A variational approach for picking optimal surfaces from semblance-like panels[a]

Luke Decker and Sergey Fomel

ABSTRACT

We propose and demonstrate a variational method for determining optimal velocity fields from semblance-like volumes using continuation. The proposed approach finds a minimal-cost surface through a volume, which often corresponds to a velocity field within a semblance scan. This allows picked velocity fields to incorporate information from gathers that are spatially near the midpoint in question. The minimization process amounts to solving a nonlinear elliptic partial differential equation, which is accomplished by changing the elliptic problem to a parabolic one and solving it iteratively until it converges to a critical point which minimizes the cost functional. The continuation approach operates by using a variational framework to iteratively minimize the cost of a velocity surface through successively less-smoothed semblance scans. The method works because a global minimum for the velocity cost functional can only exist when the semblance scan varies smoothly in space and convexly in the parameter being scanned. Using a discretization of the functional with a limited-memory Broyden-Fletcher-Goldfarb-Shanno (ℓ-BFGS) algorithm we illustrate how the continuation approach is able to avoid local minima that would typically capture the iterative solution of an optimal velocity field determined without continuation. Incorporating continuation enables us to find a lower cost final model which is used for seismic processing of a field data set from the Viking Graben. We then utilize a field data set from the Gulf of Mexico to show how the final velocity model determined by the method employing continuation is largely independent of the starting velocity model, producing something resembling a global minimum. Finally, we illustrate the versatility of the variational picking approach by demonstrating how it may be used for automatic interpretation of a seismic horizon from the Heidrun Field.

INTRODUCTION

The concept of picking normal moveout (NMO) velocity from the dominant semblance trend in velocity spectral display panels was pioneered by Taner and Koehler (1969). These panels measure how well applying a parameterized operation, like the NMO correction, over a parameter sweep transform data in some metric. In the case of

Taner and Koehler (1969), this meant NMO gather flatness as measured by semblance. Numerous other geophysical applications exist including dip moveout (DMO) velocity analysis (Hale, 1984; Deregowski, 1986; Yilmaz, 2001); migration velocity analysis (Fowler, 1988; Deregowski, 1990; Fomel, 2003; Decker et al., 2017; Decker and Fomel, 2021); image registration with different time shifts (Hale, 2013); data alignment by local similarity scan and similar applications (Fomel, 2007a, 2009; Bader et al., 2019); deconvolution with dynamic frequency wavelets (Decker and Fomel, 2018); picking geobodies, faults, and seismic horizons in automatic interpretation (Wu and Fomel, 2018a,b; Yan and Wu, 2021); and determining the principal anisotropic axis in image gathers (Decker and Zhang, 2020).

Because manually selecting such trends can be a labor-intensive process, a rich tradition of research has focused on increasing the ability of computers to learn the dominant trends in semblance-like panels with reduced need for human intervention. This has involved overcoming difficulties in the picking process where artificially or anomalously high semblance values do not reflect a high-quality fit. These artificially high values are often caused by artifacts related to lateral velocity variations (Hubral and Krey, 1980) and care must be taken to avoid outputting an unphysical velocity model. Early efforts, like those of Sherwood and Poe (1972), involved interpolating a line between the largest semblance values detected, but this required interpreter supervision to avoid artifacts and unphysical velocity trends. de Bazelaire (1988) used the theory of geometric optics (Born and Wolf, 1959) to devise an "optical stack" method that automatically output velocity information. Doicin et al. (1994) modified the optical stack approach by applying constraints which force the method to avoid unphysical velocity selections. Alder and Brandwood (1999) expanded on the method of Doicin et al. (1994) by using three-dimensional interpolation and locally scaled regression to determine a dense three-dimensional velocity field. This work was further extended by Siliqi et al. (2003) to solve for two parameters simultaneously: velocity and an annellipticity term which can help account for the effects of dipping reflectors or seismic anisotropy. Arnaud et al. (2004) explored how these picking algorithms may be implemented in situations where the phase or amplitude of a seismic reflection changes with offset. Larner and Celis (2007) demonstrated how selective-correlation velocity analysis could be used to improve the resolution of velocity spectra and lead to more accurate picks. Research has progressed on methods for automatic multiparameter scanning (Tao et al., 2012), including the efficient application of the three-dimensional version of the Fourier integral operator butterfly algorithm (Candès et al., 2009) to the problem by Hu et al. (2015).

The approach outlined here follows a branch of inquiry originated by Toldi (1989), who introduced a method for iteratively finding the best path through semblance using a conjugate-gradient solver given a priori knowledge of the likely velocity field and enforcing penalties for large changes in velocity. Symes and Carazzone (1991) proposed applying a variational principle to invert for velocity models from seismic reflection data using differential semblance. Differential semblance is advantageous because, in the case of noiseless data, it possesses a single minima and exhibits convex behavior near that minimum, but with a reduction in velocity resolution compared to

regular semblance (Symes, 1998, 1999; Mulder and ten Kroode, 2002; Li and Symes, 2007). Harlan (2001) simplified the method of Toldi (1989), replacing the a priori constraints of that method with a stiffness penalty and explicitly casting the ideal path through the semblance panel as the maximization of the variational integral:

$$\max_{v(\mathbf{x})} \int \alpha\left[v(\mathbf{x}), \mathbf{x}\right] d\mathbf{x}, \tag{1}$$

where $v(\mathbf{x})$ is a smooth surface defining the velocity and α is semblance. This maximization was performed using a Gauss-Newton algorithm (Luenberger and Ye, 1984). Although Harlan (2001) proposed a framework for solving for $v(\mathbf{x})$, a multidimensional surface, no examples were provided.

Inspired by the work of Deschamps and Cohen (2001) in virtual endoscopy, Fomel (2009) continued with the variational approach of Harlan (2001). Fomel (2009) noted that in the one-dimensional case of a velocity path through a single gather, Equation 1 could be reformulated to appear analogous to a ray-tracing equation solving for the minimal travel time of the following integral,

$$\min_{v(t)} \int_{t_o}^{t_f} \exp\left(-\alpha\left[v(t), t\right]\right) \sqrt{\lambda^2 + \left(\frac{dv}{dt}\right)^2}\, dt, \tag{2}$$

where λ is a parameter that modifies the cost of changing position in t relative to changing position in v. Using variational methods (Lanczos, 1966; Greenberg, 1978; Gelfand and Fomin, 2000), the optimal $v(t)$ in Equation 2 is determined by solving the eikonal equation (Babich and Buldyrev, 1972; Yilmaz, 2001) for $U(t, v)$,

$$\left(\frac{\partial}{\partial v} U(t, v)\right)^2 + \frac{1}{\lambda^2}\left(\frac{\partial}{\partial t} U(t, v)\right)^2 = \exp\left(-2\alpha\left[v(t), t\right]\right), \tag{3}$$

which, given a $v(t_o)$, may be done using a finite-difference algorithm (Iserles, 1996). After determining $U(v, t)$, the method finds $v(t)$ by tracking backward along ∇U from the $v(t_f)$ that minimizes $U(v, t_f)$. Oscillations in $v(t)$ are dampened using shaping regularization (Fomel, 2007b).

We propose to expand the approach of Fomel (2009) to picking a multidimensional surface by minimizing a functional resembling semblance-weighted total variation regularization. This formulation enables direct use of information from spatially adjacent semblance panels to determine a continuous velocity field without explicitly enforcing smoothing. It improves upon the existing one-dimensional, gather-by-gather approach of Fomel (2009) by incorporating spatially adjacent information into the picking algorithm. The proposed approach is equivalent to a nonlinear elliptic partial differential equation, which can be challenging to solve directly. Instead, we propose finding minimizing surfaces iteratively. Because common iterative methods seek out the nearest minimizer for the velocity functional, which can be highly multimodal, successful implementation may require an accurate starting model. Additionally,

gradient or steepest descent methods, which can be used for iteratively finding a minima for a functional, may require many iterations to converge.

Continuation, or graduated optimization (Blake and Zisserman, 1987; Chapelle et al., 2006; Chaudhuri and Solar-Lezama, 2011; Mobahi and Fisher, 2015; Hazan et al., 2016; Xue et al., 2016), is a method of non-convex optimization. Using this approach, local minima may be avoided by solving a series of successively more challenging, or less convex, approximations to an optimization problem. The solution of a smoother, or more convex, problem is used as the starting model for more rugose one. Increasing the convexity of a minimization problem is frequently accomplished by convolving the objective function with a Gaussian kernel (Wu, 1996). Gaussian convolution may be prohibitively expensive, but may be efficiently approximated by triangle smoothing (Claerbout, 1993).

The limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm (ℓ-BFGS) refers to a class of quasi-Newton schemes for accelerating the convergence of an iterative minimizer without the need for a large amount of computer memory (Nocedal, 1980; Liu and Nocedal, 1989; Li and Fukushima, 2001). ℓ-BFGS works by using several gradient computations to build an approximation for the Hessian of a system. Such memory considerations become essential for large problems, as the Hessian is on the order of the number of samples in the input vector squared. Using the approximate Hessian enables the algorithm to draw on information about the curvature of cost function level sets, and thus provide a step direction leading more directly to a minimum.

In the following sections, we propose an extension of Fomel (2009) for picking velocity surfaces from semblance volumes. The problem is then discretized, and a ℓ-BFGS algorithm is used to accelerate convergence. We then demonstrate how continuation may be applied to the picking algorithm to bypass local minima and make the algorithm behave more like a global optimizer. Finally, the automatic picking algorithm is applied to an automatic interpretation problem to illustrate its versatility and applicability beyond picking velocity surfaces from semblance scans.

THEORY

As a natural extension of the one-dimensional functional of Fomel (2009) (Equation 2) to higher dimensions, consider the functional

$$\tilde{G}[v] = \int_{\Omega} e^{-\alpha[v(\mathbf{x}),\mathbf{x}]} \left(\sqrt{\lambda^2 + |\nabla v|^2} + \frac{\epsilon}{2} |\nabla v|^2 \right) d\Omega. \tag{4}$$

In Equation 4, we assume that λ and ϵ are strictly positive real numbers, and $\Omega \subset \mathbb{R}^n$ is a convex, bounded set with Lipschitz boundary $\partial\Omega = \Gamma$. Possible solutions v exist in the set of real valued functions $V \subset H^1(\Omega)$, the infinite-dimensional vector space of functions whose gradient components are square integrable, such that $\forall v \in V, \nabla v \cdot \hat{\mathbf{n}} = 0$ for $\hat{\mathbf{n}} \in \Gamma$, which is equivalent to a zero-flux boundary condition.

$\alpha \in C^1(V \times \Omega, [0, \bar{\alpha}])$ with $0 < \bar{\alpha} < \infty$, or α is a continuous function defined for all plausible velocities v in V and at all spatial locations $\mathbf{x}$ in study area Ω, such that its output is between 0 and some finite positive number $\bar{\alpha}$. $\alpha[v, \mathbf{x}]$ is assumed to be a Lipschitz continuous function in v such that $|\alpha_v| \leq L_{\alpha 1}$, or the magnitude of its partial derivative with respect to v is bounded by a constant, $L_{\alpha 1}$. Furthermore, the partial derivative of α in v is also assumed to be Lipschitz continuous with $|\alpha_{vv}| \leq L_{\alpha 2}$, or the magnitude of its second partial derivative with respect to v is bounded by a constant, $L_{\alpha 2}$. We use $|\mathbf{x}|$ to signify $\sqrt{\mathbf{x} \cdot \mathbf{x}}$ for $\mathbf{x} \in \mathbb{R}^n$, which reduces to the absolute value function if $x \in \mathbb{R}$. We wish to find the $v \in V$ that minimizes this equation and will serve as the picked velocity model.

α is a semblance-like volume over spatial domain Ω. λ and ϵ are regularization parameters. Increasing λ makes the minimizing functions more closely follow highs in α, while decreasing it makes the output minimizing functions smoother. ϵ is a small positive term which penalizes changes in minimizing functions and ensures they exists in $H^1(\Omega)$ (Dobson and Vogel, 1997). Although we found that ϵ may be set to negligibly small values on the order of machine precision in practice, a complete mathematical treatment of this problem, which is beyond the scope of this manuscript but may be found in Chapter 5 of Decker (2021) and draws inspiration from Dacorogna (2004) and Smyrlis and Zisis (2004), shows how the ϵ term is essential for guaranteeing the existence of and convergence to minimizers of Equation 4 in an infinite-dimensional setting. Again, larger values of λ lead to minimizing surfaces which may vary significantly to follow high values of α more closely. Larger values of ϵ ensure that minimizing surfaces have less total variation. The examples in this paper use $1 \leq \lambda \leq 5$ and $\epsilon = 10^{-3}$.

Minimizing surfaces of Equation 4 occur when its Fréchet derivative, defined as

$$\nabla \tilde{G}(v) = - \alpha_v[v(\mathbf{x}), \mathbf{x}] e^{-\alpha[v(\mathbf{x}), \mathbf{x}]} \left(\sqrt{\lambda^2 + |\nabla v|^2} + \frac{\epsilon}{2} |\nabla v|^2 \right)$$
$$- \nabla \cdot \left(e^{-\alpha[v(\mathbf{x}), \mathbf{x}]} \left[\frac{1}{\sqrt{\lambda^2 + |\nabla v|^2}} + \epsilon \right] \nabla v \right), \tag{5}$$

is equal to zero. In the right-hand side of Equations 4 and 5, gradient and divergence operations are applied in $\mathbf{x} \in \Omega$, where $\Omega \subset \mathbb{R}^n$ is the domain of velocity models, and are based on derivatives for each spatial dimension of $\mathbf{x}$. On the left-hand side of Equation 5, the gradient operation refers to the functional gradient, which is applied to a functional $\tilde{G}[v]$ which maps $v \in H^1(\Omega)$ to a real number. This functional gradient, $\nabla \tilde{G}[v]$, is itself a function defined on Ω.

Directly determining minimizing surfaces involves solving a nonlinear elliptic partial differential equation, which can be challenging. Instead, we choose to adopt an iterative approach by introducing a pseudo-time coordinate, τ, and solving the parabolic equation

$$\frac{\partial v}{\partial \tau} = -\nabla \tilde{G}(v), \tag{6}$$

by allowing v to evolve in τ until it converges to a critical v^* where $\left\|\nabla\tilde{G}(v^*)\right\| = 0$. This is a common approach for finding the minima of a functional (Mawhin and Willem, 2010) and suggests the use of an iterative gradient descent scheme. This method provides a sequence $\{v_i\}_{i\in\mathbb{N}}$ of subsequent approximations for v^*, the critical point of $\tilde{G}(v)$. The update structure given an initial velocity model function v_0 is

$$v_{i+1} = v_i + \rho_i h_i, \tag{7}$$

where v_i are updated velocity models and h_i are search directions. Control parameters, also known as step sizes, $\{\rho_i\}_{i\in\mathbb{N}}$ satisfy $a \le \rho_i \le b$ for positive real numbers a, b whose values are related to the smoothness of α (Decker, 2021). Experiments in this paper use $a = 10^{-8}$ and $b = 10^{-2}$. At each step in Equation 7, $\nabla\tilde{G}(v_i)$ is evaluated. If $\left\|\nabla\tilde{G}(v_i)\right\| = 0$ the method stops because it has arrived at a critical point or minimizer. Otherwise, h_i is chosen to satisfy $\left\langle\nabla\tilde{G}(v_i), h_i\right\rangle_{H^1(\Omega)} < 0$, where $\langle\cdot,\cdot\rangle_{H^1(\Omega)}$ is the H^1 inner product on Ω. In the method of steepest descent $h_i = -\nabla\tilde{G}(v_i)$. v_{i+1} is computed and the method proceeds to the next step. Example pseudocode for a picking algorithm using continuation can be found in Appendix ??.

