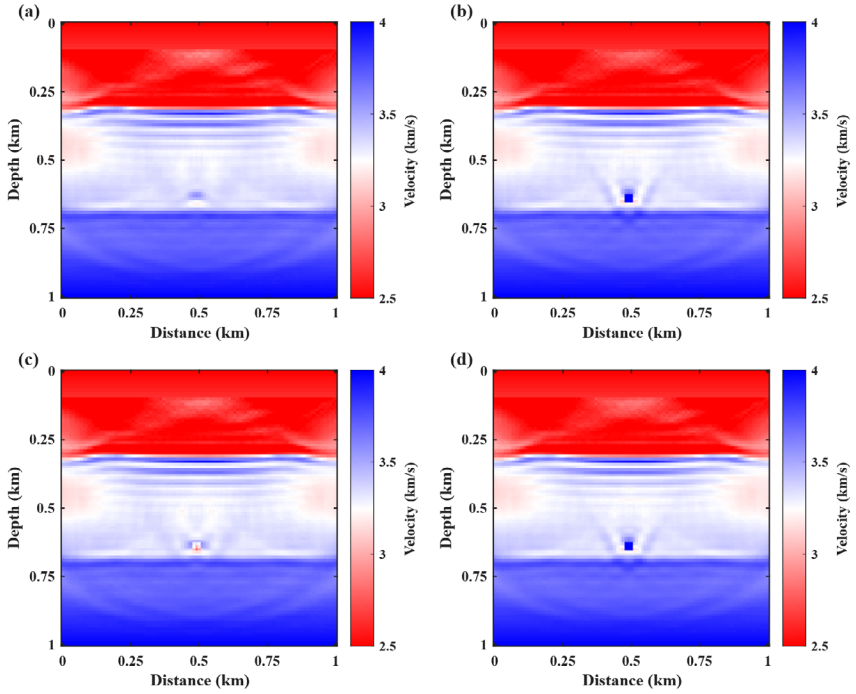


Figure 6. Inverted velocity models of the three-layer model after 50 iterations. (a) Conventional full-waveform inversion (FWI); (b) staining-enhanced FWI with the staining-based local gradient computed once at activation (frozen); (c) masked conventional FWI (same mask, no staining); (d) staining-enhanced FWI with the staining-based local gradient refreshed every 10 iterations.

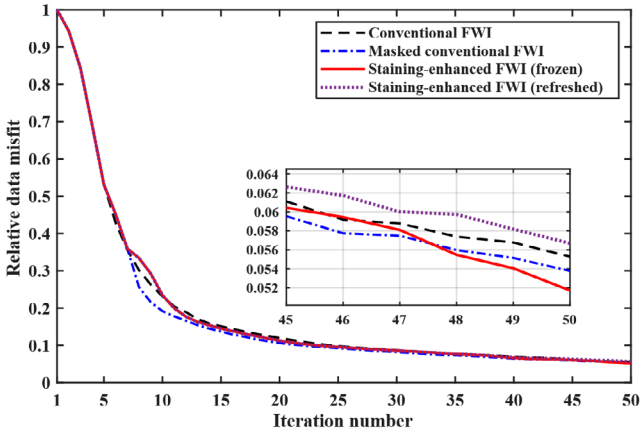


Figure 7. Relative-misfit convergence curves over 50 iterations for conventional full-waveform inversion (FWI), masked conventional FWI, and staining-enhanced FWI with the frozen and the refreshed staining-based local gradient

Table 1. Quantitative comparison of the inverted models of the three-layer example (Figure 6) after 50 iterations

Method	Mean velocity in Ω (m/s)	RMS error (m/s)	Local SSIM	Final relative misfit
True model	4,000	–	–	–
Conventional full-waveform inversion (FWI)	3,427	207.9	0.563	0.055
Masked conventional FWI	3,131	266.2	0.228	0.054
Staining-enhanced FWI (frozen)	4,000	182.8	0.652	0.052
Staining-enhanced FWI (refreshed)	3,992	172.7	0.676	0.057

Note: The root mean square (RMS) velocity error and the local structural similarity index measure (SSIM) are computed within a window extending five grid points beyond Ω; the final relative misfit is taken from Figure 7.

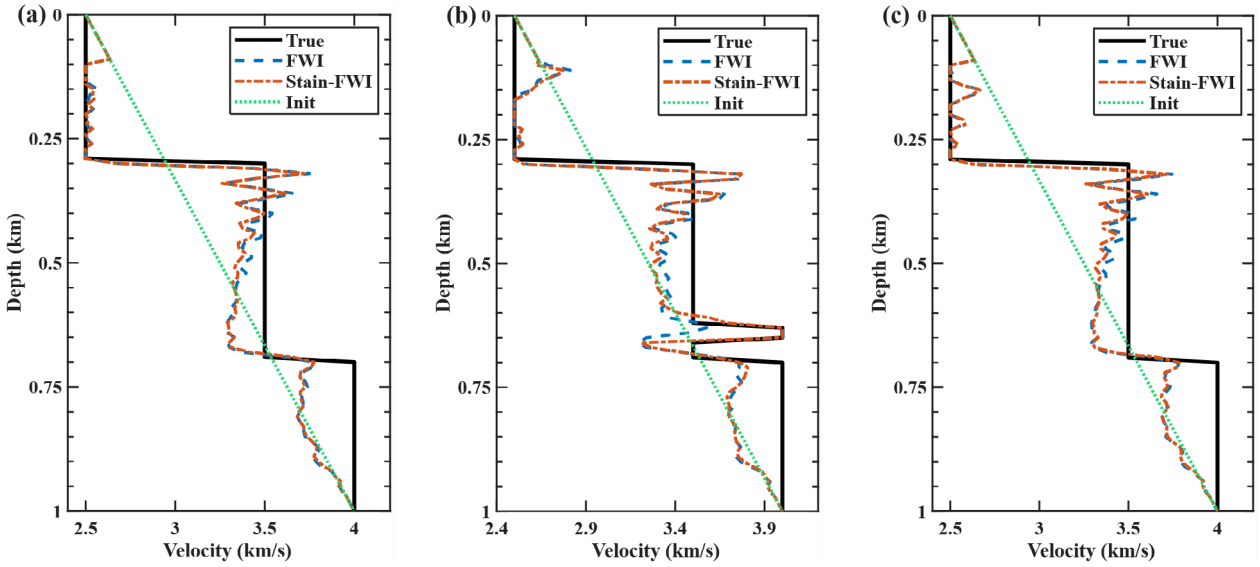


Figure 8. Single-trace velocity profile (true/initial/conventional full-waveform inversion [FWI]/staining-FWI) at (a) x = 0.3 km, (b) x = 0.5 km, (c) x = 0.7 km

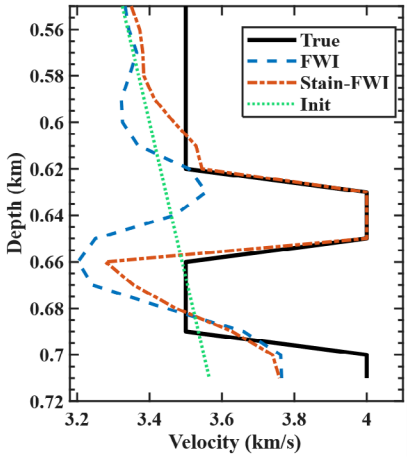


Figure 9. Zoomed single-trace velocity comparison (true model/conventional full-waveform inversion [FWI]/staining-enhanced FWI) around the target depth (z ≈ 0.64 km).

