

## Horizon extraction using ordered clustering on a directed and colored graph

Zhining Liu<sup>1</sup>, Chengyun Song<sup>2</sup>, Kunhong Li<sup>3</sup>, Bin She<sup>1</sup>, Xingmiao Yao<sup>4</sup>, Feng Qian<sup>1</sup>, and Guangmin Hu<sup>5</sup>

### Abstract

Extracting horizons from a seismic image has been playing an important role in seismic interpretation. However, how to fully use global-level information contained in the seismic images such as the order of horizon sequences is not well-studied in existing works. To address this issue, we have developed a novel method based on a directed and colored graph, which encodes effective context information for horizon extraction. Following the commonly used framework, which generates horizon patches and then groups them into horizons, we first build a directed and colored graph by representing horizon patches as vertices. In addition, edges in the graph encode the relative spatial positions of horizon patches. This graph explicitly captures the geologic context, which guides the grouping of the horizon patches. Then, we conduct premerging to group horizon patches through matching some predefined subgraph patterns that are designed to capture some special spatial distributions of horizon patches. Finally, we have developed an ordered clustering method to group the rest of the horizon patches into horizons based on the pairwise similarities of horizon patches while preserving geologic reasonability. Experiments on real seismic data indicate that our method can outperform the autotracking approach solely based on the similarity of local waveforms and can correctly pick the horizons even across the fault without any crossing, which demonstrates the effectiveness of exploring the spatial information, i.e., the order of horizon sequences and special spatial distribution of horizon patches.

### Introduction

Horizon picking plays an important role in seismic interpretation because it can provide seismic interpreters with an understanding of the seismic data from a macro-perspective. Moreover, it is an essential step in solving many geophysical problems, such as seismic attribute extraction (Chopra and Marfurt, 2007; Liu et al., 2017), facies analysis (Song et al., 2017; Zhao et al., 2018), geophysical inversion (Wu, 2017), and generation of relative geologic time (RGT) volumes (Stark, 2003, 2004; Labruyere and Carn, 2015).

The above fundamental problem has attracted a surge of research interest. A common way is to follow the framework of seed-point-based autotracking, which is to extract horizons by correlation of local waveforms between neighboring seismic traces. After doing this trace

by trace, a complete horizon would be created. Therefore, lots of methods are devoted to designing a robust and effective metric used to measure the similarity of local waveforms so that the process of tracking can be propagated from the seed point. Fomel (2010) proposes to calculate local dips measured by the plane-wave-destruction algorithm and then spread the information contained in the seed trace inside the volume for horizon autotracking. Yu et al. (2011) generate an orientation vector field (OVF) from the seismic image using a bank of optical filters to guide the tracking. Khan and Alam (2012) use four sets of parameters, which are the dip vector, a local grid, amplitude, and semblance, to find the best picks on neighboring traces. Dossi et al. (2015) choose to use the cosine of the instantaneous phase and track events based on lateral phase continuity. Wu

<sup>1</sup>University of Electronic Science and Technology of China, School of Information and Communication Engineering, Chengdu 611731, China.  
E-mail: liuzhininguestc@gmail.com; bin.stepbystep@gmail.com; fengqian@uestc.edu.cn.

<sup>2</sup>Chongqing University of Technology, School of Computer Science and Engineering, Chongqing 400054, China. E-mail: scyer123@163.com.

<sup>3</sup>University of Electronic Science and Technology of China, School of Resources and Environment, Chengdu 611731, China. E-mail: moplushao@foxmail.com.

<sup>4</sup>University of Electronic Science and Technology of China, School of Resources and Environment, Chengdu 611731, China and University of Electronic Science and Technology of China, Center for Information Geoscience, Chengdu 611731, China. E-mail: yxm@uestc.edu.cn.

<sup>5</sup>University of Electronic Science and Technology of China, School of Information and Communication Engineering, Chengdu 611731, China, University of Electronic Science and Technology of China, School of Resources and Environment, Chengdu 611731, China, and University of Electronic Science and Technology of China, Center for Information Geoscience, Chengdu 611731, China. E-mail: hgm@uestc.edu.cn.

Manuscript received by the Editor 11 September 2018; revised manuscript received 15 June 2019; published ahead of production 16 September 2019; published online 08 November 2019. This paper appears in *Interpretation*, Vol. 8, No. 1 (February 2020); p. T1–T11, 7 FIGS., 4 TABLES.

<http://dx.doi.org/10.1190/INT-2018-0161.1>. © 2020 Society of Exploration Geophysicists and American Association of Petroleum Geologists. All rights reserved.