An iterative algorithm for minimizing Equation 4 to pick an optimal surface from a semblance-like volume is implemented in the `Madagascar` software library (Fomel et al., 2013) using NumPy. Parallelism is achieved using Numba, and NumPy functions are rewritten as simple, array-based, object-free operations and performance tested using Numba on a Texas Advanced Computing Center (TACC) Stampede2 compute node with 48 Skylake cores for code optimization of each component helper function. The Fréchet derivative, or functional gradient $\nabla\tilde{G}(v)$, given by Equation 5 is computed using finite-difference derivative operations and the Numba functions written for the task. The finite-difference approach was chosen for numerical efficiency. Representing the large, semblance-like panels that are used for $\alpha[v(\mathbf{x}), \mathbf{x}]$ as FEniCS interpolated functions for finite element implementation was found to be overly expensive, and finite-difference computations yielded satisfactory results at reduced computational cost. The calculated gradient is utilized with the ℓ-BFGS two-loop recursion method outlined by Nocedal (1980) for approximating the inverse Hessian and calculating a search direction. After experimentation on different data sets, we found that a memory size of the past 3 gradient computations for use in the approximation of the inverse Hessian achieved the most rapid cost convergence results. The step size is determined using the output search direction with a golden-section search (Kiefer, 1953; Mordecai and Wilde, 1966). The method is allowed to iterate and update the model until either a maximum number of iterations are achieved, any step size in the search direction fails to decrease the cost by more than a set amount, or the L^2 magnitude of the calculated gradient drops below a threshold. We also found that the Hessian may change rapidly in this problem, leading some search directions output by the ℓ-BFGS scheme to increase the cost of surfaces for all step sizes. In this circumstance, the previous gradient computations are "forgotten" and the ℓ-BFGS algorithm is reset, enabling a fresh approximation of the inverse Hessian.

Careful readers will also notice a dimensional analysis ambiguity when calculating best-fit surfaces from typical semblance panels using the Fréchet derivative defined in Equation 5. Usually, the first axis in these data is seismic trace recording time, while other axes are spatial with units of distance. In order to convert derivatives in these different units into similar ones, spatial derivatives are multiplied by the velocity model for the previous iteration so they have units of s^{-1}.

Convergence of this algorithm requires at least local convexity (Dennis and Mor, 1977; Nocedal, 1980; Li and Fukushima, 2001), which is achieved by enforcing smoothness on $\alpha[v(\mathbf{x}), \mathbf{x}]$ using triangle smoothing filters (Claerbout, 1993). As a practical point, artificially large values in $\alpha[v(\mathbf{x}), \mathbf{x}]$ not related to high-quality fit for the parameter being tested must be muted to prevent them from capturing the cost minimizing surface. In this paper, we manually define muting functions.

Convergence can only be guaranteed to a local minimizer. For a unique minimizer to exist, the integrand of Equation 4 would need to be strictly convex in every v, ∇v combination for all $\mathbf{x} \in \Omega$ (Dacorogna, 2004), which places unrealistic burdens on α. In order to overcome the multimodality of this problem, we employ continuation, or graduated optimization. This involves smoothing and scaling the semblance-like volume several times. Smoothing makes the semblance volume more convex, and scaling increases the attraction of semblance highs for $v(\mathbf{x})$ surfaces. If sufficiently strong smoothing is applied, the problem may become convex (Hazan et al., 2016). In this paper, the authors choose a set of minimum smoothing radii that capture the finest scale of changes in the semblance-like volume that appear to be representative of the underlying model. The maximum smoothing radii are determined by multiplying the minimum radii by a sufficiently large number so that smoothing makes panels of the semblance volume appear strictly convex. We choose a number of smoothing levels, then proceed incrementally from maximum to minimum smoothing in a linear manner. The semblance scaling factor for each level is chosen to be proportional to the ratio between smoothing at that level and minimum smoothing. A constant-gradient velocity field is used as the starting model on the most-smoothed semblance volume, and the minimizing velocity surface is iteratively determined using the two-loop recursion ℓ-BFGS scheme until convergence on that model is achieved. The minimizing surface from one level of smoothing is then used as the starting model on the next-less-smoothed semblance volume. The method proceeds until finding the cost minimizing velocity on the least-smoothed model. Because the method is allowed to operate on semblance-like volumes with different levels of smoothing, it is able to find the macro-trends for the best-fit surface before determining finer components. This enables the method to behave more like a global optimizer and enjoy applicability to problems beyond picking velocity surfaces from semblance-like volumes as experiments in subsequent sections illustrate.

FIELD DATA EXAMPLES

Viking Graben

We apply the method to a field data example from the Viking Graben featured in Keys and Foster (1998). Figure 1a contains a set of common midpoint (CMP) gathers which were preprocessed to have multiples removed for times less than 2 seconds (s) using the parabolic Radon transform. We apply the DMO method from Fowler (1988) to generate a series of constant-velocity stacks from which the stack power, or envelope, is computed. This will serve as $\alpha[v, \mathbf{x}]$. A lower mute is applied to suppress high envelope values corresponding to multiples below 2 s. Triangle smoothing and scaling, featuring smoothing radii ranging in strength from $r_v = 5$, $r_t = 30$, $r_x = 15$ to $r_v = 50$, $r_t = 300$, $r_x = 150$ and scaling factors ranging from 1 to 10, are applied to the muted volume for use in continuation. The least-smoothed version is shown in Figure 1b. We generate 10 volumes of increasingly aggressive smoothing and scaling and apply the ℓ-BFGS variational velocity picking algorithm with continuation using those volumes. The picking algorithm is allowed to proceed for a maximum of 20 iterations at each continuation level. The initial continuation level is shown in Figure 2a. The initial model shown in solid black is a linear $v(t)$ spanning the range of scanned velocities and record times. The final model of this continuation level is shown in dashed white. The sixth continuation level is shown in Figure 2b, the ninth in Figure 2c, and the tenth and final level in Figure 2d. In each of these panels, the solid black line represents the starting model for the continuation level, which was the final model of the previous one, and the dashed white line is the final model. Figure 2d also features the linear initial model in solid red, which is the same as the starting model in Figure 2a, and the picking output without continuation in dashed green. This model is generated applying the ℓ-BFGS picking approach only on this level of smoothing, again with a maximum of 20 iterations. This non-continuation final model is shown in Figure 3a, and the final model with continuation is displayed in Figure 3b. To further illustrate how the two final models track stack power at different positions within the study area, overlays of the starting model (solid fuchsia), non-continuation final model (dashed black) and continuation final model (dashed white) are plotted on stack power at midpoints of 5 km in Figure 4a, 12 km in Figure 4b and 18.25 km in Figure 4c. The cost functional in Equation 4 is evaluated for each update of both methods using the least-smoothed stack power volume visible in Figure 1b.

Notice that the cost for each model never attains zero. This is because the cost defined by Equation 4 is related to the regularized, incremental surface area of the velocity model, $\sqrt{\lambda^2 + |\nabla v|^2}$ weighted by $e^{-\alpha(t,x,v)}$ at each model position with an additional total variation term $\frac{\epsilon}{2} |\nabla v|^2$. Here, $\alpha(t, x, v)$ is the stack power volume at zero-offset time t, midpoint x, and model velocity v. Since stack power is positive and bounded above by some $\bar{\alpha}$, $e^{-\bar{\alpha}} \leq e^{-\alpha(t,x,v)}$. The surface area term is always greater than or equal to λ, and the total variation term is greater than or equal to zero. Thus, the lowest cost a model could possibly attain is $\Delta t \, \Delta x \, \lambda \, e^{-\bar{\alpha}}$ where Δt is the difference
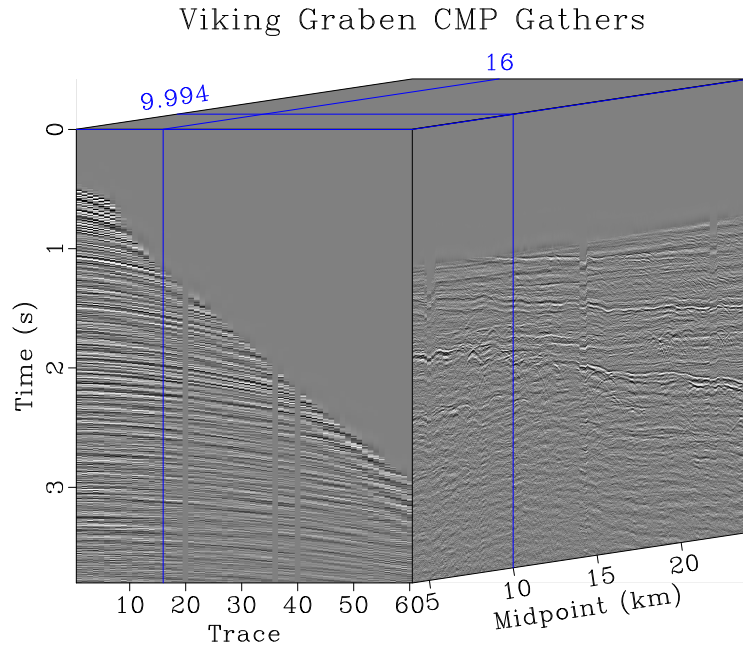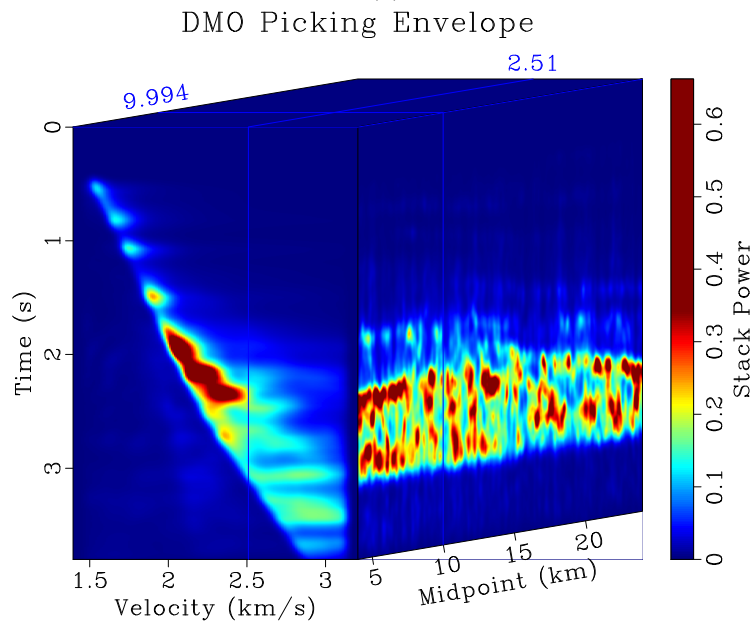
Figure 1: (a) Common midpoint gathers from the Viking Graben; (b) Constant velocity DMO stack power.

between maximum and minimum zero-offset time for the stack power volume and Δx is the difference between the maximum and minimum midpoint. For this example, that bounding minimal cost is roughly 40, but that would entail a constant velocity model through a stack power volume uniformly containing maximum stack power values which, on examination of Figure 1b, is clearly not the case. In effect, the cost function, and consequently the update at each iteration, seeks the model that balances tracking high stack power values with minimal model variation.
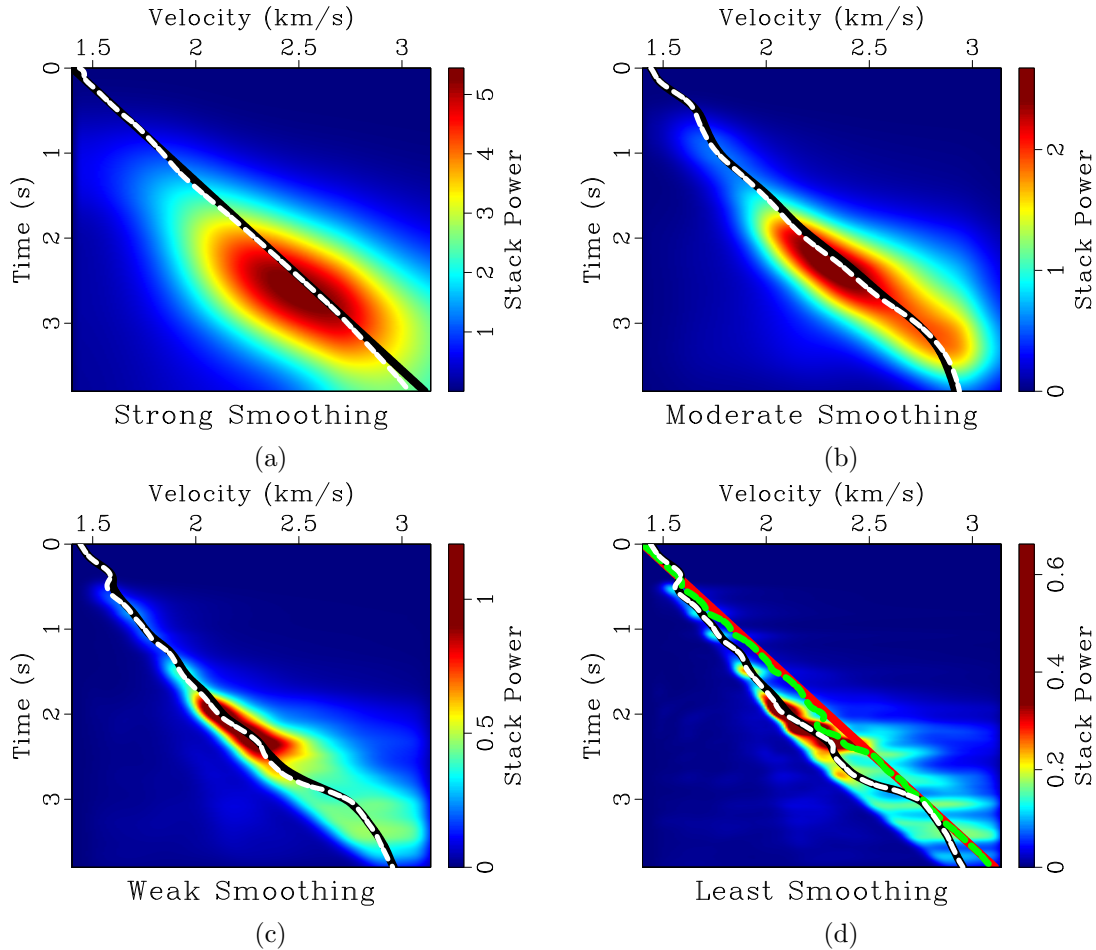
Figure 2: Illustration of continuation picking for midpoint at 9.994 km. Solid black line plots starting model for each level, dashed white line shows final model. (a) Initial continuation level with strongest smoothing; (b) Middle continuation level with moderate smoothing; (c) Second to last continuation level with weak smoothing; (d) Final continuation level also plotting the initial (solid red) and final models (dashed green) for picking without continuation.