Figures 6b and 6d test the frozen-gradient approximation in Section 2.2: in Figure 6d, the staining-based local gradient was recomputed every 10 iterations from the current model and data residual. The two models are nearly indistinguishable (Table 1), although the raw amplitude of the recomputed local correlation decreased from 2.46 at activation to 0.09 at iteration 45, consistent with its linear dependence on the residual (Appendix A1). Because refreshing roughly doubles the cost of the iterations in which it is performed, computing the local gradient once at activation is adequate here; the consequences for the stationary points are discussed in Appendix A3.

Figure 7 compares the convergence of the four inversions. The curves coincide until staining is activated at iteration 5, and the final relative misfits differ by less than 0.01 (Table 1), because the data misfit is only weakly sensitive to such a small target.

Figure 8 shows single-trace velocity profiles extracted through the anomaly at the horizontal midpoint of the model, comparing the true model, the initial model, the conventional FWI result, and the staining-FWI result. Across most of the depth range, the conventional FWI and staining-FWI curves coincide, and both track the true model reasonably well; the key difference is concentrated in the anomaly's depth interval, where the conventional FWI curve shows pronounced oscillation/ringing around the true value while the staining-FWI curve reproduces the anomaly's velocity level and its two edges far more cleanly. This is the trace-level counterpart of the two-dimensional artifact seen in Figure 6a and directly illustrates the mechanism proposed in Section 2.2: because the staining wavefield used in the target's local gradient excludes the unrelated background reflections that generate the ringing, the corresponding part of the model converges to the correct value instead of oscillating around it. Figure 9 presents zoomed single-trace velocity comparisons around the target depth ($z \approx 0.64$ km), further highlighting the improved edge definition obtained with staining-enhanced FWI.

To assess how the method behaves when the target region Ω is imperfectly known, the sensitivity of the stained gradient to horizontal mislocation of Ω is tested on a background model containing a single true scatterer of known position and radius. The marked region Ω—matched in radius to the true scatterer—is offset from the true position by amounts ranging from −80 m to +80 m (nine offsets in total). For each offset, the stained gradient is formed by correlating a genuine, mask-independent data residual (the difference between conventional forward simulations of the true, anomaly-bearing model and the background model) against the corresponding ESP-staining wavefield, stacked over five shots spanning the target laterally to average out shot-specific near-surface imaging artifacts.

Figure 10 summarizes the peak-location and peak-amplitude behavior across the full range of tested offsets. Within approximately ±40 m of the true position, the stained-gradient peak remains anchored close to the true target location, essentially independent of exactly where Ω is placed within that range, rather than tracking the (incorrect) marked-region center. Beyond approximately ±60 m—where Ω and the true scatterer's spatial supports no longer meaningfully overlap—the peak-to-true-position distance increases sharply while the peak-to-mask-center distance plateaus, indicating that the focused energy has shifted to follow the (incorrect) marked location instead of the true target. This transition offset closely matches the sum of the true scatterer's and Ω's radii (40 m each in this test), consistent with a simple spatial-overlap criterion for when staining still finds the correct target. The normalized peak amplitude follows the same pattern, decreasing roughly symmetrically as Ω moves away from the true position in either direction. To verify that this behavior carries over from the gradient to the final inversion, full staining-enhanced inversions were performed with Ω laterally shifted by 20 m and 60 m in the +x direction, with all other settings identical to those in Figure 6 (Figure 11). For the 20 m shift, which lies within the overlap range identified above, the recovered anomaly remains at its true position and its shape and velocity are close to those obtained with the correctly placed Ω (Figure 11b). The staining term therefore does not drag the update toward the marked region; it only enhances the update where the target actually exists. For the 60 m shift, where Ω and the true scatterer no longer overlap, the recovery of the anomaly is clearly degraded (Figure 11d) and a weak artifact appears inside the shifted Ω. Hence, the method tolerates moderate errors in the target marking, provided that Ω still overlaps the true target, consistent with the spatial-overlap criterion derived from the gradient analysis.

A null test was further conducted to examine whether the staining term can create a spurious anomaly where none exists (Figure 12). The anomaly was removed from the true model, while Ω was kept at the original target location. The staining term was recomputed from the anomaly-free data, and all other settings were unchanged. As shown in Figure 12c, staining-enhanced FWI recovers the background layers without generating any spurious anomaly inside Ω. This result, together with the shifted-mask tests, indicates that the improvement in the target region originates from the information carried by the data through the staining wavefield, rather than from the prior specification of the target region itself.

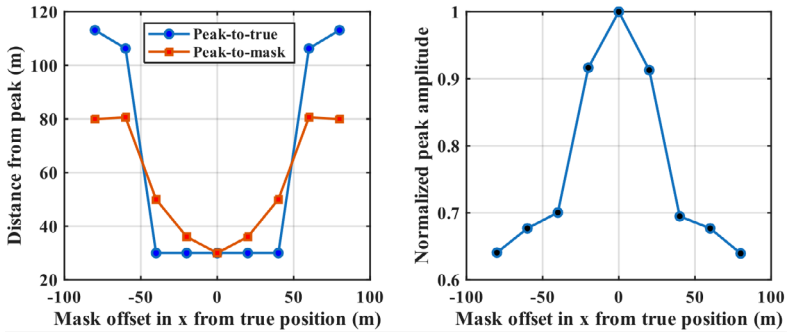


Figure 10. Sensitivity of the stained-gradient peak to horizontal mask offset. (a) Distance from the stained-gradient peak to the true target position (blue) and to the marked-region center (red), as a function of offset. (b) Normalized peak amplitude as a function of offset.