(2017) proposes a method called directional structure tensors to estimate reflection orientations in a new  $upq$  space, where  $u$  is the initial estimation of the reflection normal vectors and the other two orthogonal vectors  $p$  and  $q$  are those that are orthogonal to  $u$  and approximately aligned within the reflections. Wu and Fomel (2018) propose to combine local reflection slopes estimated using structure tensors and laterally multigrid slopes estimated using dynamic time warping to correlate seismic traces within multiple laterally coarse grids to guide the lateral tracking of the horizons. Targeting autotracking for weak seismic reflections, Di et al. (2018) develop a structural-orientation vector estimation algorithm to find the most likely locations in the neighboring traces. Different from these hand-crafted metrics, Harrigan et al. (1992) use the multilayer perceptron that is trained on several labeled traces to automatically extract the features of the targeted horizons. Furthermore, Leggett et al. (2003) use a hybrid artificial neural network that combines the self-organizing feature map and multilayer perceptron network paradigms. Instead of using the multilayer perceptron, Huang (1997) uses Hopfield neural network with multiple constraints to detect local horizon patterns. With the rapid development of deep learning, several works (Di, 2018; Lowell and Paton, 2018a, 2018b; Wu and Zhang, 2018) investigate its application for horizon interpretation and demonstrate its effectiveness.

However, even with carefully designed metric, horizon picking is still facing many difficulties. For example, when there are discontinuities present or chaotic regions with a low signal-to-noise ratio, this seed-point-based way would not work well for its limited locality. Therefore, recently, many researchers focus on introducing global-level information rather than locality within neighbors. An often-used choice is to track multiple horizons simultaneously (Hoyes and Cheret, 2011). An implicit assumption behind this technique is that one single horizon should covary with other horizons. Therefore, while extracting one horizon, other horizons would enforce additional constraints on the process of autotracking. For example, we should avoid a crossing of two horizons while interpreting horizons due to the principle of superposition (Monsen et al., 2007). With these constraints, ambiguities would be possibly cleared. Here, we consider this kind of information as the *context* in the process of horizon tracking. A natural way of representing the context of a horizon or horizon patch is in terms of its relationship to other objects. Plenty of methods have been developed to take full advantage of the context. Yu et al. (2011) propose to generate the OVF from the seismic image using a bank of optical filters to guide the pick selection, and based on the OVF, the minimum spanning tree is used for finding a complete horizon. Forte et al. (2016) present an attribute-based autotracking algorithm to mark all of the coherent events, which consists of two critical steps, patching (connect consecutive horizons) and grouping. Furthermore, Forte et al. (2016) design a different grouping strategy for sets of

consecutive horizons with different gap arrangements. Wu and Fomel (2018) compute a weighted least-squares fitting of features used for highlighting the target horizon. A context of smoothness that the horizons are locally laterally continuous can be easily introduced under this framework.

But, here, we urge that until now there has been a lack of an effective way to encode the rich and complex context and a unified framework to fuse all of the information for extracting horizons. In this paper, we also follow the patching-and-grouping framework (first, it generates horizon patches based on lateral continuity, and, then, it groups horizon patches into complete horizons), which has been used in several related works (Borgos et al., 2003; Qian et al., 2014; Forte et al., 2016; Lou and Zhang, 2018). In the stage of grouping, we choose to use a directed and colored graph to encode the context for each horizon patch. Peaks and troughs of waveforms are represented with different colors, and directed edges represent the relative spatial positions (over/under) between each two horizon patches. After building this graph, we consider the grouping of horizon patches as a problem of grouping nodes on the graph in two steps. First, we conduct premerging based on matching some predefined subgraph patterns. Then, we develop an ordered clustering method to group the rest of the horizon patches into horizons. By testing our method on real seismic data, we demonstrate that the context provided by the directed and colored graph is crucial in the grouping of horizon patches and that the ordered clustering method can successfully generate grouping results while preserving stratigraphic order well.

The remainder of the paper is organized as follows. First, we describe the details of the proposed method. Then, we test the method on real seismic data and show a promising result. Next, we discuss the advantages and weaknesses of our method before we finally conclude and describe the future work.

## Method

The proposed method follows the patching-and-grouping framework. Several other works also choose this for its simplicity and effectiveness. Figure 1 shows the workflow of our method. The following sections illustrate the details of each step.