We use the continuation velocity field to DMO stack the data, as shown in Figure 5a. Then we apply Kirchhoff migration to the data, as displayed in Figure 5b. The Dix formula (Dix, 1955) is used to compute interval velocity for the continuation model. This is a representation of local velocity through material at each region in the subsurface. Picked velocities are RMS velocities, corresponding to a mean velocity
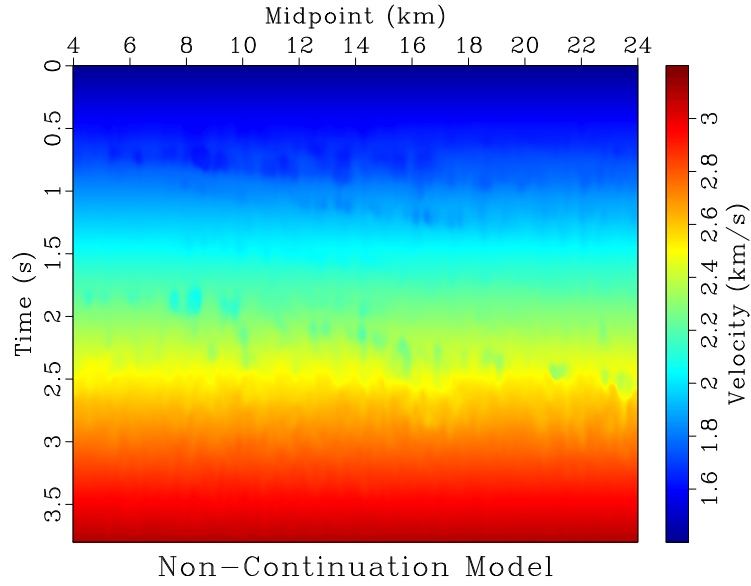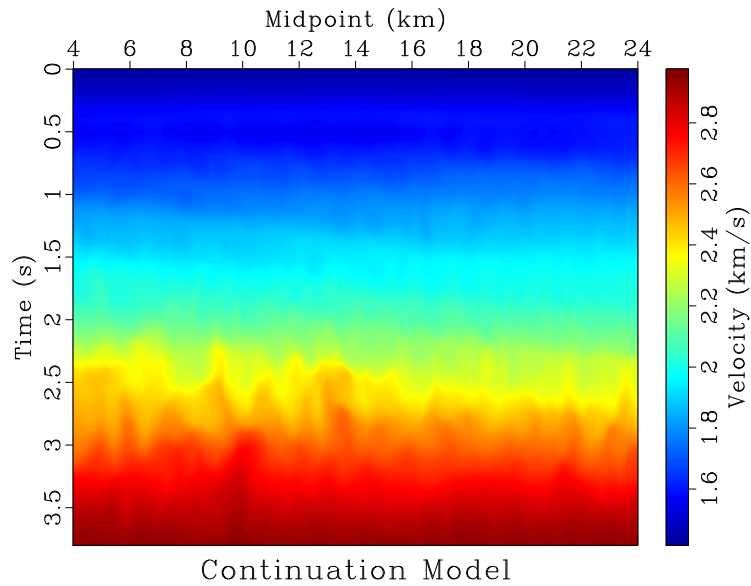
Figure 3: (a) Velocity model determined by variational picking algorithm without continuation; (b) Velocity model determined by variational velocity picking algorithm utilizing continuation.
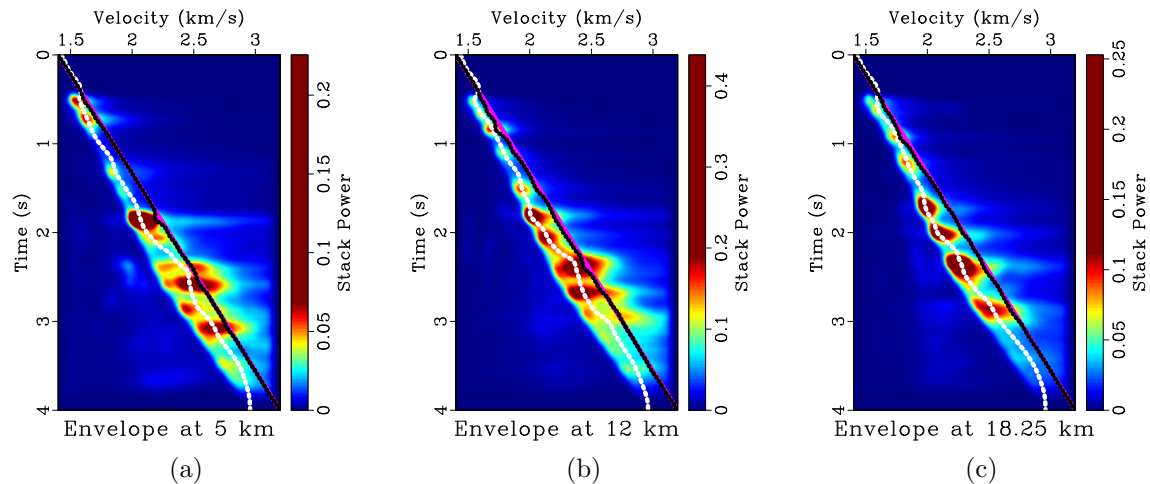
Figure 4: Illustration of the picking output at: (a) 5 km; (b) 12 km; (c) 18.25 km. Solid fuchsia line plots the starting model, dashed black line plots the non-continuation final model from Figure 3a, dashed white line plots the continuation final model from Figure 3b. Background plots the stack power with minimal smoothing shown in Figure 1b.

value over the complete ray path. The interval velocity is used for conversion from the time to the depth domain following the manner of Sripanich and Fomel (2018). The depth stretched interval velocity for the continuation model is plotted in Figure 6a and the corresponding depth stretched image in Figure 6b.

Notice that continuation behaves as a more global optimization scheme, enabling the solution to avoid local minima. In fact, on the first continuation level shown in Figure 2a, the cost actually increases when calculated on the least-smoothed stack power volume, as shown in Figure **??**. This enables the continuation velocity model to find the dominant trend where it deviates from the starting model, and produce a final model with a low associated cost. The model found without continuation is only able to follow stack power highs where the starting model is already near those highs, as can be seen around 2.25 s in Figure 2d, and in the three positions illustrated in Figures 4a, 4b, and 4c, where the black non-continuation final model primarily either overlies or is close to the initial model shown in fuchsia. The velocity field computed without continuation finds a local minimum with higher cost than the one found using continuation, as shown in Figure **??**.

The velocity field determined using continuation, Figure 3b, appears geologically plausible. Velocity values vary smoothly even though smoothness was not explicitly imposed. Furthermore, the velocity values in that model are in agreement with velocities computed from vertical seismic profile (VSP) travel time curves for two wells along the seismic line featured in Keys and Foster (1998). Seismic events in the DMO stack, Figure 5a, and seismic image, Figure 5b, generated using that model tend to be well-focused. Additionally, depth domain interval velocity traces for the continuation
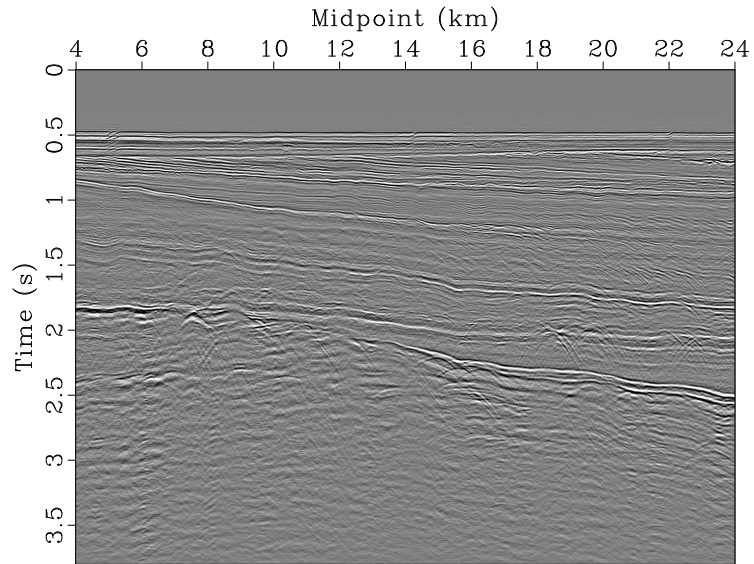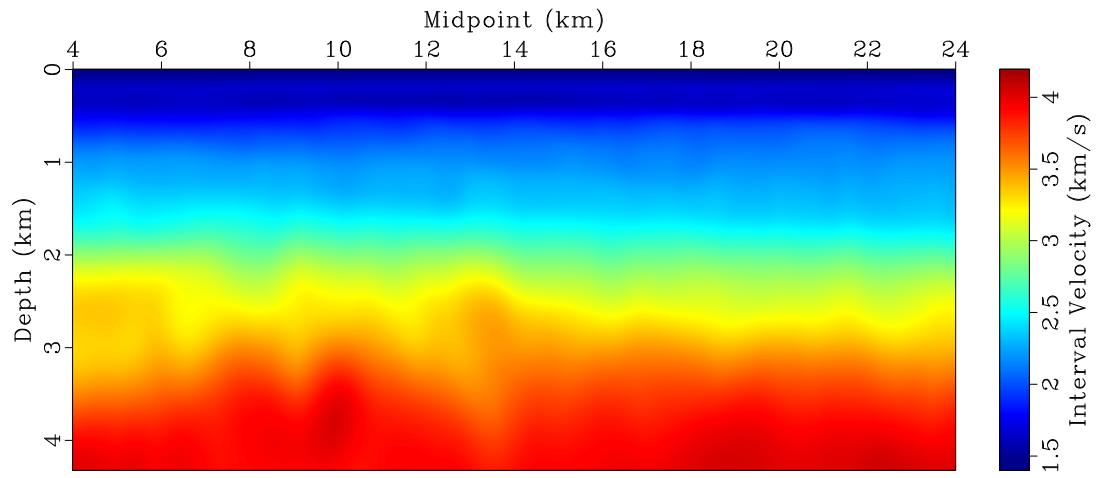
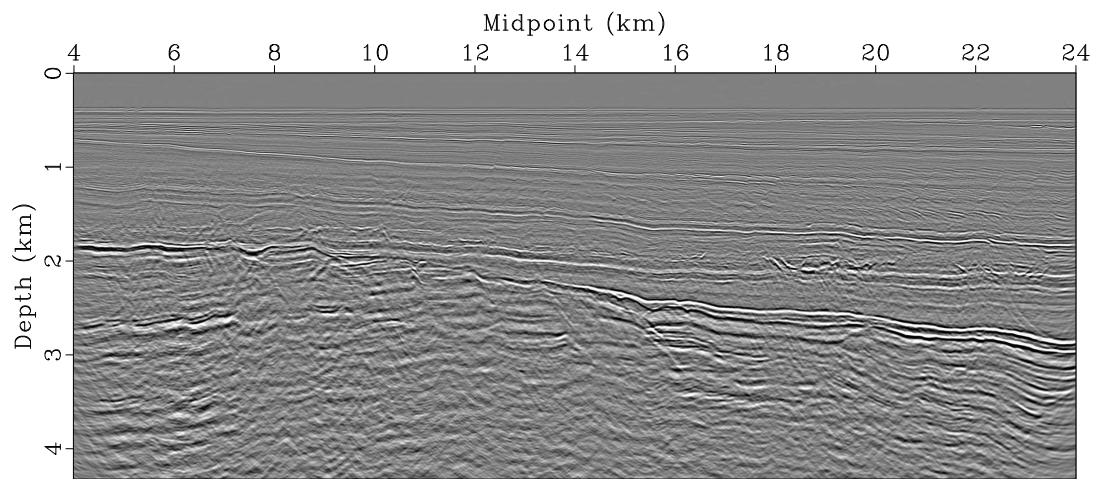Continuation Model DMO Stack

(a)
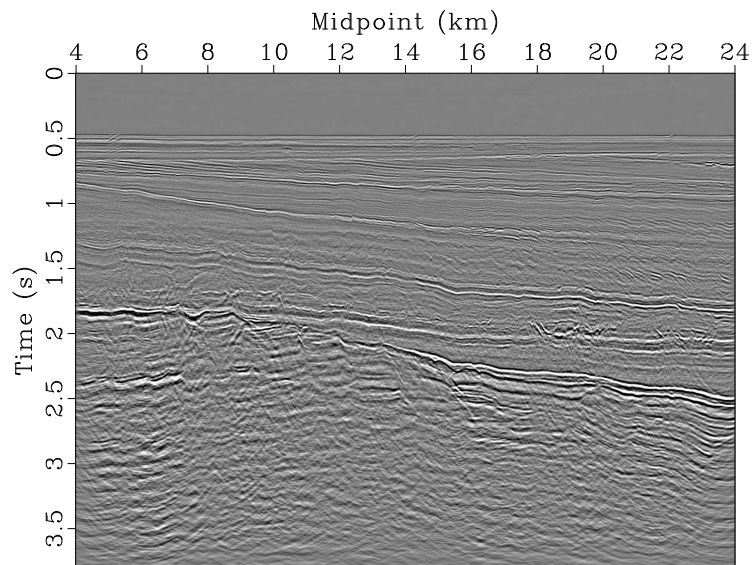
Continuation Model Image

(b)

Figure 5: (a) DMO stack of the data set in Figure 1a using the continuation velocity model from Figure 3b; (b) Kirchhoff time migrated images of the DMO stack in Figure 5a using the continuation velocity model in Figure 3b.