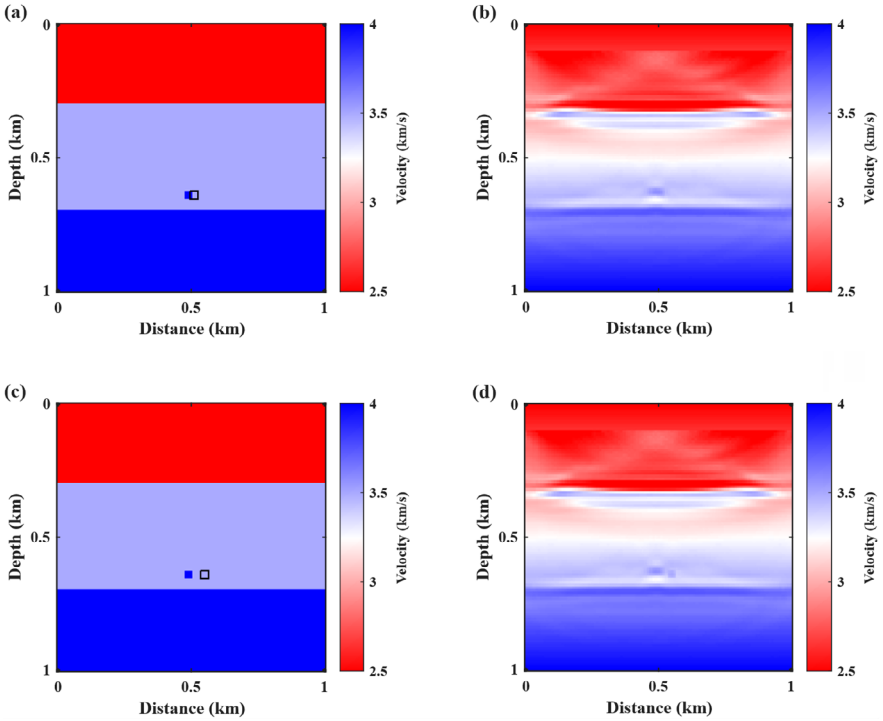


Figure 11. Sensitivity of staining-enhanced full-waveform inversion to a laterally mispositioned staining region. (a) True velocity model with the staining region shifted by 20 m in the +x direction (dashed box); (b) inversion result obtained with the 20 m shifted staining region; (c) true velocity model with the staining region shifted by 60 m in the +x direction (dashed box); (d) inversion result obtained with the 60 m shifted staining region.

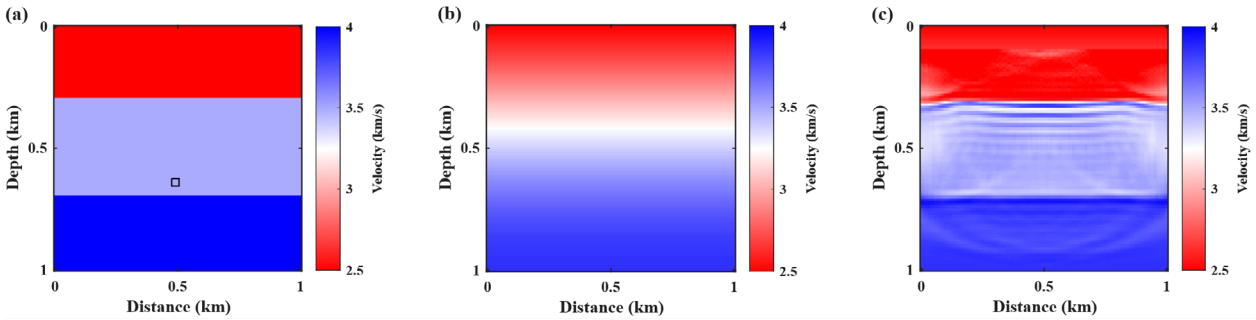


Figure 12. Null test of staining-enhanced full-waveform inversion (FWI) on the three-layer model without the anomaly: (a) true velocity model without the anomaly, with the staining mask region (dashed box) placed at the original target location; (b) initial velocity model, obtained by smoothing (a); (c) inversion result of staining-enhanced FWI.

3.2. Sigsbee sub-model

The Sigsbee sub-model, comprising 601 × 291 grid points with 10 m spacing, represents a mid-model region extracted from the standard Sigsbee salt-body benchmark and modified by embedding three small marked scatterer targets at mid-model depth, indicated by markers at 1, 3, and 5 km (Figure 13, true model; Figure 14, initial model).

A moving-spread surface acquisition is simulated with 41 shots and 321 receivers per shot, with the receiver spread shifting according to shot position. The source is a 20 Hz wavelet, and the total record length is 1.5 s (3,001 samples at 0.5 ms). Absorbing boundaries were applied on all sides of the model, and direct-arrival muting (travel-time-based, with a taper) was applied to both observed and modeled shot gathers prior to residual formation. FWI was run for 60 iterations using an acoustic GPU-accelerated forward/adjoint solver, an illumination-normalized PR^{+}-conjugate-gradient search direction, and the three-point parabolic step-length search. Following the staged strategy of Section 2.3, the run began as conventional FWI; once the relative misfit fell below 0.45 of its initial value, the gradient was switched to the staining-fused form, using an ESP-staining-based local gradient for the three marked targets and a Gaussian-tapered spatial mask.

Figure 15 presents the inverted velocity models after 60 iterations. Figure 15a displays the staining-enhanced FWI result, and Figure 15b shows the corresponding conventional FWI result obtained with identical acquisition, starting model, and iteration count.

Figure 16 compares full-depth vertical single-trace velocity profiles extracted through each of the three marked targets at x = 1, 3, and 5 km. Away from the targets, the staining-FWI and conventional FWI profiles are nearly indistinguishable, both closely tracking the true layered background velocity trend—consistent with the staining correction being strictly local and not perturbing the background inversion. At each target location, only the staining-FWI profile reproduces the sharp velocity spike present in the true model; the conventional FWI profile exhibits only a modest, broadened deviation. This illustrates that the marked small-scale targets remain substantially under-resolved by the conventional gradient in this complex model, consistent with the behavior observed in the simple-model test.

Figure 17 presents zoomed views of the same three trace locations, with both the depth window and velocity axis restricted to the immediate neighborhood of each target. Only the staining-FWI curve reproduces the target velocity spike at all three locations, whereas the conventional FWI curve remains close to the smooth background trend.

Figure 18 displays the relative-misfit convergence curves for both runs. The staining-enhanced run converges slightly faster and reaches a modestly lower final relative misfit compared with the conventional run. This suggests

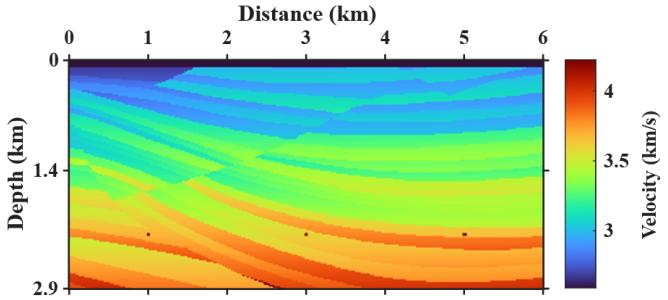


Figure 13. True velocity model for the Sigsbee sub-model, with three small marked target scatterers at mid-depth (markers at 1, 3, and 5 km)

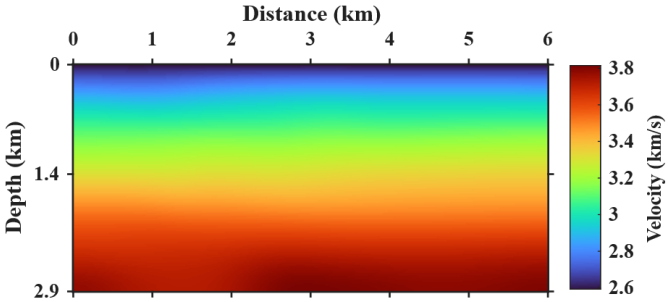


Figure 14. Smoothed initial velocity model used to start the full-waveform inversion iterations

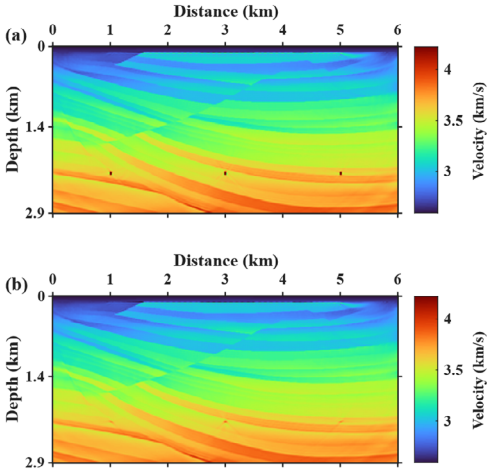


Figure 15. Sigsbee sub-model inversion result comparison. (a) Inverted velocity model after 60 iterations of staining-enhanced full-waveform inversion (FWI). (b) Inverted velocity model after 60 iterations of conventional FWI.