### Patching

The goal of the patching procedure is to connect consecutive horizons points (e.g., peaks, troughs, and zero crossings) representing the same seismic event of interest. Usually, the patching generates a set of horizon patches that serve as the basis elements for subsequent processing (e.g., grouping). Most horizon tracking methods choose the local peak or trough that can be defined as the point of which the amplitude is larger or smaller than the neighboring points. A more advanced method is to use 3D curvature analysis (Di et al., 2017) through treating a waveform as a signal of convex and concave components to prepare these points. Moreover,

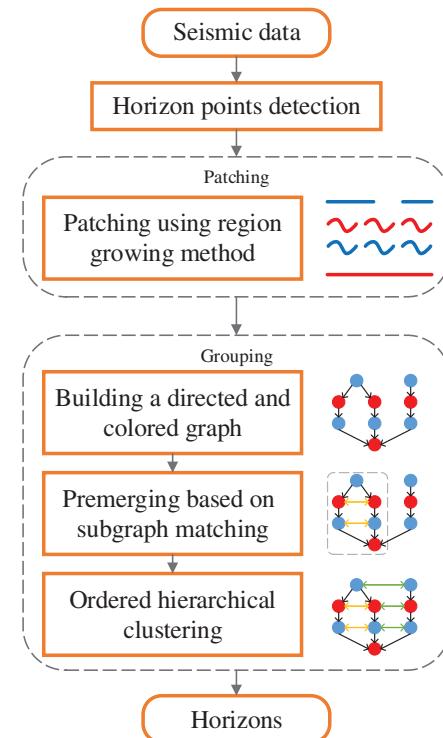
the peak-over-troughs and the trough-over-peaks can also be extracted using this method. After picking these horizon points, a clustering method is adopted to divide these points into several separate patches based on the lateral continuity. The basic assumption behind patching is that with the good lateral continuity of the horizons, these horizon points belonging to the same horizon patch would be connected to each other and distant from the horizon points consisting of another horizon (Qian et al., 2014). Therefore, we can easily generate several horizon patches using a clustering method based on the spatial distribution of horizon points. In this paper, we choose the region growing method (Adams and Bischof, 1994). First, we initialize all horizon points to the status “unvisited”. Starting with a randomly picked horizon point  $x$  and a list  $U$  that only contains  $x$ , we visit the vicinity of  $x$  and check whether there exists any other unvisited horizon point that is the same type as  $x$ . If so, we would consider these points and  $x$  as belonging to the same horizon patch, and add them except for  $x$  into  $U$  and remove  $x$  from  $U$  in the same time because  $x$  has been visited. Next, we repeat this growing procedure on the points in  $U$  until  $U$  is empty, which suggests that one horizon patch has been found. Then, we move on looking for other horizon patches on the rest of the horizon points with the status of unvisited. Algorithm 1 summarizes the method. By visiting all horizon points, we would be able to obtain several horizon patches that have clear gaps between every two of them because horizons points with spatial closeness would obviously be clustered into one horizon patch. Two hyperparameters are involved in this algorithm: the size of lateral window  $w$  and the threshold of the maximum time difference for growing  $\delta$ , which define the area (an 8-connected  $w \times \delta$  window in 2D or a 26-connected  $w \times w \times \delta$  window in 3D) for searching other horizon points around the target point. Here, we set the two hyperparameters to 3. With these patches, a grouping procedure would be required to form complete horizons.

### Grouping

After generating these horizon patches, we apply the grouping procedure on these patches to generate the final results. Here, we mainly consider the following four criterions for precise grouping:

- C1: Only the same type (peak or trough) of horizon patches can be grouped together.
- C2: After grouping, the newly formed horizon patch should not violate the over/under relationships of stratigraphic sequences. Here, the over/under relationships indi-

cate the relative geologic age of stratigraphic sequences, which are inferred based on the principle of superposition that if one sedimentary layer is stacked on the top of another sedimentary layer, then it must be younger than the one below (Monsen et al., 2007).



**Figure 1.** Workflow of the proposed method.

---

### Algorithm 1. Region growing method for patching.

---

**Input:** horizon points set  $H$ , lateral window size  $w$  and vertical window size  $\delta$

**Output:** horizon patches  $P$

```

1 Initialize an array  $P$ , of which size is the same with  $H$ , to unvisited;
2 Initialize a scalar  $patch\_index = 0$ ;
3 Initialize an empty list  $U$ ;
4 while  $\text{any}(P)$  is unvisited do % the loop for tracking all horizon patches
5   Randomly select a horizon point  $x$  that is unvisited from  $P$ ;
6    $U = [x]$ ;
7   while  $U$  is not empty do % the loop for tracking one horizon patch
8     Randomly select a horizon point  $y$  from  $U$ ;
9     Check neighboring horizon points within the 2D  $w \times \delta$  window or 3D  $w \times w \times \delta$  window centering the point  $y$ , and add all found unvisited horizon points that are the same type with  $y$  into  $U$ ;
10     $U.\text{delete}(y)$ ;
11     $P[y]=patch\_index$ 
12  end
13   $patch\_index+=1$ ;
14 end
15 return  $P$ ;