Continuation Model Interval Velocity in Depth

(a)

Continuation Model Image in Depth

(b)

Figure 6: (a) Interval velocity corresponding to continuation model in Figure 3b transformed to the depth domain; (b) Continuation model image in Figure 5b transformed to the depth domain.

model displayed in Figure 6a match the trend of sonic log p-wave velocities at the two well locations along the line (Keys and Foster, 1998).
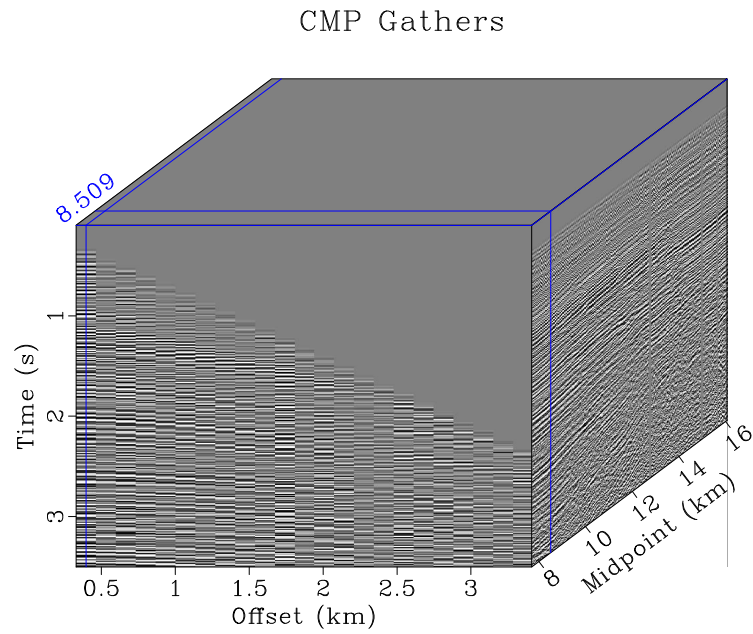
Gulf of Mexico

We now use a field data set from the Gulf of Mexico to illustrate how the continuation approach coupled with the iterative algorithm for velocity picking proposed here can act like a global minimizer, significantly reducing the dependence of the final model on the starting model. Figure 7a contains a set of CMP gathers. A NMO velocity scan is performed on the gathers to calculate semblance over a range of plausible NMO velocities and displayed in Figure 7b. Ten levels of increasingly aggressive triangle smoothing and scaling, again featuring smoothing radii ranging in strength from $r_v = 5$, $r_t = 30$, $r_x = 15$ to $r_v = 50$, $r_t = 300$, $r_x = 150$ and scaling factors ranging from 1 to 10, are applied to this NMO scan to create an $\alpha[v(\mathbf{x}), \mathbf{x}]$ for each of ten continuation levels.
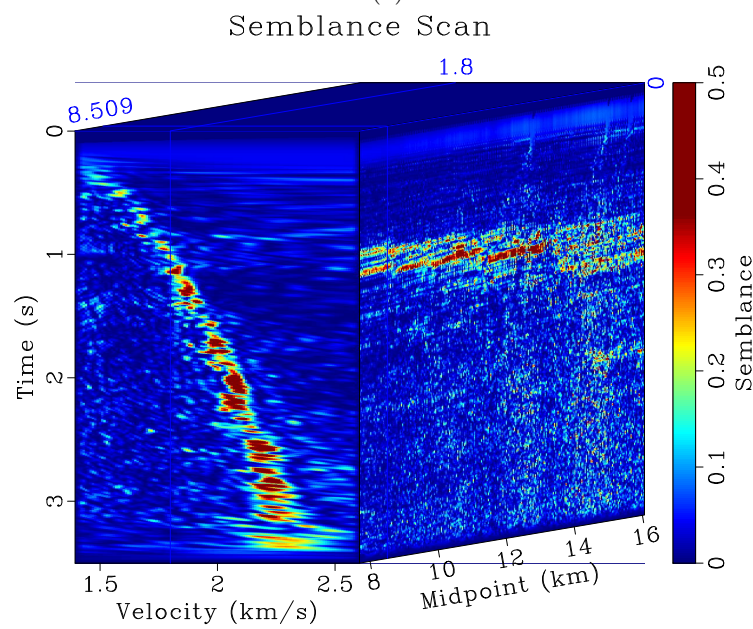
To illustrate how the continuation approach reduces the dependence of the final model on the starting model, we generate 125 constant-gradient velocity models which are clipped to the range of the velocity scan. A random selection of nine of these is shown in Figures 8a through 8i. These 125 starting models are used with the continuation picking approach over the ten smoothness levels and their cost convergence, or model cost by iteration according to Equation 4 as measured on the least-smoothed semblance volume in Figure 7b, is plotted in Figure 9a. If convergence is achieved before the maximum possible number of iterations, the final cost is repeated in these plots through the maximum iteration number. The lowest cost achieved is plotted as a dashed black line in that plot. For comparison, we also use the collection as starting models for iterative picking on the least smooth $\alpha[v(\mathbf{x}), \mathbf{x}]$ without continuation. That cost convergence is plotted in Figure 9b. Again, the lowest cost achieved by the continuation picking approach is plotted as a dashed black line and, if convergence is achieved before the maximum possible number of iterations, the final cost is repeated in these plots through the maximum iteration number.

The final continuation model with the lowest cost ($\tilde{G}(v) = 129.597$) is plotted in Figure 10a, and the continuation model with the highest cost ($\tilde{G}(v) = 129.636$) is shown in Figure 10b. The lowest cost ($\tilde{G}(v) = 130.278$) non-continuation final model can be seen in Figure 10c, and the highest cost ($\tilde{G}(v) = 143.778$) non-continuation final model is displayed in Figure 10d. To show how the continuation approach tends to produce similar final models, we compute the H^1 difference between the lowest cost continuation final model shown in Figure 10a and every update for every continuation model update and plot them in log scale in Figure 11a. If convergence is achieved before the maximum possible number of iterations, the final H^1 misfit is repeated in these plots thru the maximum iteration number. The H^1 norm of a function $u(\mathbf{x}) \in$

$H^1(\Omega)$, $\Omega \subset \mathbb{R}^n$ is defined as $\|u(\mathbf{x})\|_{H^1(\Omega)} = \sqrt{\int_\Omega \left(u(\mathbf{x})^2 + \sum_{i=1}^n \left(\frac{\partial}{\partial x_i} u(\mathbf{x}) \right)^2 \right) d\Omega}$,

CMP Gathers

(a)

Semblance Scan

(b)

Figure 7: (a) CMP gathers from a Gulf of Mexico field data set; (b) NMO velocity scan for the CMP gathers in Figure 7a.
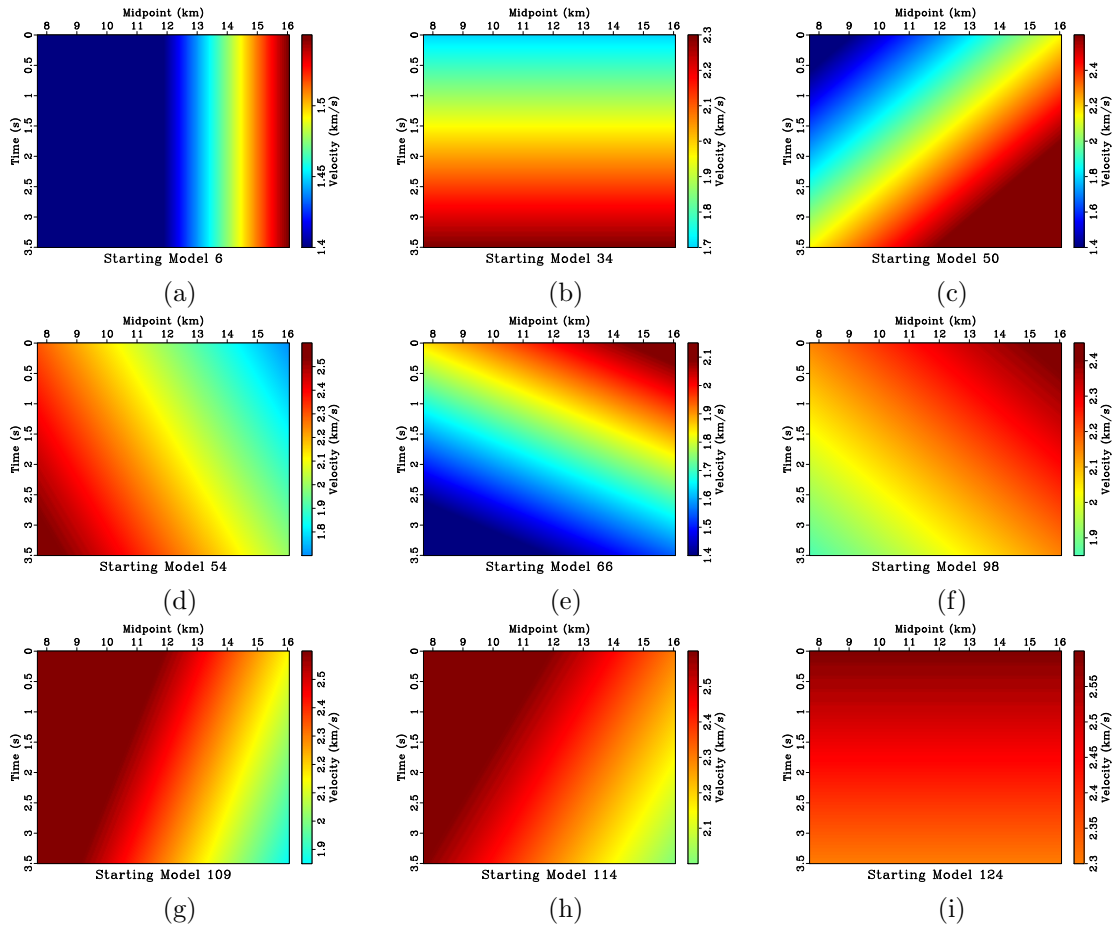
Figure 8: Random selection of nine constant-gradient velocity models from the 125 used as starting models.
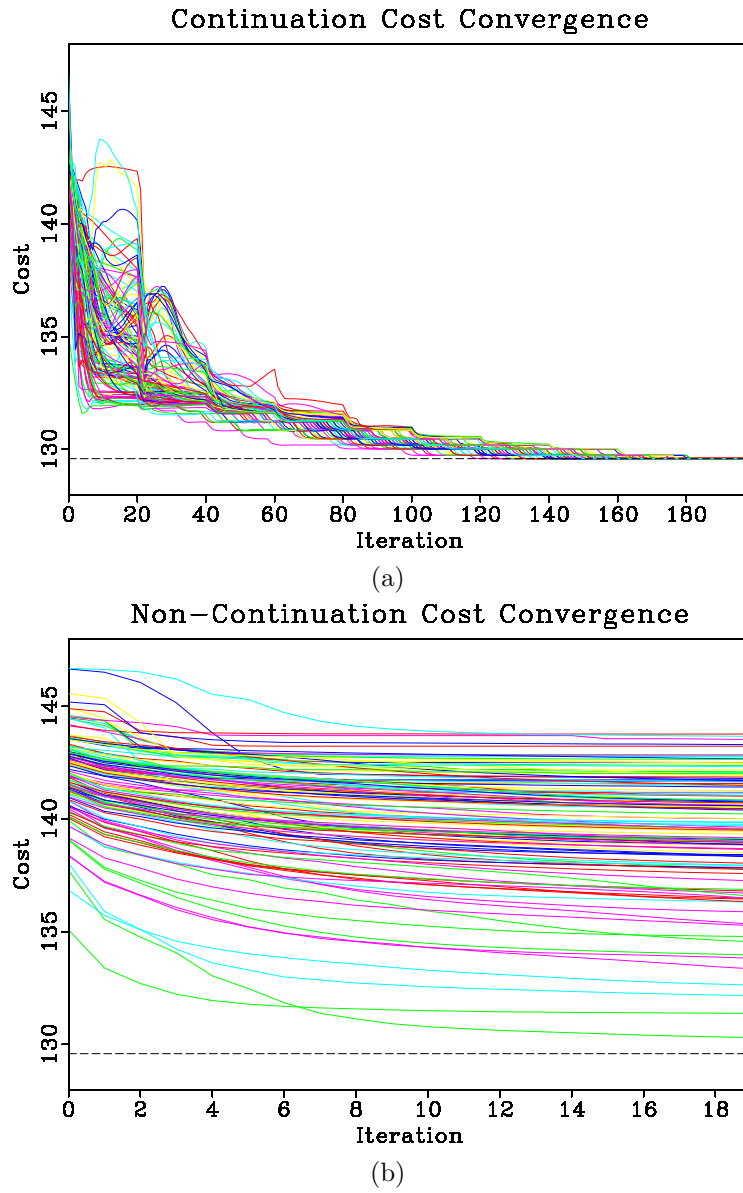
(a)

(b)

Figure 9: Cost convergence $\tilde{G}(v_i)$ for the 125 constant-gradient velocity models calculated on the least-smoothed semblance volume (a) using the continuation approach; (b) without continuation. The lowest cost achieved using continuation is plotted as a dashed black line.
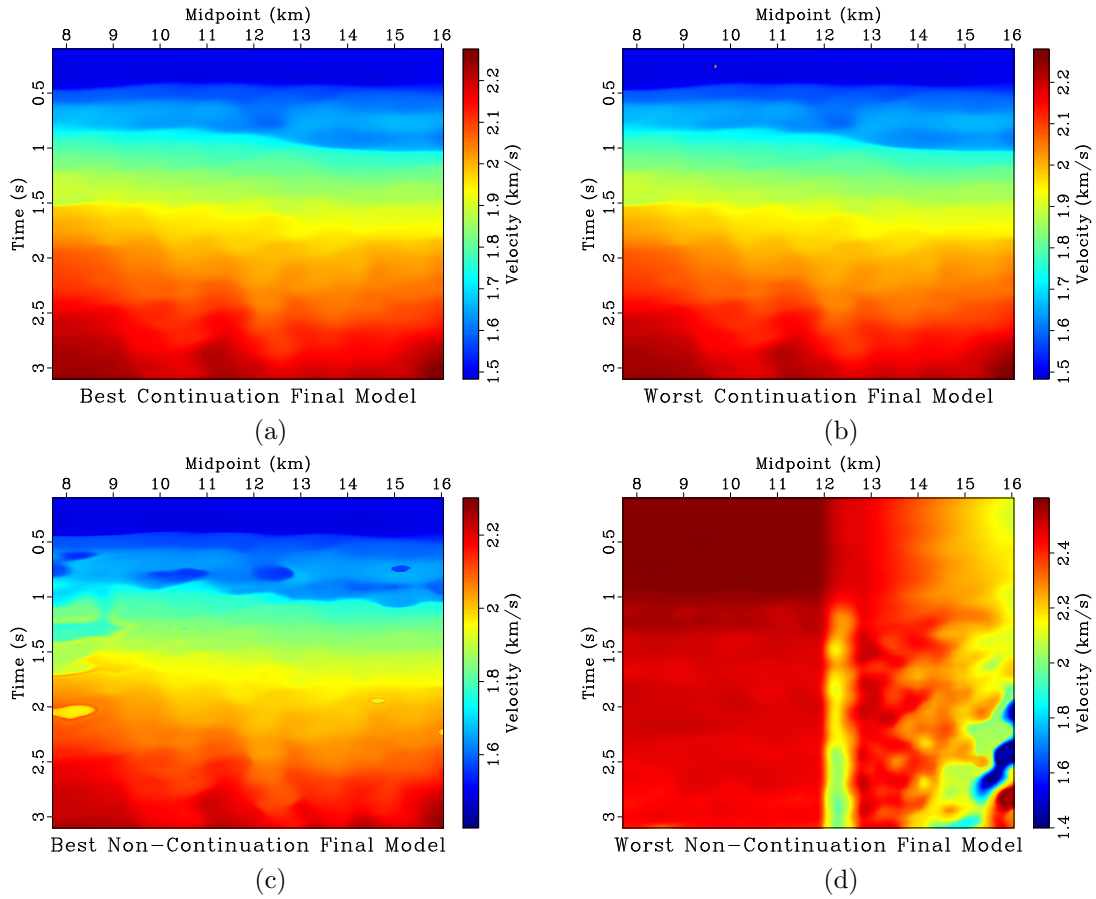
Figure 10: Lowest and highest cost final models from velocity picking using the 125 constant-gradient starting models: (a) lowest cost continuation final model; (b) highest cost continuation final model; (c) lowest cost non-continuation final model; (d) highest cost non-continuation final model.

and the H^1 difference or misfit of two functions $u, w \in H^1(\Omega)$ is $||u - w||_{H^1(\Omega)}$. This norm is used because it is the norm of the Hilbert space where minimizing surfaces of Equation 4 exist, and the norm in which convergence to minimizers should occur. The mean H^1 difference between the lowest-cost continuation final model and each of the 125 continuation final models is displayed in Figure 11b. These visualizations are also generated for log scale misfit between the lowest-cost continuation final model and each non-continuation update, as shown in Figure 11c, and the mean H^1 misfit between non-continuation final models and the lowest-cost continuation final model in Figure 11d. Notice that in the continuation H^1 convergence plot shown in Figure 11a, the lowest-cost model can be seen in red descending to an isolated end point around -11 near iteration 145. This is because, in the next iteration, it achieves the lowest cost final model, and thus the H^1 misfit is zero, the log of which is undefined.