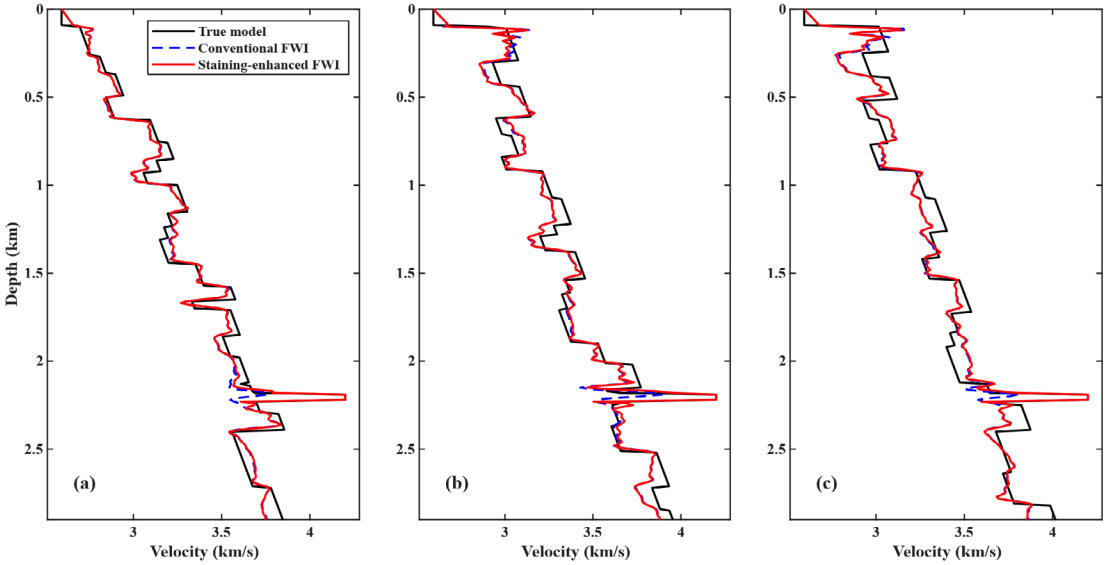


Figure 16. Full-depth vertical single-trace velocity comparison (true/staining-full-waveform inversion [FWI]/conventional FWI) at each of the three marked-target locations, shown as three separate panels: (a) x = 1 km, (b) x = 3 km, (c) x = 5 km

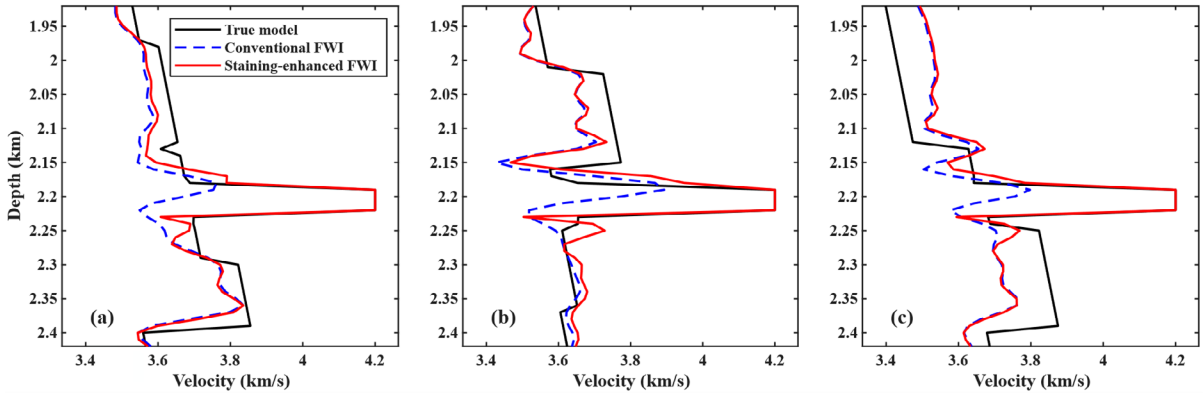


Figure 17. Zoomed single-trace velocity comparison (true model/conventional full-waveform inversion [FWI]/staining-enhanced FWI) around the target depth (z ≈ 2.18 km) at the three marked scatterer locations in the Sigsbee sub-model: (a) x = 1 km, (b) x = 3 km, (c) x = 5 km

that the improved local update at the marked scatterers also contributes modestly to the overall data fit.

To test the robustness to noise, band-limited Gaussian random noise, filtered with the amplitude spectrum of the source wavelet, was added to the observed shot gathers before direct-wave muting, giving an SNR of 5 dB. The SNR is defined as SNR = 20 $\log_{10}$(root mean square$_{signal}$/root mean square$_{noise}$), computed over the non-muted samples of all muted gathers with a single global noise level. The same noise realization was used for all inversions.

Figure 19 compares the 21st shot gather before and after noise contamination. With an SNR of 5 dB, the effective reflections, particularly the weak deep events, are severely masked by the band-limited noise, which shares the frequency band of the signal and therefore cannot be removed by simple filtering. Figure 20a and 20b show the inversion results obtained from the noisy data. Compared with the noise-free results, the noise causes slight blurring of the velocity interfaces and weak artifacts between layers, but the main structures are still recovered, indicating that both inversions are reasonably robust to noise. More importantly, the proposed staining-enhanced FWI still delineates the marked target anomalies more clearly than conventional FWI, in terms of both their boundaries and their velocity contrasts. This indicates that the benefit of the

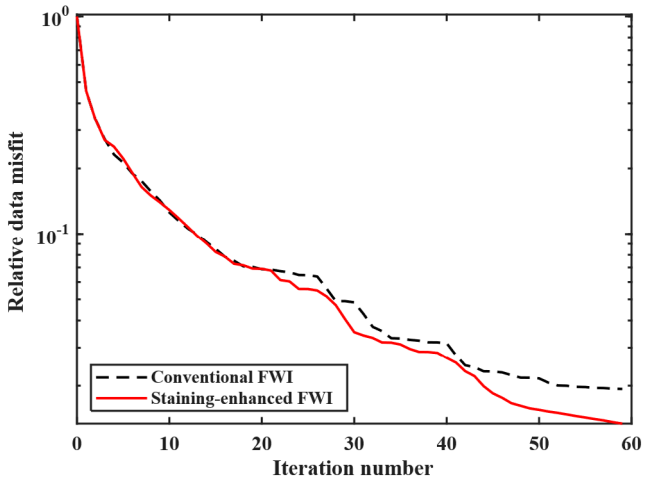


Figure 18. Relative-misfit convergence curves for staining-enhanced full-waveform inversion (FWI) versus conventional FWI over 60 iterations on the Sigsbee sub-model.

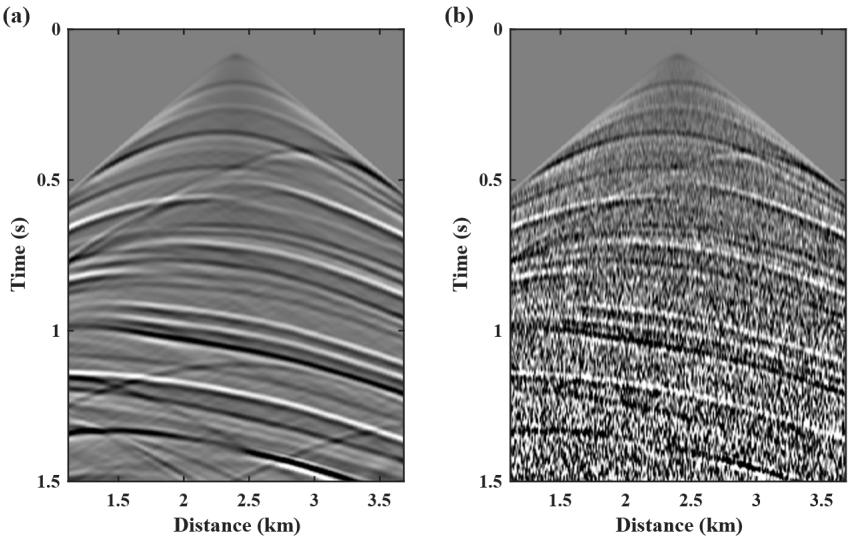


Figure 19. Comparison of the 21st shot gather of the Sigsbee model before and after noise contamination: (a) noise-free gather; (b) gather contaminated with band-limited Gaussian random noise (signal-to-noise ratio = 5 dB). The direct waves were muted, and the same time-power gain and display scale were applied to both panels.

staining term persists under strong noise contamination. Figure 20c and 20d show the differences between the results obtained with noisy and noise-free data; for both methods, the noise-induced velocity changes are mostly within ±23 m/s (99th percentile) and are concentrated in the shallow part of the model. Table 2 quantifies this effect: the noise increases the root mean square velocity error over the whole model by 1.8 m/s for staining-enhanced FWI and by 2.6 m/s for conventional FWI, whereas within 7 × 7-point windows centered on the three targets the error remains essentially unchanged; the same conclusion holds for larger 21 × 21-point windows. Staining-enhanced FWI is therefore not more sensitive to noise than conventional FWI and retains a markedly lower error at the targets.

4. Discussion

The results in Section 3 are consistent with the mechanism proposed in Section 2.2 and with the earlier CDV-based staining-FWI observations of Hu *et al.*:[2,32] the staining-enhanced gradient differs most from the conventional gradient precisely at the marked small-scale target, where the conventional correlation is corrupted by cross-talk from the surrounding structure, while the two gradients—and the resulting inverted models—are essentially indistinguishable away from the target. This locality is an important practical property: staining is a targeted correction, not a global regularizer, so it does not need to be validated against, or risk degrading, the parts of the model that conventional FWI already resolves adequately.

The Sigsbee sub-model misfit-convergence result additionally suggests that the benefit of staining is not confined to the target's immediate neighborhood. Because the three marked targets are embedded within the same wave-equation system as the rest of the model, a more accurate local update at the targets may also slightly improve the overall data fit. Disentangling how much of the misfit-curve improvement is attributable strictly to the target region versus to this indirect global effect would require an additional controlled test and is left for future work.

Because Ω is specified a priori, its mislocation relative to the true small-scale target is a natural concern. The gradient-level and full-inversion tests in Section 3.1 show that the stained update follows the true target as long as Ω still overlaps it spatially (offsets within approximately ±40 m), whereas beyond approximately ±60 m the focused energy follows the marked location instead. The transition

Table 2. Root mean square velocity error (m/s) of the Sigsbee sub-model inversions with noise-free and noisy data (signal-to-noise ratio = 5 dB), computed over the whole model and within 7 × 7-point windows centered on the three marked targets

Method	Whole model, noise-free	Whole model, noisy	Targets, noise-free	Targets, noisy
Staining-enhanced full-waveform inversion	62.8	64.5	105.4	101.4
Conventional full-waveform inversion	68.5	71.1	314.1	315.6

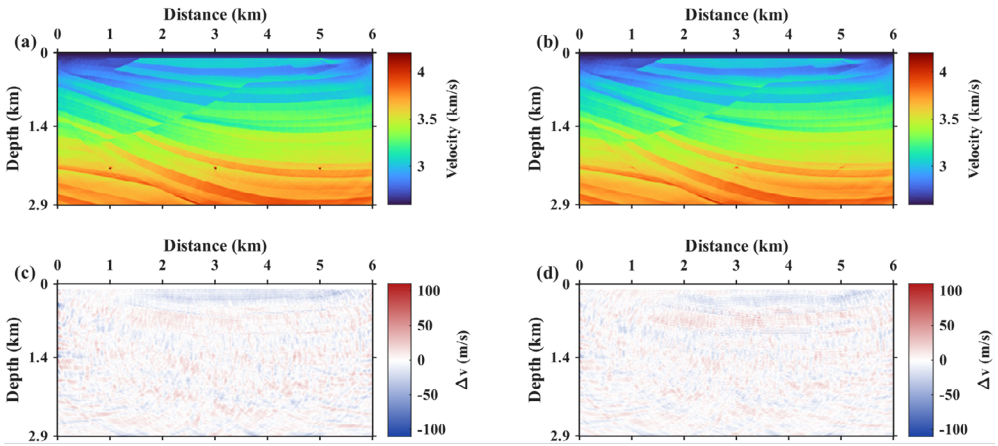


Figure 20. Effect of noise on the Sigsbee sub-model inversions (signal-to-noise ratio = 5 dB): (a) staining-enhanced full-waveform inversion (FWI) and (b) conventional FWI using the noisy data; (c) and (d) differences between the results obtained with noisy and noise-free data (noisy minus noise-free) for staining-enhanced and conventional FWI, respectively.

offset matches the sum of the radii of the true scatterer and Ω, consistent with a simple spatial-overlap criterion for the capture radius.