The lowest cost continuation velocity model, which is subsequently referred to as the "best model," is used for seismic processing of this data set. The left panel of Figure 12 visualizes the best model overlaid on the semblance scan for a CMP gather at 8.5 km. The right panel illustrates that gather after NMO correction with the best model has been applied. Figure 13a contains stacked data following NMO correction using the best model. Diffraction data are extracted from that stack by using plane-wave destruction filters (Fomel, 2002; Fomel et al., 2007; Decker et al., 2013) to predict reflection signal and remove it from zero-offset data, leaving the set of diffractions and noise, which is displayed in Figure 13b. The best model is used for Kirchhoff time migration on the NMO stack data, whose image is shown in Figure 14a, and on the diffraction data to create the diffraction image in Figure 14b.

The continuation approach to picking velocities is able to overcome local minima and generate final models of similar cost and appearance relative to those created without continuation. As can be seen in the cost convergence plots for the continuation approach, Figure 9a, and the cost convergence plot without continuation, Figure 9b, the continuation approach leads to uniformly low-cost models relative to the approach without continuation. Indeed, the lowest-cost model achieved with continuation possesses a cost of $\tilde{G}(v) = 129.597$ while the highest cost model possesses a cost of $\tilde{G}(v) = 129.636$. These costs are significantly lower than the starting model costs for the 125 constant-gradient models, which range from roughly $\tilde{G}(v_o) = 135$ to $\tilde{G}(v_o) = 148$. Examining Figure 9a, none of the final models from the 125 has a final cost that is visibly different from the lowest-cost model. Comparing this result to those of cost convergence without continuation, shown in Figure 9b, the lowest-cost non-continuation model has a higher cost ($\tilde{G}(v) = 130.278$) than the highest-cost continuation model. There is also a significantly larger spread in final costs.

The continuation approach tends to generate to models that appear similar. The lowest-cost continuation model, plotted in Figure 10a, and the highest-cost continuation model in Figure 10b are visibly the same. Areas where the continuation models tend to differ can be seen in the mean continuation final model H^1 misfit plot of Figure 11b. This appears to be primarily confined to two approximately horizontal features near 0.5 and 1 s, which coincide with areas where velocity changes rapidly in the
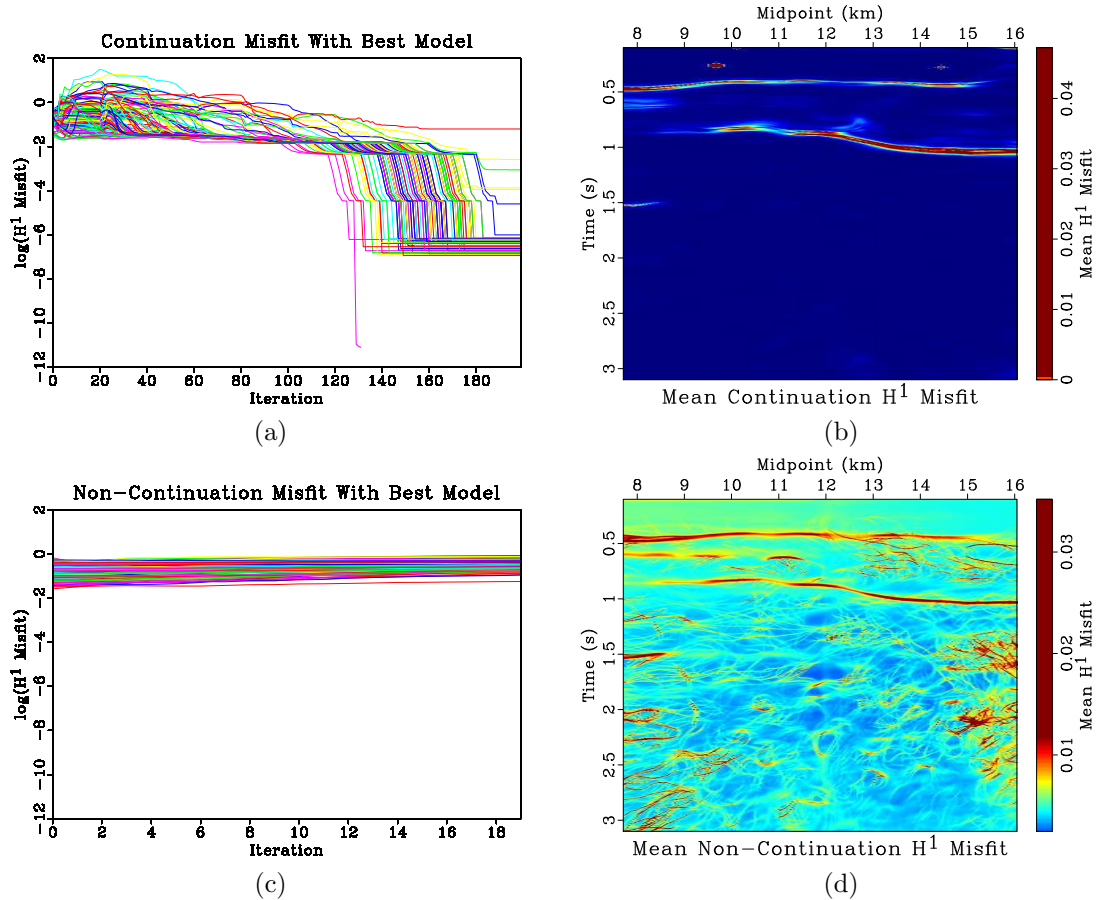
Figure 11: (a) log plot of H^1 misfit, or H^1 difference between the lowest-cost continuation final model shown in Figure 10a and continuation picking model updates for all 125 constant-gradient starting models; (b) mean H^1 difference between continuation picking final models and the lowest-cost continuation final model in Figure 10a; (c) log plot of H^1 difference between the lowest-cost continuation final model shown in Figure 10a and non-continuation picking model updates for all 125 constant-gradient starting models; (d) mean H^1 difference between non-continuation picking final models and the lowest-cost continuation final model in Figure 10a.
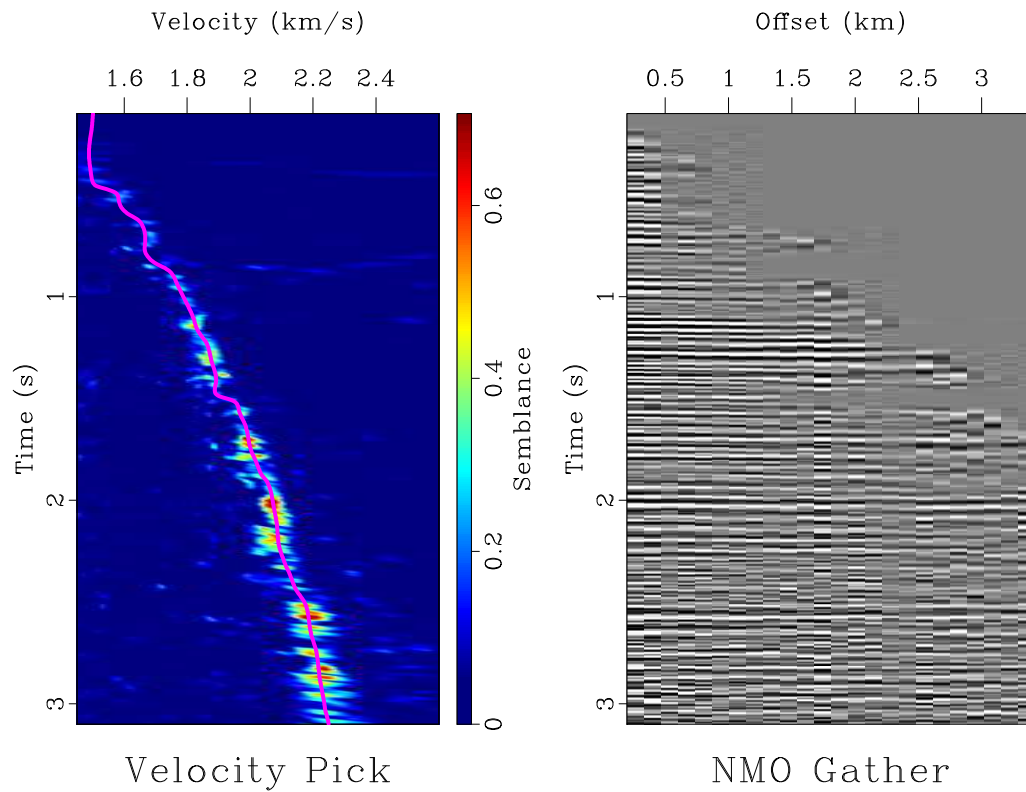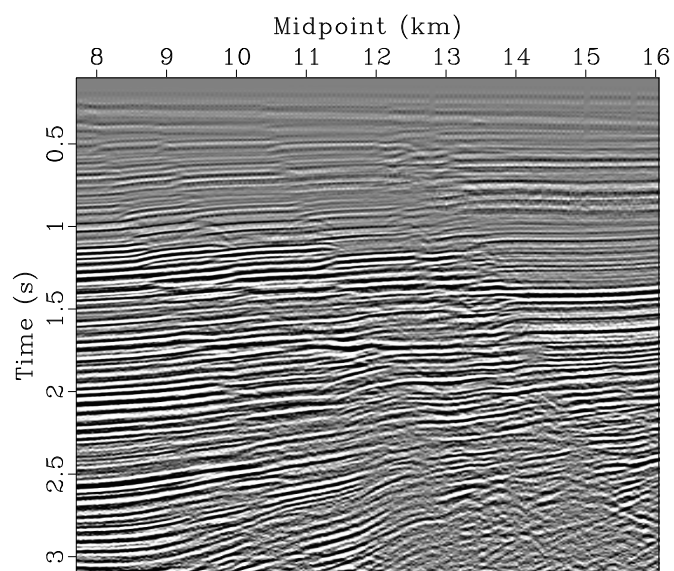
Figure 12: Illustration of the effects of the picking algorithm on a CMP centered at 8.5 km. Left panel contains a semblance scan for the CMP gather centered here overlaid by lowest cost final continuation velocity model from Figure 10a in solid fuchsia. Right panel contains the NMO corrected gather using that lowest cost final velocity model.

Best Model NMO Stack

(a)
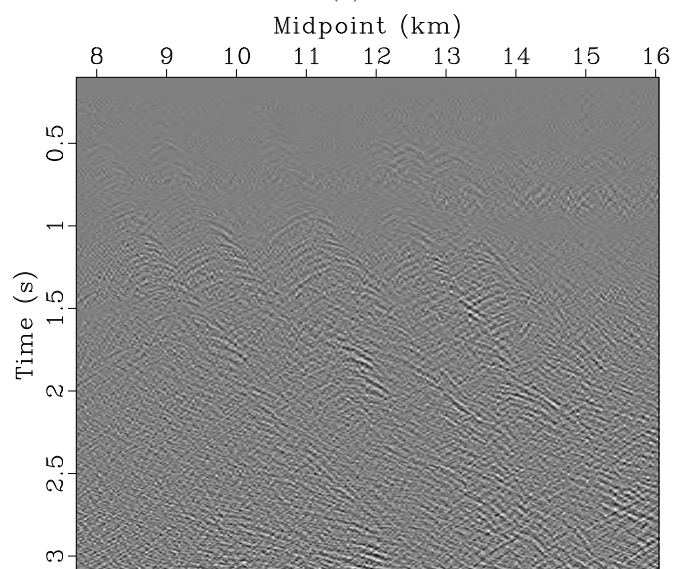
Best Model Separated Diffractions

(b)

Figure 13: (a) NMO corrected stack generated using the lowest cost continuation model shown in Figure 10a; (b) Diffraction data extracted from the NMO stack in Figure 13a using plane-wave destruction filters.

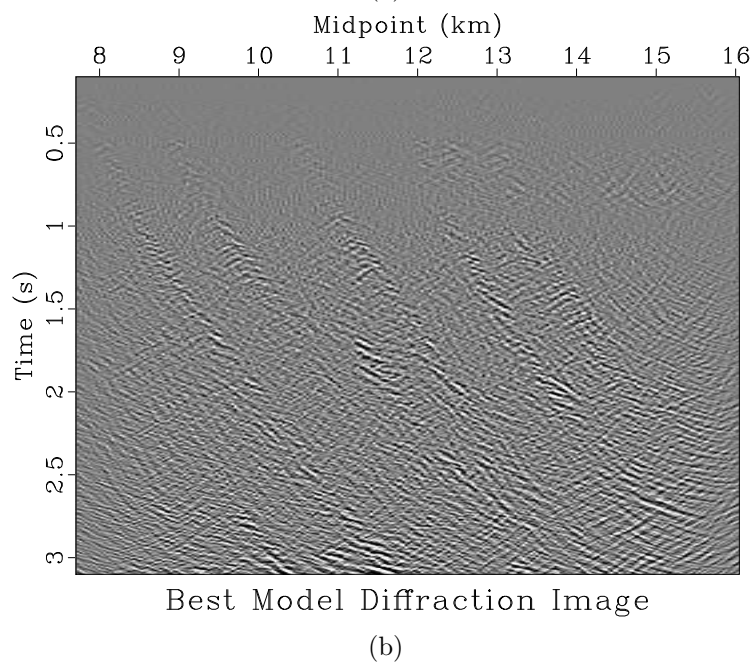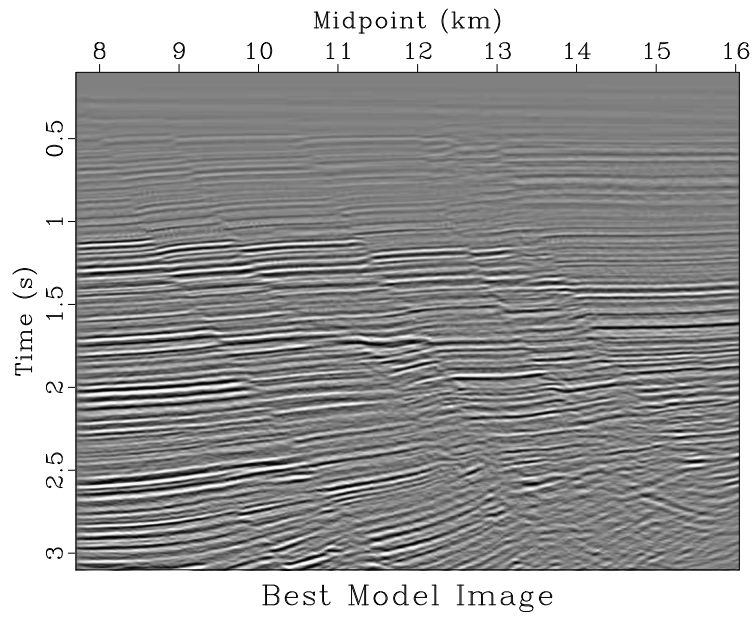Best Model Image

(a)

Best Model Diffraction Image

(b)

Figure 14: Kirchhoff time images generated using the lowest cost continuation model shown in Figure 10a corresponding to (a) the complete NMO stack data displayed in Figure 13a; (b) the diffraction data shown in Figure 13b.

lowest-cost continuation final model, Figure 10a. The lowest-cost non-continuation model in Figure 10c has a reasonably similar appearance to the lowest-cost continuation final model, although it possesses some anomalous "blobs" that are not geologically plausible. The highest-cost non-continuation final model in Figure 10d bears little, if any, resemblance to the lowest-cost continuation final model, and does not appear particularly geological or informative of the subsurface. Mean H^1 misfit for the non-continuation final models in Figure 11d tends to be significantly larger than that for final models determined using continuation. This plot has interesting dendritic features whose cause we do not understand, but may be related to the structure of local minima. Examining the H^1 convergence plot of the continuation approach, all but four of the 125 initial constant-gradient velocity models achieve misfits with the lowest-cost continuation model lower than 10^{-6} (note that seismic processing operations as well as these calculations were performed using single precision arithmetic, so some of this misfit may be due to rounding error). Most of these models follow paths that significantly decrease their H^1 misfit with the lowest-cost model as the number of iterations increase, causing them to move "toward" the lowest-cost model. Conversely, the evolution of H^1 misfit for non-continuation models has no obvious decreasing trend. Rather, the misfit of these models actually tends to increase, indicating they are moving "away" from the lowest-cost model. No non-continuation models possess misfit values less than 10^{-2}, and the smallest misfit is actually attained by a starting model.

The semblance scan overlaid by lowest cost continuation velocity in the left panel of Figure 12 shows a picked velocity that tracks the dominant trend in the semblance scan. The right panel of that figure shows the corresponding NMO corrected gather. Events in that gather are flat and laterally coherent, indicating that the picked velocity does a good job of performing NMO correction. The NMO stack, which is generated by applying the NMO correction corresponding to the lowest cost continuation velocity to all gathers and then summing over offset, is shown in Figure 13a. Energy in the stack is both focused and laterally coherent, again indicating that the NMO correction performs as intended. The diffraction data in Figure 13b appear as expected – most energy present in that data have the hyperbolic moveout associated with seismic diffraction. Migration using the lowest cost continuation picked velocity produces a complete image in Figure 14a that has well-defined, laterally coherent reflections that are often interrupted by discontinuities indicative of faulting. The diffraction features present in Figure 14b are collapsed to points, providing further confirmation of the quality of the picked velocities. These diffractions tend to delineate the faults which can be seen in the discontinuities of Figures 14a and 14b.

The lowest-cost final velocity model output by the continuation picking method produces quality complete and diffraction seismic images when used as part of a seismic processing workflow on the Gulf of Mexico field data set featured in this section.

Heidrun Field Horizon Picking

To illustrate the versatility of the proposed variational picking method, we apply it to an automatic horizon interpretation problem. Seismic horizons are laterally continuous features in seismic images representing isochrons, or constant levels of geologic time (Vail, 1977). Identifying and mapping seismic horizons is a key step in seismic interpretation, enabling earth scientists to delineate subsurface structures, stratigraphy, and the volumetrics of potential subsurface reservoirs (Wu and Hale, 2013). Because this activity is such an essential step in subsurface evaluation, much research has focused on developing computer algorithms to automatically accomplish the task (Lomask et al., 2006; Fomel, 2010; Hoyes and Cheret, 2011; Wu and Hale, 2015; Wu and Fomel, 2018b; Xue et al., 2018; Peters et al., 2019). The picking method presented here is not intended to replace this body of research, but rather to be seen as a tool that could work in tandem with it. One approach for automatic horizon interpretation involves outputting a volume containing the likelihood for a horizon to exist at each point in space. For example, Shi et al. (2020) use waveform embedding to extract seismic horizons using a deep convolutional network. The output of this deep convolutional network is a horizon probability volume, and the horizon is determined by selecting the positions with high probability values. The variational picking approach of this paper works well with advanced techniques which output semblance-like volumes, including probability volumes, used to determine a horizon surface. Because smoothness is not explicitly imposed with the variational approach proposed in this paper, it has the ability to track the rugose features frequently observed in seismic data.

The semblance-like volume for horizon likelihood employed here is less advanced than some of those appearing in the literature. In our approach, which is detailed in Appendix **??**, a trace representative of a seismic horizon is selected, cosine tapering is applied to its edges, and it is padded with zeros. To measure how well that ideal waveform matches with other traces throughout the seismic image volume, the crosscorrelation is calculated throughout the volume over shift τ. The minimum crosscorrelation value is subtracted so the output semblance-like volume is non-negative. Automatic picking may be performed on that volume to determine the shifts defining the horizon. Those shifts may be converted back to time in the image domain by adding a reference time for the reference horizon trace, taken here to be the trace's midpoint.

Figure 15a contains a seismic image from the Heidrun Field off the coast of Norway, and Figure 15b contains a zoomed portion of that image. We will focus on automatically picking the horizon defined by the "double white peak" in Figure 15b. A waveform representing an ideal version of that horizon is selected, and that horizon reference trace is plotted in Figure 16.

Crosscorrelation values between the horizon reference trace and the seismic image are computed over a variety of shifts according to Equations B-1 and B-2. The shift coordinate τ is transformed to image time t to be representative of horizon location
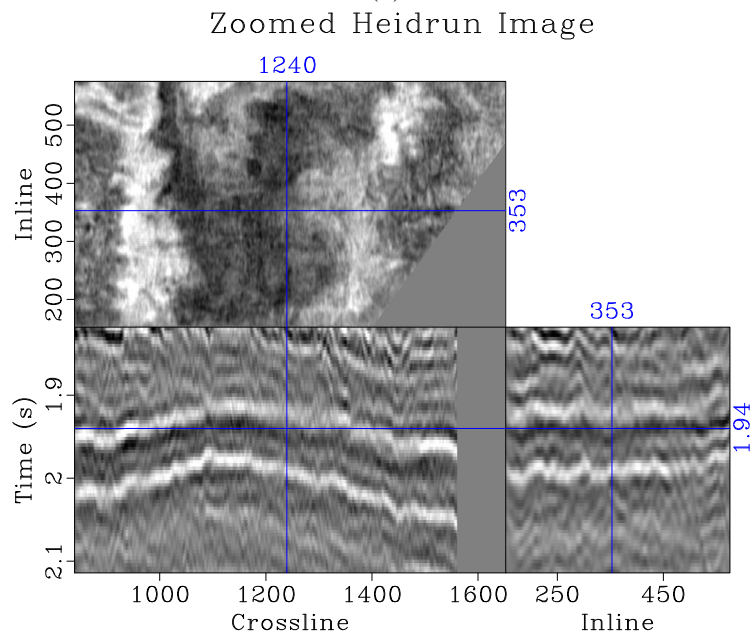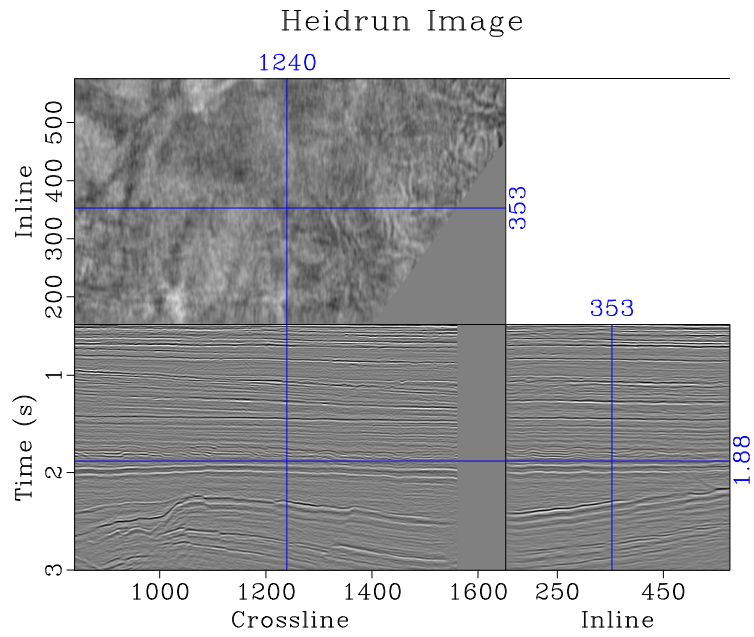
(a)

(b)

Figure 15: (a) Seismic image from the Heidrun Field; (b) zoomed portion of that image centered on the horizon that will be automatically picked.
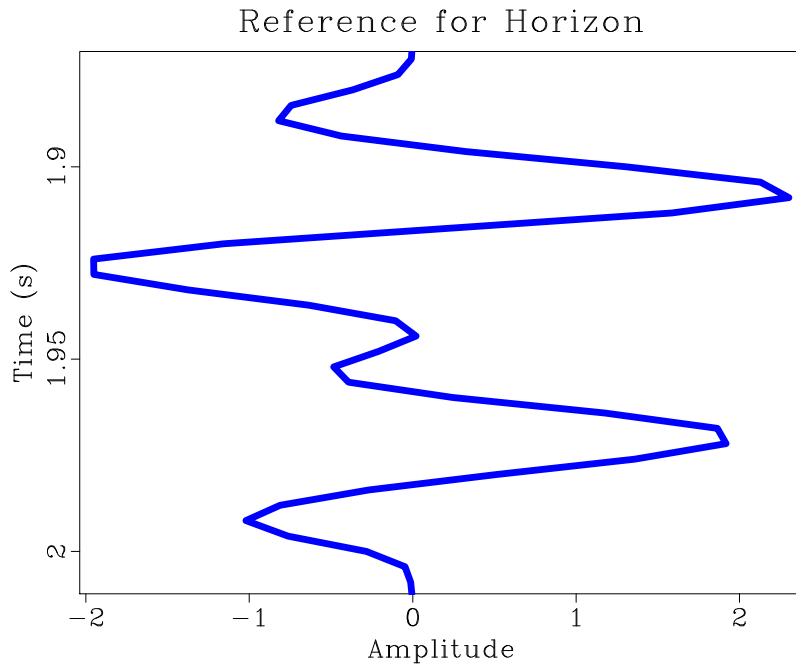
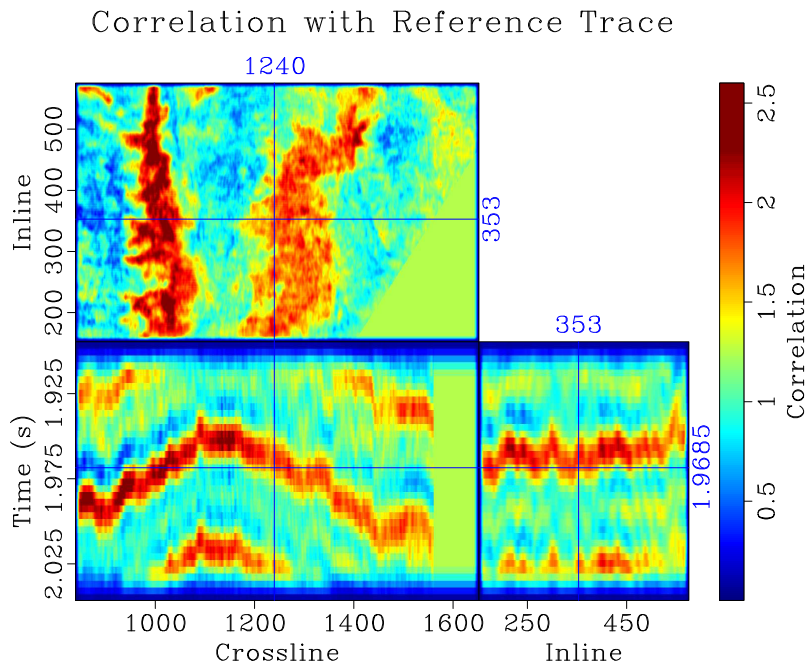Figure 16: Reference trace for horizon that will be picked from Figure 15b.

Figure 17: Cross correlation between reference trace and seismic image.

and displayed in Figure 17. Continuation picking is applied to the crosscorrelation volume using a flat starting model of 1.9745 s using 12 continuation levels. The evolution of the picked horizon every ten iterations, beginning with the starting model, is shown in Figures 18a through 18i. The final picked horizon is shown with constant horizon time contours in Figure 19. The cost convergence is calculated for each model update using the semblance-like volume in Figure 17 and displayed in Figure 20.
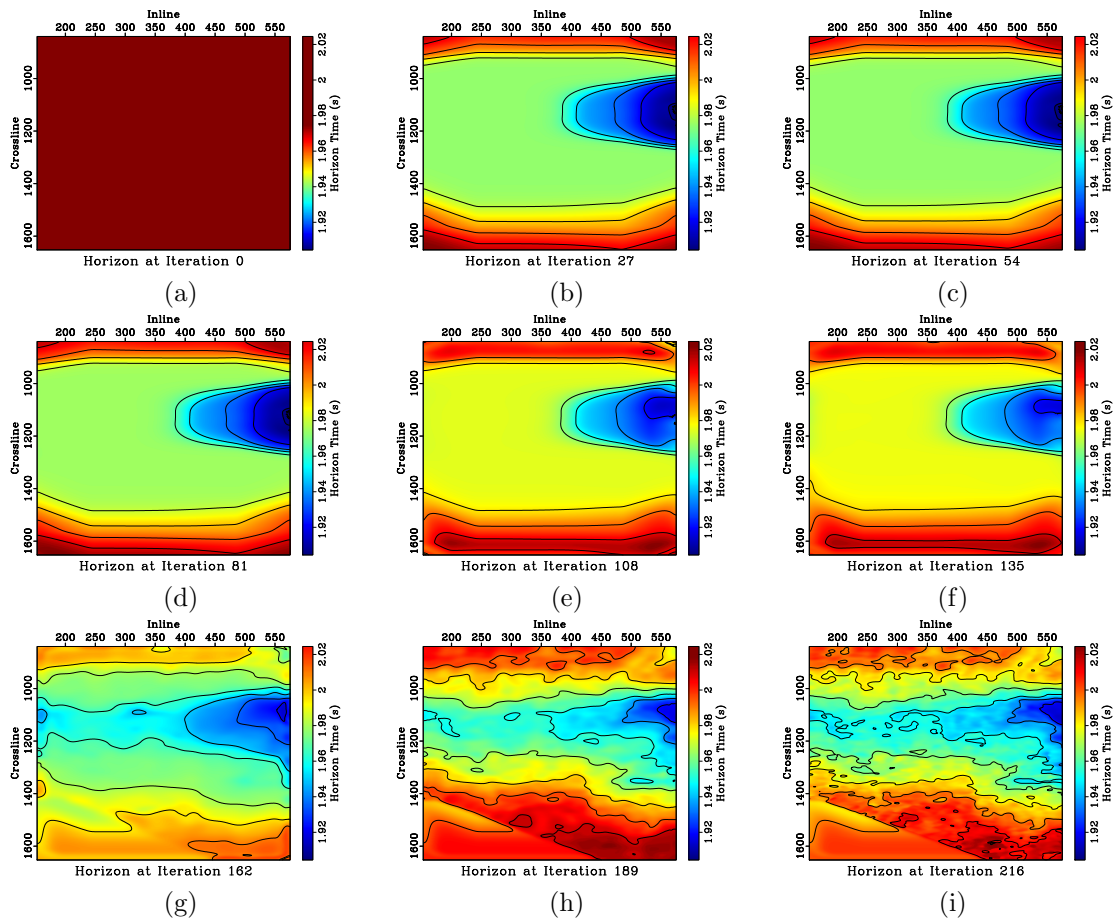
Figure 18: Evolution of continuation picking horizon overlaid by constant horizon time contours.

To illustrate the picked horizon in the seismic volume, we generate a series of overlay images. Note that the position of the horizon is taken to be at the halfway point of the reference trace in Figure 16, and thus we would expect it to approximately track the trough between the double white peaks in the image. Constant-inline slices are shown for inline 100 in Figure 21a and inline 300 in Figure 21b. Constant-crossline slices are similarly generated and shown for crossline 1200 in Figure 21c and crossline 1400 in Figure 21d.