This behavior defines a practical capture radius for the method: Ω does not need to coincide exactly with the true target, but it does need to be delineated with an accuracy comparable to the target's own spatial extent. This is consistent with, and quantifies, the practical guidance given in Section 2.2—that Ω can be obtained from illumination/resolution analysis, geological priors, or a preliminary conventional FWI stage—methods whose typical localization accuracy is generally comparable to or finer than the scale of the small targets they are used to identify. It also clarifies the method's failure mode: when Ω does not overlap the true target, the recovery of the target degrades and a weak artifact can appear inside Ω (Figure 11d), whereas an Ω placed where no anomaly exists does not generate a spurious anomaly in the full inversion (Figure 12). We therefore recommend that Ω be cross-checked against an independent illumination or resolution indicator before being used to fuse a staining gradient in a field-data application.

Two implementation choices adopted for computational efficiency—noted in Section 2.2—are worth discussing explicitly. First, the staining-based local gradient was computed once, at the activation iteration, rather than re-derived at every FWI iteration. Refreshing it every 10 iterations from the current model and residual changed the recovered target only marginally in the three-layer test (Figures 6b and 6d), although the raw amplitude of the recomputed local correlation decreased by more than an order of magnitude as the residual decreased. Because the frozen term does not decay with the residual, the fused vector of **Equation 14** does not vanish at a minimizer of E; stationarity is instead guaranteed by the descent safeguard (Appendix A3), and a residual-scaled variant whose local term vanishes together with the residual is given in Appendix A4. The frozen approximation would need re-examination if staining were activated from the first iteration, when the model evolves rapidly. Second, the misfit-triggered activation strategy (switching on staining only once relative misfit drops below 0.65) means the reported results characterize a specific two-stage strategy rather than staining used throughout; an ablation over the activation threshold, and a comparison against activating staining from iteration 1, would help separate the contribution of staining itself from the contribution of the staged strategy.

Other limitations include sensitivity of the staining wavefield's accuracy to the perturbation coefficient η; the need to pre-define the marked target locations Ω_{target}, which in a field-data setting would require some prior knowledge or an auxiliary detection/localization step; and the additional cost of the staining computation relative to conventional acoustic FWI, which—while roughly half that of the CDV formulation—is still a nonzero overhead that scales with how often the staining gradient is refreshed. Finally, the present tests use the acoustic wave equation; extending the approach to elastic or anisotropic FWI, where the adjoint-state gradient and the staining wavefield equations both acquire additional terms, is a natural next step.

5. Conclusion

We have proposed a staining-enhanced acoustic FWI scheme for small-scale targets, in which the ESP staining wavefield replaces the conventional forward wavefield in the gradient correlation inside a marked target region, while the conventional adjoint-state gradient is retained elsewhere. The main conclusions are as follows.

(i) The staining-fused quantity is not the gradient of the least-squares misfit but a modified search direction whose local term is one component of the Hessian–vector product along the target perturbation; descent of the original objective is therefore enforced by checking each search direction against the true gradient, which guarantees a monotonic decrease of the misfit and stationarity only where the true gradient vanishes (Appendix A3).

(ii) On a three-layer model and on a modified Sigsbee sub-model, staining-enhanced FWI recovers the marked small-scale targets more accurately than conventional FWI, with a higher local structural similarity index measure, fewer residual-diffraction artifacts, and a modestly lower final data misfit, while the surrounding background is recovered comparably. A masked-conventional baseline using the same mask and weighting does not reproduce this improvement, and a null test without the anomaly produces no spurious target, showing that the benefit arises from the staining wavefield rather than from the prior marking of the target region. Refreshing the staining-based local gradient during the inversion changes the recovered target only marginally, which justifies computing it once at activation.

(iii) The method tolerates a mislocation of the marked region up to a capture radius comparable to the target size; beyond this radius, the correction focuses at the marked location, so the target region must be delineated with an accuracy comparable to the extent of the target.

(iv) With band-limited random noise at an SNR of 5 dB, the inversion remains stable, and staining-enhanced FWI still delineates the target more clearly than conventional

FWI, and the noise-induced increase in velocity error is no larger than that of conventional FWI (Table 2).

(v) Because the staining wavefield is obtained from the difference between two real-valued simulations, its cost is about half that of the CDV staining formulation.

Acknowledgments

None.

Funding

This work is supported by the National Science and Technology Major Project for New Oil and Gas Exploration and Development (Grant No. 2025ZD1400503-004).

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Jiajun Song, Yingming Qu
Formal analysis: Jiajun Song
Investigation: All authors
Methodology: Jiajun Song, Yingming Qu
Writing–original draft: Jiajun Song, Yingming Qu
Writing–review & editing: All authors

Availability of data

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

References

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

   doi: 10.1190/1.1441754

2. Hu C, Jia X, Li Q. Staining algorithm for full waveform inversion. In: Proceedings of the SEG International Exposition and 86th Annual Meeting. Society of Exploration Geophysicists; 2016:1444-1448.

   doi: 10.1190/segam2016-13850857.1

3. Tromp J. Seismic wavefield imaging of Earth's interior across scales. *Nat Rev Earth Environ*. 2019;1(1):40-53.

   doi: 10.1038/s43017-019-0003-8

4. Yang D, Dong X, Huang J, *et al.* High-resolution full waveform seismic imaging: Progresses, challenges, and prospects. *Sci China Earth Sci*. 2025;68:315-342.

   doi: 10.1007/s11430-024-1498-0

5. Yong P, Brossier R, Métivier L, Virieux J. Localized adaptive waveform inversion: theory and numerical verification. *Geophys J Int*. 2023;233(2):1055-1080.

   doi: 10.1093/gji/ggac496

6. Wang W, McMechan GA, Ma J. Reweighted variational full-waveform inversions. *Geophysics*. 2023;88(4):R499-R512.

   doi: 10.1190/geo2021-0766.1

7. Zhou L, Dong S, Han L, Zhang P, Hu Y. Full waveform inversion based on dynamic data matching of convolutional wavefields. *Front Earth Sci*. 2023;11:1134871.

   doi: 10.3389/feart.2023.1134871

8. Xiong K, Lumley D, Zhou W. The importance of data preconditioning strategies for envelope full-waveform inversion methods: Demonstration on marine seismic data. *Geophysics*. 2024;89(4):R415-R428.

   doi: 10.1190/geo2023-0322.1

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

   doi: 10.1029/2022JB025964

10. Schuster GT, Chen Y, Feng S. Review of physics-informed machine-learning inversion of geophysical data. *Geophysics*. 2024;89(6):T337-T356.

    doi: 10.1190/geo2023-0615.1

11. Niu Y, Qu Y. A physics-constrained autoencoder for full-waveform inversion method based on axial self-attention. *J Seism Explor*. 2026;35(1):269-281.

    doi: 10.36922/JSE025480119

12. Bear G, Liu CP, Lu R, *et al.* The construction of subsurface illumination and amplitude maps via ray tracing. *Lead Edge*. 2000;19(7):726-728.

    doi: 10.1190/1.1438700

13. Gelius LJ, Lecomte I, Tabti H. Analysis of the resolution function in seismic prestack depth imaging. *Geophys Prospect*. 2002;50(5):505-515.