The variational picking method outlined here appears to successfully pick the desired horizon from a poorly informed, flat starting model. Examining the picked horizon image overlays (Figures 21a through 21d) the horizon successfully tracks the

Figure 19: Final picked horizon with constant horizon time contours.

Figure 20: Cost convergence for continuation horizon picking using the semblance-like volume shown in Figure 17.
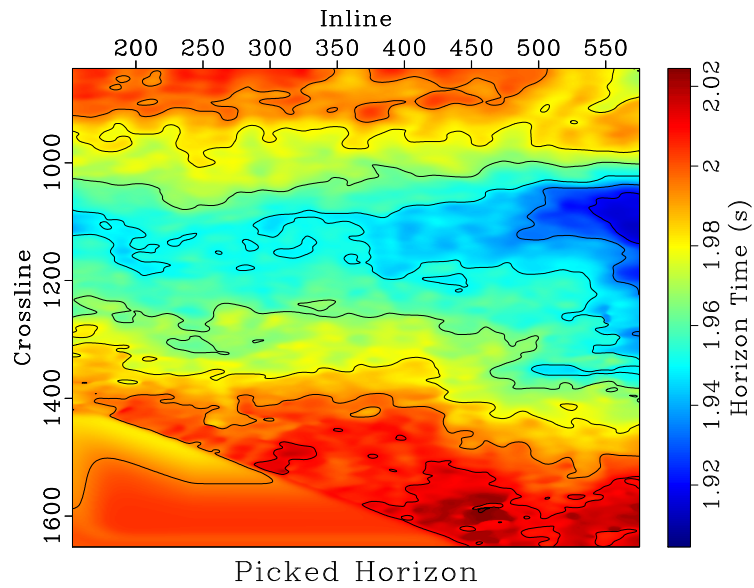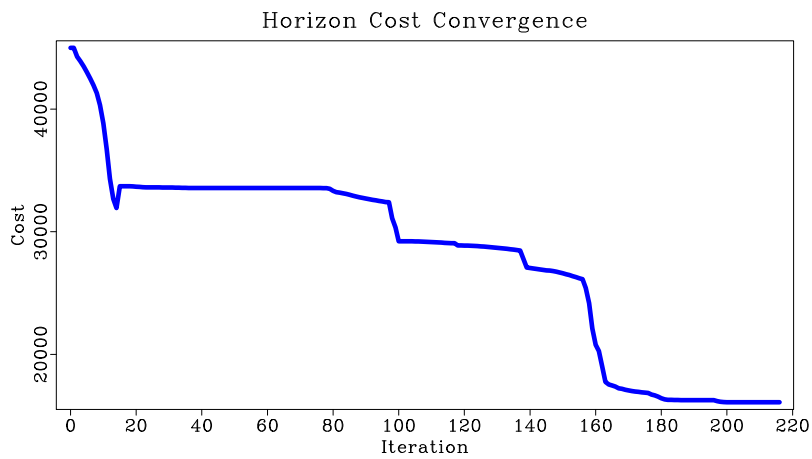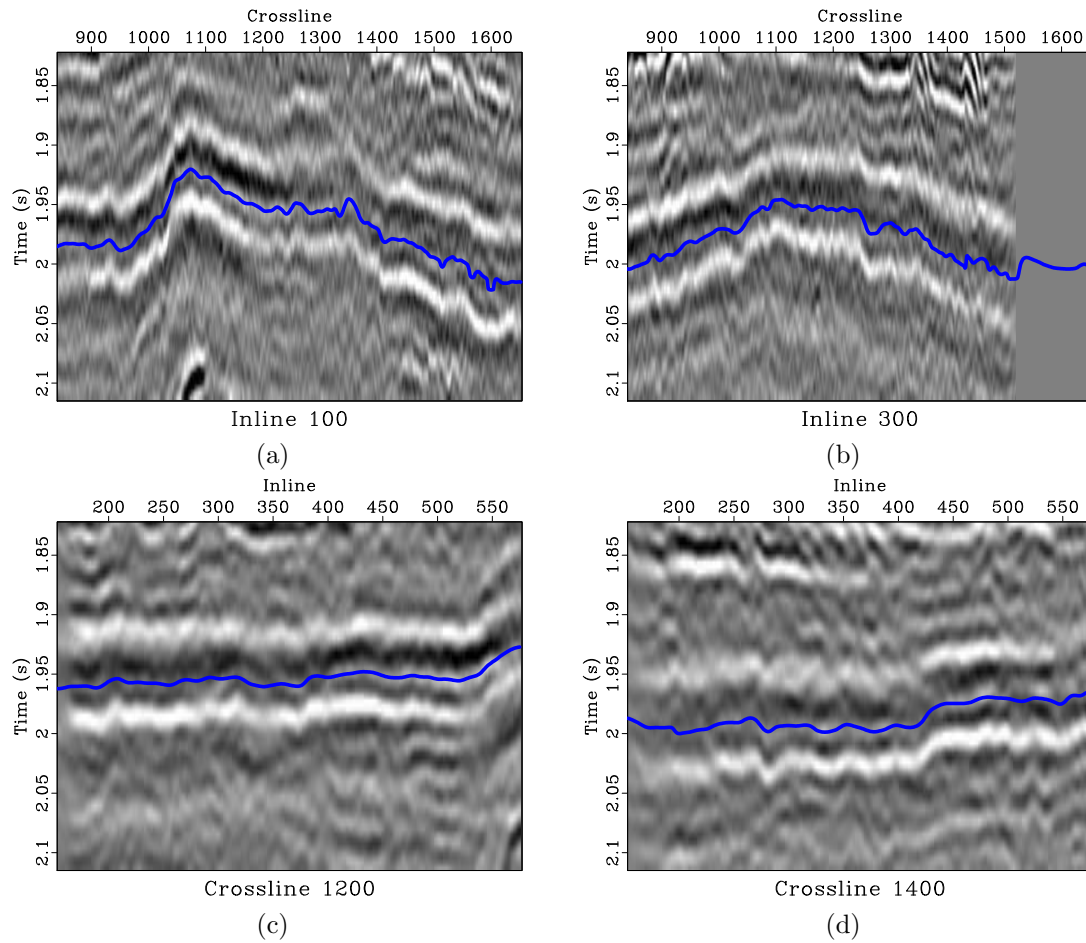
Figure 21: Picked horizon from Figure 19 overlaid on Heidrun seismic image from Figure 15b for: (a) Inline 100; (b) Inline 300; (c) Crossline 1200; (d) Crossline 1400.

trough between the two bright, white peaks, even when that position changes relatively rapidly, as is the case near crossline 1050 in Figure 21a, or when the character of the horizon wavelet changes somewhat, as is the case near inline 560 in Figure 21d. The display of the picked horizon in Figure 19 shows how the method is able to identify a plunging anticline structure in the horizon. The horizon-quality-of-fit volume defined by Equation B-2 is primitive, and may not be well-suited for the automatic interpretation of other horizons which are faulted or where there are significant changes in the wavelet. However, when applied to the horizon in this study, and used in conjunction with the continuation-variational picking method, it is able to produce a representation of the horizon from the single seed wavelet shown in Figure 16. When used with more sophisticated measures of horizon location probability, the variational approach outlined here could help produce high-quality automatic interpretations of subsurface features with minimal human intervention.

CONCLUSIONS

We propose a variational method for picking velocity surfaces from semblance-like volumes. When coupled with a continuation approach, this method is able to avoid many local minima. Using a discretized version of the variational statement with a ℓ-BFGS algorithm enables the method to converge more rapidly.

The variational method outlined here for determining optimal surfaces from semblance-like volumes is more computationally expensive than existing methods for determining optimal lines from semblance panels. However, the ability of the method to incorporate spatially adjacent information without explicitly imposing smoothing on the model justifies the added expense. Direct application of the method suffers from the multimodality and non-convexity of the proposed objective function and, unless a particularly well-informed starting model is used, the method is likely to converge to a local minimum. Such local minima may differ significantly from the global minimum. This difficulty may be overcome by applying the variational picking scheme with a continuation approach. Although continuation increases the cost of the method through the creation of additional smoothed semblance volumes and the application of the picking scheme to each smoothed volume, the ability of the method to avoid local minima and find a superior final model justifies the expense.

Applying the proposed approach to field data sets from the Viking Graben and the Gulf of Mexico shows how it is able to determine geologically plausible velocity models. The use of continuation both enables the scheme to behave more like a global minimizer, and allows models output by the approach to vary substantially from the starting model. Using the lowest-cost output model for each data set in seismic processing workflows produces quality images and further shows how the method may be incorporated as a tool for seismic processors.

The versatility of the variational method is demonstrated by deploying it to automatically pick a seismic horizon from the Heidrun Field. Because smoothing is not

explicitly imposed on the model, and is only used during the continuation process, the approach is able to determine a seismic horizon that changes rapidly in space to follow the reflector.

This method solves for minimizing surfaces in H^1, a space of smoothly varying functions. Minimizing surfaces are guaranteed to exist in this space because of the presence of a strictly positive ϵ in Equation 4, which is essential for demonstrating the existence of minimizers and proving that iterative schemes converge to those minimizers in an infinite-dimensional setting. Because RMS velocities are, in a sense, the integral of interval velocities, which exist in L^2, or equivalently are square integrable functions, all RMS velocities will exist in H^1. Therefore, this is not a limitation. However, if one is interested in solving for best-fit surfaces which are not smoothly varying but may feature discontinuities, they would want to set λ to a high value and ϵ to 0. Although having $\epsilon = 0$ violates the assumptions used in proving the existence of and convergence to minimizers of the variational method, in practice, minimizers are found to exist and can be iteratively determined.

The variational statement can be thought of as a modified soap bubble problem. It determines minimal area surfaces using a weight which rewards tracking highs in the semblance-like volume. If intervals within the semblance-like volume without high values exist, the method should output a minimal-area, or constant-gradient, surface connecting the high values bounding the region. This situation corresponds to intervals in reflection velocity scans without reflection energy. The method assumes that highs in the semblance-like volume correctly track the value of the parameter. Thus, the largest source of manual intervention in the examples shown in this paper comes from defining the functions to mute artificially high values from the semblance-like volumes where this assumption does not hold. Developing methods to automatically generate those mutes is a promising direction for future study.

Beyond the demonstrated applications shown here for processing 2D seismic lines and automatically interpreting seismic horizons, numerous promising directions for future study involving the application of this variational approach exist. Examples include using the method to generate starting models for full waveform inversion, extending the software to picking 3D velocity volumes from 4D semblance hypervolumes, calculating time-shifts to match time-lapse seismic volumes, or applying the method to other situations where one wishes to determine a laterally continuous surface. These extensions could be accomplished using the same variational statement featured in this paper. Modifying that variational statement could enable the creation of a method for simultaneously determining surfaces for multiple parameters, which would be useful for generating anisotropic models.

ACKNOWLEDGEMENTS

APPENDIX A: PSEUDOCODE FOR A PICKING ALGORITHM UTILIZING CONTINUATION

Algorithm 1 contains pseudocode which could be used to create a variational picking algorithm. Algorithm 2 shows how a variational picking algorithm may be used in a continuation framework. Here, ρ_o and ρ_f are strictly positive numbers denoting the minimum and maximum possible step sizes at each iteration, γ and ξ are positive early termination parameters, N is the maximum number of iterations at each continuation level, $v(\mathbf{x})$ is a velocity model, $v_0(\mathbf{x})$ is a starting velocity model, $g(\mathbf{x})$ is the functional gradient, and $h(\mathbf{x})$ is the search direction. M is the number of continuation levels, $\mathbf{r}_j$ is a vector holding the triangle smoothing radius for each spatial dimension at continuation level j, and σ_j is the semblance scaling factor at continuation level j. We assume that scaling and smoothing increase with increasing j. $\alpha_0[v, \mathbf{x}]$ is the least-smoothed semblance-like volume, and $\alpha[v, \mathbf{x}]$ denotes a semblance-like volume. These algorithms assume one has the following methods defined:

- COST$(v(\mathbf{x}), \alpha[v, \mathbf{x}], \lambda, \epsilon)$
  Returns the cost, $\tilde{G}[v]$ according to Equation 4, associated with model $v(\mathbf{x})$ for semblance-like volume $\alpha[v, \mathbf{x}]$ using regularization parameters λ and ϵ. The returned object is a scalar.

- GRADIENT$(v(\mathbf{x}), \alpha[v, \mathbf{x}], \lambda, \epsilon)$
  Returns the functional gradient, $\nabla \tilde{G}[v]$ according to Equation 5, associated with model $v(\mathbf{x})$ for semblance-like volume $\alpha[v, \mathbf{x}]$ using regularization parameters λ and ϵ. The returned object is a function with the same domain as $v(\mathbf{x})$.

- UPDATESEARCHDIRECTION$(g(\mathbf{x}), \mathbf{s})$
  Returns an updated search direction using gradient $g(\mathbf{x})$ according to a quasi-Newton scheme, for example ℓ-BFGS. In this method, $\mathbf{s}$ is an abstract state vector for the quasi-Newton scheme. The object $\mathbf{s}$ contains hyperparameters related to the method as well as gradient or search direction information from previous iterations which is used in the construction of the updated search direction. Application of the method is assumed to update abstract state $\mathbf{s}$ without returning it as an output. The returned object is a function with the same domain as $g(x)$.

- LINESEARCH$(h(\mathbf{x}), v(\mathbf{x}), \alpha[v, \mathbf{x}], \rho_o, \rho_f, \lambda, \gamma)$
  Returns the step size ρ between minimum and maximum step sizes ρ_o and ρ_f using a method of one's choosing such that

$$\rho = \underset{\rho \in [\rho_o, \rho_f]}{\arg \min} \text{COST}\left(v(\mathbf{x}) + \rho h(\mathbf{x}), \alpha[v, \mathbf{x}], \lambda, \epsilon\right). \quad \text{(A-1)}$$

The returned object is a scalar.