    doi: 10.1046/j.1365-2478.2002.00331.x

14. Lecomte I. Resolution and illumination analyses in PSDM: A ray-based approach. *Lead Edge*. 2008;27(5):650-663.

    doi: 10.1190/1.2919584

15. Xie XB, Wu RS, Fehler M, Huang L. Seismic resolution and illumination: A wave-equation based analysis. In: Proceedings of the SEG 75th Annual International Meeting. 2005:1862-1865.

    doi: 10.1190/1.2148066

16. Xie XB, Jin SW, Wu RS. Wave-equation based seismic illumination analysis. *Geophysics*. 2006;71(5):S169-S177.

    doi: 10.1190/1.2227619

17. Xie XB, Yang H. A full-wave equation based seismic illumination analysis method. In: Proceedings of the 70th EAGE Conference and Exhibition. June 9-12, 2008; Rome, Italy.

doi: 10.3997/2214-4609.20148022

18. Yang H, Xie XB, Jin S, Luo M. Target oriented full-wave equation based illumination analysis. In: Proceedings of the SEG 78th Annual International Meeting. 2008:2216-2220.

doi: 10.1190/1.3059326

19. Cao J, Wu RS. Full-wave directional illumination analysis in the frequency domain. *Geophysics*. 2009;74(4):S85-S93.

doi: 10.1190/1.3131383

20. Cao J. Resolution/illumination analysis and imaging compensation in 3D dip-azimuth domain. In: Proceedings of the SEG 83rd Annual International Meeting. 2013:3931-3936.

doi: 10.1190/segam2013-0380.1

21. Chen B, Xie XB. An efficient method for broadband seismic illumination and resolution analyses. In: Proceedings of the SEG 85th Annual International Meeting. 2015:4227-4231.

doi: 10.1190/segam2015-5926976.1

22. Yao G, Wu D, Wang SX. A review on reflection-waveform inversion. *Pet Sci*. 2020;17(2):334-351.

doi: 10.1007/s12182-020-00431-3

23. Sun X, Xu W, Xiong Y, Li Y, Song X. Feasibility study of full waveform inversion of Love-SH waves for detecting small-scale shallow anomalies. *AIP Adv*. 2025;15(4):045320.

doi: 10.1063/5.0260272

24. Chen B, Jia X. Staining algorithm for seismic modeling and migration. *Geophysics*. 2014;79(4):S121-S129.

doi: 10.1190/geo2013-0262.1

25. Dale L, Slack JMW. Fate map for the 32-cell stage of Xenopus laevis. *Development*. 1987;99(4):527-551.

doi: 10.1242/dev.99.4.527

26. Ginhoux F, Greter M, Leboeuf S, *et al*. Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science*. 2010;330(6005):841-845.

doi: 10.1126/science.1194637

27. Li Q, Jia X. Time-partition staining algorithm for target-oriented migration. In: Proceedings of the SEG 85th Annual International Meeting. 2015:4370-4374.

doi: 10.1190/segam2015-5888935.1

28. Chen B, Ning H, Xie XB. Broadband seismic illumination and resolution analyses based on staining algorithm. *Appl Geophys*. 2016;13(3):480-490.

doi: 10.1007/s11770-016-0555-z

29. Li Q, Jia X. Generalized staining algorithm for seismic modeling and migration. *Geophysics*. 2017;82(1):T17-T26.

doi: 10.1190/geo2015-0652.1

30. Li Y, Jia X. High-SNR staining algorithm for one-way wave equation based modeling and imaging. *Commun Comput Phys*. 2020;28(1):187-206.

doi: 10.4208/cicp.OA-2018-0122

31. Chen B, Ning H, Xie XB. Investigating the effect of shallow scatterings from small-scale near-surface heterogeneities to seismic imaging: A resolution analysis based method. *Chinese J Geophys*. 2016;59(5):1762-1775.

doi: 10.1002/cjg2.20228

32. Hu C, Jia X. Staining algorithm for full-waveform inversion. In: Proceedings of the SPG/SEG Beijing 2016 International Geophysical Conference. 2016:610-611.

doi: 10.1190/IGCBeijing2016-187

33. Yu Y, Jia X. A frequency-domain staining algorithm for full waveform inversion with low SNR data. In: Proceedings of the 82nd EAGE Annual Conference & Exhibition. 2021:1-5.

doi: 10.3997/2214-4609.202011372

34. Liu C, Qu Y, Li Z, Zeng S, Yang T, Zhao W. Staining algorithm for least-squares reverse time migration. *J Appl Geophys*. 2023;219:105243.

doi: 10.1016/j.jappgeo.2023.105243

35. Tao T, Qu Y, Li Z, Dong Y, Ren Y. Viscoacoustic VTI media reverse time migration method in the complex-domain based on the staining algorithm. *J Appl Geophys*. 2024;225:105352.

doi: 10.1016/j.jappgeo.2024.105352

36. Liu C, Qu Y, Li Z, Zeng S, Yang T, Zhao W. Regional staining reverse time migration algorithm based on energy-constrained factors. *Acta Geophys*. 2024;72(3):1807-1817.

doi: 10.1007/s11600-023-01205-2

37. Song J, Qu Y, Li Z, Tao T. Viscoacoustic reverse time migration based on the region staining algorithm. In: Proceedings of the 86th EAGE Annual Conference & Exhibition. 2025:1-5.

doi: 10.3997/2214-4609.202510910

38. Qu Y, Li Z. New paradigms in seismic imaging: advances in localized, precise, and intelligent solutions. *J Geophys Eng*. 2026.

doi: 10.1093/jge/gxag107

39. Li Z, Qu Y. Research progress on seismic imaging technology. *Pet Sci*. 2022;19(1):128-146.

doi: 10.1016/j.petsci.2022.01.015

Appendix A. Supplementary derivations

A1. Staining correlation as a component of the Hessian–vector product

The directional derivative of the adjoint-state gradient g of **Equation 4** along a model perturbation δv, i.e., the Hessian–vector product $\boldsymbol{H}\delta\boldsymbol{v}$, follows from perturbing each factor in **Equation 4**:

$$\delta g(\boldsymbol{x}) = -\frac{6\delta v}{v^4}\sum_{x_s}\int_0^T \frac{\partial^2 p}{\partial t^2}\lambda dt + \frac{2}{v^3}\sum_{x_s}\int_0^T \frac{\partial^2 \delta p}{\partial t^2}\lambda dt + \frac{2}{v^3}\sum_{x_s}\int_0^T \frac{\partial^2 p}{\partial t^2}\delta\lambda dt \tag{A1}$$

where δp is governed by **Equation 9** and $\delta\lambda$ is the perturbation of the adjoint wavefield, which contains both the change of the data residual (the Gauss–Newton contribution) and the change of the adjoint wave operator. When δv is the ESP perturbation of **Equation 7**, the second term on the right-hand side of **Equation A1** coincides, inside Ω_{target}, with the local correlation of **Equation 11**, that is,

$$g_{\text{stain}}(\boldsymbol{x}) = [\boldsymbol{H}\delta\boldsymbol{v}]_{\delta p}(\boldsymbol{x}),\ \boldsymbol{x}\in\Omega_{\text{target}} \tag{A2}$$

where $[\cdot]_{\delta p}$ denotes the source-side component of the Hessian–vector product. The staining correlation therefore carries curvature information along the target perturbation and is not the gradient of E. Moreover, the adjoint wavefield is a linear function of the data residual $\boldsymbol{r} = L(\boldsymbol{v}) - \boldsymbol{d}_{\text{obs}}$: with $\mathcal{A}$ the wave operator of **Equation 3** and $\boldsymbol{R}^T$ the operator that injects the residual at the receivers,

$$\lambda = \mathcal{A}^{-\mathrm{T}}\boldsymbol{R}^T\boldsymbol{r} \quad\Rightarrow\quad \left\|g_{\text{stain}}\right\| \le C\left\|\boldsymbol{r}\right\| \tag{A3}$$

where the constant C depends on p and δp but not on $\boldsymbol{r}$. The raw staining correlation thus vanishes together with the data residual.