- SMOOTH($\alpha\left[v, \mathbf{x}\right], \mathbf{r}$)
  Applies triangle smoothing to $\alpha\left[v, \mathbf{x}\right]$ using smoothing radii for each dimension of α contained in $\mathbf{r}$. The returned object is a smoothed function with the same domain as $\alpha\left[v, \mathbf{x}\right]$

Algorithm 1 Variational Picking

1: **function** VARIATIONALPICK($v_0(\mathbf{x}), \alpha\left[v, \mathbf{x}\right], N, \rho_o, \rho_f, \lambda, \epsilon$)
2:     Initialize arrays $v(\mathbf{x})$, $g(\mathbf{x})$, $z(\mathbf{x})$
3:     Initialize search direction state vector $\mathbf{s}$
4:     $v(\mathbf{x}) = v_0(\mathbf{x})$ ▷ initialize model with starting model
5:     $c = $ COST($v(\mathbf{x}), \alpha\left[v, \mathbf{x}\right], \lambda, \epsilon$))   ▷ compute starting model cost
6:     **for** $i = 1 : N$ **do**   ▷ Iterate through maximum number of iterations
7:         $g(\mathbf{x}) = $ GRADIENT($v(\mathbf{x}), \alpha\left[v, \mathbf{x}\right], \lambda, \epsilon$)   ▷ compute gradient
8:         $h(\mathbf{x}) = $ UPDATESEARCH($g(\mathbf{x}), \mathbf{s}$)   ▷ update search direction and state vector $\mathbf{s}$
9:         $\rho = $LINESEARCH($h(\mathbf{x}), v(\mathbf{x}), \alpha\left[v, \mathbf{x}\right], \rho_o, \rho_f, \lambda, \epsilon$)   ▷ determine step size
10:         $\tilde{c} = $COST($v(\mathbf{x}) + \rho h(\mathbf{x}), \alpha\left[v, \mathbf{x}\right], \lambda, \epsilon$))   ▷ Cost of updated model
11:         **if** $\tilde{c} > c$ **then**
12:             **break**   ▷ increasing cost implies divergence, return model without update
13:         **end if**
14:         $v(\mathbf{x}) = v(\mathbf{x}) + \rho h(\mathbf{x})$ ▷ update model
15:         **if** $||\rho h(\mathbf{x})|| < \gamma$ or $c - \tilde{c} < \xi$ **then**
16:             **break**   ▷ sufficiently small model update or cost decrease implies convergence
17:         **end if**
18:         $c = \tilde{c}$ ▷ update cost
19:     **end for**
20:     **return** $v(\mathbf{x})$ ▷ return final model
21: **end function**

Algorithm 2 Continuation Picking

1: **function** CONTINUATIONPICK($v_0(\mathbf{x}), \alpha_0\left[v, \mathbf{x}\right], M, N, \{\sigma_j\}_{j=1}^{M}, \{\mathbf{r}_j\}_{j=1}^{M}, \rho_o, \rho_f, \lambda, \epsilon$)
2:     $v(\mathbf{x}) = v_0(\mathbf{x})$ ▷ initialize model with starting model
3:     **for** $j = M : 1$ **do**   ▷ iterate from highest continuation level to lowest
4:         $\alpha\left[v, \mathbf{x}\right] = \sigma_j$ SMOOTH ($\alpha_0\left[v, \mathbf{x}\right], \mathbf{r}_j$)   ▷ smooth and scale semblance
5:         $v(\mathbf{x}) = $ VARIATIONALPICK ($v(\mathbf{x}), \alpha\left[v, \mathbf{x}\right], N, \rho_o, \rho_f, \lambda, \epsilon$)   ▷ perform variational picking, update model
6:     **end for**
7:     **return** $v(\mathbf{x})$ ▷ return final model
8: **end function**

APPENDIX B: A PRIMITIVE, SEMBLANCE-LIKE VOLUME FOR HORIZON PICKING

Suppose we have a trace representative of a seismic horizon, apply cosine tapering to its edges, and pad it with zeros. This ideal horizon reference trace is called $h(t)$. To create a measure of how well this ideal waveform matches with other traces throughout a seismic volume, the crosscorrelation is calculated throughout the volume over shift τ,

$$\gamma(\tau, \mathbf{x}) = \int h(t + \tau) d(t, \mathbf{x}) d\tau, \tag{B-1}$$

where $d(t, \mathbf{x})$ is the seismic image. In order to ensure that the volume we generate is non-negative, let

$$\alpha(\tau, \mathbf{x}) = \gamma(\tau, \mathbf{x}) - \min_{(\tau, \mathbf{x})} \left[\gamma(\tau, \mathbf{x}) \right]. \tag{B-2}$$

Automatic picking may be performed on the semblance-like volume $\alpha(\tau, \mathbf{x})$ to determine the shifts defining the horizon corresponding to reference trace $h(t)$. Those shifts may be converted back to time in the image domain by adding a reference time for the ideal horizon trace, taken here to be that trace's midpoint.

This is a basic method for creating a horizon probability volume, which is intended for use in this paper as part of a proof of concept for automatic horizon interpretation using the variational picking algorithm. It could encounter difficulties in areas where significant lateral wavelet variation or faulting occurs.

REFERENCES

Alder, F., and S. Brandwood, 1999, Robust estimation of dense 3-D stacking velocities from automated picking: 69th Annual International Meeting: Society of Exploration Geophysicists Expanded Abstracts, 11621165.

Arnaud, J. D., J. P. Dunand, and V. Curinier, 2004, High density picking for accurate velocity and anisotropy determination: 74th Annual International Meeting: Society of Exploration Geophysicists Expanded Abstracts, 1627–1629.

Babich, V. M., and V. S. Buldyrev, 1972, Short-Wavelength Diffraction Theory Asymptotic Methods, 1 ed.: Springer-Verlag, volume **4** *of* Springer Series on Wave Phenomena.

Bader, S., X. Wu, and S. Fomel, 2019, Missing log data interpolation and semiautomatic seismic well ties using data matching techniques: Interpretation, **7**, T347–T361.

Blake, A., and A. Zisserman, 1987, Visual reconstruction.

Born, M., and E. Wolf, 1959, Principles of Optics: Pergamon Press Inc.

Candès, E., L. Demanet, and L. Ying, 2009, A fast butterfly algorithm for the computation of Fourier integral operators: Multiscale Modeling & Simulation, **7**, 1727–1750.

Chapelle, O., M. Chi, and A. Zien, 2006, A Continuation Method for Semi-Supervised SVMs: Proceedings of the 23rd International Conference on Machine Learning, Association for Computing Machinery, 185192.

Chaudhuri, S., and A. Solar-Lezama, 2011, Smoothing a program soundly and robustly: Proceedings of the 23rd International Conference on Computer Aided Verification, Springer-Verlag, 277292.

Claerbout, J., 1993, Earth soundings analysis: Processing versus inversion: Stanford University Press. Stanford Exploration Project.

Dacorogna, B., 2004, Introduction to the calculus of variations: Imperial College Press. EBL-Schweitzer.

de Bazelaire, E., 1988, Normal moveout revisited: Inhomogeneous media and curved interfaces: Geophysics, **53**, 143–157.

Decker, L., 2021, Parameter selection in sesimc data analysis problems: PhD thesis, The University of Texas at Austin.

Decker, L., and S. Fomel, 2018, A finite-element method for blind deconvolution with dynamic frequency wavelets: SEG Technical Program Expanded Abstracts 2018, 4563–4567.

——, 2021, A Probabilistic Approach to Seismic Diffraction Imaging: Lithosphere, **2021**. (6650633).

Decker, L., A. Klokov, and S. Fomel, 2013, Comparison of seismic diffraction imaging techniques: plane wave destruction versus apex destruction: 83rd Annual International Meeting, SEG, Expanded Abstracts, 4054–4059.

Decker, L., D. Merzlikin, and S. Fomel, 2017, Diffraction imaging and time–migration velocity analysis using oriented velocity continuation: Geophysics, **82**, no. 2, U25–U35.

Decker, L., and Q. Zhang, 2020, Quantifying and correcting residual azimuthal anisotropic moveout in image gathers using dynamic time warping: Geophysics, **85**, O71–O82.

Dennis, Jr., J. E., and J. J. Mor, 1977, Quasi-Newton methods, motivation and theory: SIAM Review, **19**, 46–89.

Deregowski, S., 1986, What is DMO?: First Break, **4**.

——, 1990, Common-offset migrations and velocity analysis: First Break, **8**.

Deschamps, T., and L. D. Cohen, 2001, Fast extraction of minimal paths in 3D images and applications to virtual endoscopy: Medical image analysis, **5**, 281–99.

Dix, C. H., 1955, Seismic velocities from surface measurements: Geophysics, **20**, 68–86.

Dobson, D. C., and C. R. Vogel, 1997, Convergence of an Iterative Method for Total Variation Denoising: SIAM Journal on Numerical Analysis, **34**, 1779–1791.

Doicin, D., C. Johnson, N. Hargreaves, and C. Perkins, 1994, Machine-guided velocity interpretation: SEG Technical Program Expanded Abstracts, 1413–1416.

Fomel, S., 2002, Applications of plane wave destruction filters: Geophysics, **67**, no. 6, 1946–1960.

——, 2003, Time migration velocity analysis by velocity continuation: Geophysics, **68**, no. 5, 1662–1672.

——, 2007a, Local seismic attributes: Geophysics, **72**, no. 3, A29–A33.

——, 2007b, Shaping regularization in geophysical-estimation problems: Geophysics, **72**, R29–R36.

——, 2009, Velocity analysis using AB semblance: Geophysical Prospecting, **57**, 311–321.

——, 2010, Predictive painting of 3d seismic volumes: Geophysics, **75**, A25–A30.

Fomel, S., E. Landa, and M. T. Taner, 2007, Post-stack velocity analysis by separation and imaging of seismic diffractions: Geophysics, **72**, U89–U94.

Fomel, S., P. Sava, I. Vlad, Y. Liu, and V. Bashkardin, 2013, Madagascar: open-source software project for multidimensional data analysis and reproducible computational experiments: Journal of Open Research Software, **1**, no. 1, e8.

Fowler, P., 1988, Seismic velocity estimation using prestack time migration: PhD thesis, Stanford University.

Gelfand, I. M., and S. V. Fomin, 2000, Calculus of variations: Dover Publications, Inc.

Greenberg, M. D., 1978, Foundations of Applied Mathematics: Dover Publications, Inc.

Hale, D., 1984, Dipmoveout by Fourier transform: Geophysics, **49**, 741–757.

——, 2013, Dynamic warping of seismic images: Geophysics, **78**, no. 2, S105–S115.

Harlan, W. S., 2001, Constrained automatic moveout picking from semblances: http://billharlan.com/pub/papers/autopick.pdf.

Hazan, E., K. Y. Levy, and S. Shalev-Shwartz, 2016, On Graduated Optimization for Stochastic Non-Convex Problems: Proceedings of The 33rd International Conference on Machine Learning, PMLR, 1833–1841.

Hoyes, J., and T. Cheret, 2011, A review of global interpretation methods for automated 3d horizon picking: The Leading Edge, **30**, 38–47.

Hu, J., S. Fomel, and L. Ying, 2015, A fast algorithm for 3D azimuthally anisotropic velocity scan: Geophysical Prospecting, **63**, 368–377.

Hubral, P., and T. Krey, 1980, Interval velocities from seismic reflection time measurements: Society of Exploration Geophysicists.

Iserles, A., 1996, A First Course in the Numerical Analysis of Differential Equations: Cambridge University Press.

Keys, R. G., and D. J. Foster, 1998, *in* 1. A Data Set for Evaluating and Comparing Seismic Inversion Methods: 1–12.

Kiefer, J., 1953, Sequential Minimax Search for a Maximum: Proceedings of the American Mathematical Society, **4**, 502–506.

Lanczos, C., 1966, The variational principles of mechanics: University of Toronto Press.

Larner, K., and V. Celis, 2007, Selective-correlation velocity analysis: Geophysics, **72**, U11–U19.

Li, D.-H., and M. Fukushima, 2001, A modified BFGS method and its global convergence in nonconvex minimization: Journal of Computational and Applied Mathematics, **129**, 15–35. (Nonlinear Programming and Variational Inequalities).

Li, J., and W. W. Symes, 2007, Interval velocity estimation via NMO–based differential semblance: Geophysics, **72**, U75–U88.

Liu, D. C., and J. Nocedal, 1989, On the limited memory BFGS method for large

scale optimization: Math. Program., **45**, 503528.

Lomask, J., A. Guitton, S. Fomel, J. Claerbout, and A. A. Valenciano, 2006, Flattening without picking: Geophysics, **71**, P13–P20.

Luenberger, D. G., and Y. Ye, 1984, Linear and nonlinear programming: Springer, **2**.

Mawhin, J., and M. Willem, 2010, Origin and evolution of the palaissmale condition in critical point theory: Journal of Fixed Point Theory and Applications, **7**, 265–290.

Mobahi, H., and J. W. Fisher, 2015, A theoretical analysis of optimization by gaussian continuation: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI Press, 1205–1211.

Mordecai, A., and D. J. Wilde, 1966, Optimality proof for the symmetric Fibonacci search technique: Fibonacci Quarterly, **4**, 265–269.

Mulder, W. A., and A. P. E. ten Kroode, 2002, Automatic velocity analysis by differential semblance optimization: Geophysics, **67**, 1184–1191.

Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: Mathematics of Computation, **35**, no. 151, 773–782.

Peters, B., J. Granek, and E. Haber, 2019, Multiresolution neural networks for tracking seismic horizons from few training images: Interpretation, **7**, SE201–SE213.

Sherwood, J. W. C., and P. H. Poe, 1972, Continuous velocity estimation and seismic wavelet processing: Geophysics, **37**, no. 5, 769–787.

Shi, Y., X. Wu, and S. Fomel, 2020, Waveform embedding: Automatic horizon picking with unsupervised deep learning: GEOPHYSICS, **85**, WA67–WA76.

Siliqi, R., D. L. Meur, F. Gramar, L. Smith, J. P. Touré, and P. Herrmann, 2003, High-density moveout parameter fields V and η. Part one: Simultaneous automatic picking: 73rd Annual International Meeting: Society of Exploration Geophysicists Expanded Abstracts, 20882091.

Smyrlis, G., and V. Zisis, 2004, Local convergence of the steepest descent method in Hilbert spaces: Journal of Mathematical Analysis and Applications, **300**, 436–453.

Sripanich, Y., and S. Fomel, 2018, Fast time-to-depth conversion and interval velocity estimation in the case of weak lateral variations: Geophysics, **83**, no. 3, S227–S235.

Symes, W. W., 1998, High frequency asymptotics, differential semblance, and velocity estimation: SEG Technical Program Expanded Abstracts 1998, 1616–1619.

——, 1999, All stationary points of differential semblance are asymptotic global minimizers: Layered acoustics: https://hdl.handle.net/1911/101915.

Symes, W. W., and J. J. Carazzone, 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, 654–663.

Taner, M. T., and F. Koehler, 1969, Velocity spectraDigital computer derivation applications of velocity functions: Geophysics, **34**, 859–881.

Tao, Y., M. Davidson, H. Swan, S. Fomel, J. Malloy, J. Howell, S. Chiu, and R. Olson, 2012, Constrained simultaneous automatic picking for VVAZ analysis: SEG Technical Program Expanded Abstracts, 1–5.

Toldi, J. L., 1989, Velocity analysis without picking: Geophysics, **54**, 191–199.

Vail, P. R., 1977, Seismic stratigraphy and global changes of sea level.: Bull. Am. Assoc. Petrol. Geol., Mem., **26**, 49–212.

Wu, X., and S. Fomel, 2018a, Automatic fault interpretation with optimal surface voting: Geophysics, **83**, O67–O82.

——, 2018b, Least-squares horizons with local slopes and multigrid correlations: Geophysics, **83**, IM29–IM40.

Wu, X., and D. Hale, 2013, Extracting horizons and sequence boundaries from 3d seismic images: SEG Technical Program Expanded Abstracts 2013, 1440–1445.

——, 2015, Horizon volumes with interpreted constraints: Geophysics, **80**, IM21–IM33.

Wu, Z., 1996, The effective energy transformation scheme as a special continuation approach to global optimization with application to molecular conformation: SIAM J. on Optimization, **6**, 748768.

Xue, Z., N. Alger, and S. Fomel, 2016, Full-waveform inversion using smoothing kernels: SEG Technical Program Expanded Abstracts 2016, 1358–1363.

Xue, Z., X. Wu, and S. Fomel, 2018, Predictive painting across faults: Interpretation, **6**, T449–T455.

Yan, S., and X. Wu, 2021, Seismic horizon extraction with dynamic programming: Geophysics, **86**, IM51–IM62.

Yilmaz, Ö., 2001, Seismic Data Analysis: Society of Exploration Geophysicists.