A2. Dependence on the perturbation coefficient and its polarity

With $\delta v = \eta v$ inside Ω_{target} (**Equation 7**), the source term on the right-hand side of **Equation 9** is proportional to η and vanishes outside Ω_{target}. Because the operator on the left-hand side of **Equation 9** is linear and independent of η, $\delta p = \eta\delta\tilde{p}$, where $\delta\tilde{p}$ solves **Equation 9** with $\eta = 1$. Substituting into **Equation 11** gives **Equation A4**:

$$\delta p = \eta\delta\tilde{p} + O\left(\eta^2\right),\quad g_{\text{stain}} = \eta\tilde{g}_{\text{stain}} + O\left(\eta^2\right) \tag{A4}$$

where the $O(\eta^2)$ terms stem from the term discarded in **Equation 8**; they are absent from the linearized **Equation 9** but present in the ESP difference of two nonlinear simulations. Defining the local term per unit perturbation and normalizing it by its maximum absolute value (**Equation 14**) gives **Equation A5**:

$$\hat{g}_{\delta p}(\boldsymbol{x}) = \frac{g_{stain}(\boldsymbol{x})/\eta}{max_{\boldsymbol{x}'\in\Omega}\left|g_{stain}(\boldsymbol{x}')/\eta\right|} = \frac{\tilde{g}_{stain}(\boldsymbol{x})}{max_{\boldsymbol{x}'\in\Omega}\left|\tilde{g}_{stain}(\boldsymbol{x}')\right|} + O(\eta) \tag{A5}$$

which is independent of both $|\eta|$ and sign(η) to first order; η affects $\hat{g}_{\delta p}$ only through the $O(\eta)$ linearization error, which is why $|\eta|$ must be small. For the positive η used in this study, the division by η does not change the normalized local gradient. The sign of $\hat{g}_{\delta p}$ is set by the data residual through λ (**Equation A3**), not by the polarity of the embedded perturbation. **Equation 9** corresponds to $\delta p = p(\boldsymbol{v} + \delta\boldsymbol{v}) - p(\boldsymbol{v})$; if the ESP difference is taken in the opposite order, the sign of $\boldsymbol{g}_{\text{stain}}$ is reversed and must be compensated in **Equation 14**.

A3. Descent property and stationary points of the safeguarded iteration

For the update $\boldsymbol{v}_{k+1} = \boldsymbol{v}_k - \alpha\boldsymbol{d}_k$ of **Equation 2**, a first-order expansion of the misfit gives **Equation A6**:

$$E\left(\boldsymbol{v}_k - \alpha\boldsymbol{d}_k\right) = E\left(\boldsymbol{v}_k\right) - \alpha\boldsymbol{g}_k^T\boldsymbol{d}_k + O\left(\alpha^2\right) \tag{A6}$$

where $\boldsymbol{g}_k$ is the true adjoint-state gradient of **Equation 4**. Hence $\boldsymbol{d}_k$ is a descent direction if and only if $\boldsymbol{g}_k^T\boldsymbol{d}_k > 0$, which is the check used in Section 2.2. The fallback direction, the illumination-preconditioned gradient of **Equation 6**, satisfies **Equation A7**:

$$\boldsymbol{g}_k^T\hat{\boldsymbol{g}}_k = \sum_x \frac{g_k^2(\boldsymbol{x})}{I(\boldsymbol{x}) + \varepsilon} > 0 \quad \text{whenever} \quad \boldsymbol{g}_k \neq 0 \tag{A7}$$

as $I(\boldsymbol{x}) \geq 0$ and $\varepsilon > 0$. The safeguarded search direction (**Equation A8**):

$$\boldsymbol{d}_k^s = \begin{cases} \boldsymbol{d}_k, & \boldsymbol{g}_k^T \boldsymbol{d}_k > 0 \\ \hat{\boldsymbol{g}}_k, & \text{otherwise} \end{cases} \tag{A8}$$

is therefore a descent direction whenever $\boldsymbol{g}_k \neq 0$, and together with the acceptance rule $E\left(\boldsymbol{v}_k - \alpha \boldsymbol{d}_k^s\right) < E\left(\boldsymbol{v}_k\right)$ of the line search (Section 2.2), it produces a monotonically decreasing misfit sequence. Conversely, if $\boldsymbol{g}_k = \boldsymbol{0}$, then $\boldsymbol{g}_k^T \boldsymbol{d}_k = 0$ for any $\boldsymbol{d}_k$, the check fails and $\boldsymbol{d}_k^s = \hat{\boldsymbol{g}}_k = \boldsymbol{0}$, so that the model is no longer updated. The stationary points of the safeguarded iteration therefore coincide with those of E, even though the fused vector of **Equation 14** itself does not vanish there when the local term $\hat{g}_{\delta p}$ is frozen. Because the check is global, it does not by itself constrain the direction of the update inside Ω_{target}; the local behavior is assessed by the numerical examples in Section 3.

A4. Residual-scaled local term

If the staining-based local gradient is refreshed during the inversion but normalized only once, at the activation iteration k_a:

$$\hat{g}_{\delta p}^{(k)}(\boldsymbol{x}) = \frac{g_{stain}^{(k)}(\boldsymbol{x})}{max_{\boldsymbol{x}' \in \Omega} \left| g_{stain}^{(k_a)}(\boldsymbol{x}') \right|}, \quad k \geq k_a \tag{A9}$$

then, by **Equation A3**, $\left|\hat{g}_{\delta p}^{(k)}\right| \leq C \|\boldsymbol{r}_k\| / max_{\Omega} \left| g_{\text{stain}}^{(k_a)} \right|$, so that the local term decays together with the data residual. For a zero-residual problem (noise-free data and an exact model parameterization), both terms of the fused vector of **Equation 14** then vanish at the solution, which restores consistency between the fused vector and the stationarity condition of E, at the cost of refreshing the staining computation. In the three-layer example, the raw peak of the recomputed local correlation decreased from 2.46 at activation to 0.09 at iteration 45, while the relative misfit decreased from 0.53 to 0.06, consistent with **Equation A3**. For problems with a nonzero residual at the solution, the safeguard in **Appendix A3** remains the mechanism that ensures stationarity.