```

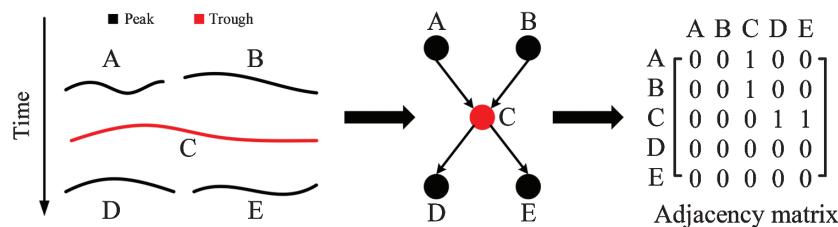
---

- C3: Horizon patches with high similarity on seismic waveforms are more likely to be grouped into the same horizon.
- C4: Horizon patches that form a special structure on the spatial distribution are more likely to be grouped into one horizon. Here, we only consider the scenario of parallel sets of consecutive horizons patches bounded by two laterally continuous horizons on the top and bottom, respectively, which suggests that these horizon patches should be grouped in a fixed way according to geologic parallelism.

C1 and C3 are commonly used criterions for guiding the grouping. However, there exist tremendous ambiguities in the process of grouping. How to take advantage of C2 and C4, which are actually powerful contexts for selecting the right patches for grouping, is the key to precisely tracking horizons. Of course, these four criteria are not always all satisfied. For example, when a small-scale fracture presents, it may fail to satisfy C3. But based on the other ones, a reasonable grouping result can also be inferred. Therefore, we first build a graph to encode the context of horizon patches from a macro perspective. Then, some horizon patches are premerged based on matching some predefined subgraphs on the graph. In this procedure, we group horizon patches based on the relative spatial distributions of a small set of horizon patches, which is a strong hint for grouping. Experimental results demonstrate that premerging indeed enhances the robustness of the method. Finally, an ordered hierarchical clustering method is adopted to group the rest of horizon patches based on a fused metric that incorporates the constraints given by stratigraphic sequences and waveform similarity. We illustrate the details of each step in the following subsections.

#### Graph building

A lot of independent horizon patches are generated in the procedure of patching, and most of them have not formed complete horizons except for those ones with good lateral continuity. Here, we build a directed and colored graph  $G$  based on these horizon patches to preserve all useful contexts within the seismic volume for



**Figure 2.** Example for building the directed and colored graph and inferring the relative spatial relationship. Each horizon patch is represented as a vertex in the graph, and different colors represent different types of horizon points (e.g., peaks and troughs). The over/under relationship of any pair of horizon patches specifies a directed edge between two vertices. With the graph, we can get the corresponding adjacency matrix based on which we can check whether one vertex can achieve another vertex using Algorithm 3.

the following grouping. A directed and colored graph is a graph that consists of a set of vertices connected by directional edges, and each vertex is assigned a color, where different colors represent different types. To construct the graph, each horizon patch is represented as a vertex in the graph and different colors of vertices are used to distinguish different types of horizon patches (e.g., peaks, troughs, and zero crossings). In this paper, only two colors are selected to represent the trough and peak, respectively. A directional edge  $(s, t)$  between two vertices  $s$  and  $t$  represents the patch denoted by  $s$  located on top of the patch denoted by  $t$  along the time growth direction. Figure 2 illustrates a simple example. Horizon patch A is located on top of horizon patch C. Therefore, an edge  $(A, C)$  is created on the graph. Similarly, for each two consecutive horizon patches along the time growth direction, an edge would be created from the former to the latter, which leads to the three edges  $(B, C)$ ,  $(C, D)$ , and  $(C, E)$ .

The initial relative spatial position is captured by vertically scanning along each trace. By visiting each sample point in increasing order of time, we can easily get a sequence that captures the relative positions of horizon patches by removing all of the sample points that do not belong to any horizon patch. For each two consecutive horizon patches, an edge would be created from the former to the latter. In practice, it is possible that some noisy horizon points would introduce errors. Therefore, a threshold  $n_0$  is set to remove these outliers. Only those edges that appear at least  $n_0$  times would be retained for the construction of the graph  $G$ , which indicates that an edge would be created if two consecutive horizon patches are overlapped over at least  $n_0$  seismic traces. Here, we set  $n_0$  to five. The procedure of scanning is explained in detail in Algorithm 2.

After scanning all of the traces using Algorithm 2, we can build a directed and colored graph (e.g., the graph shown in Figure 2). With the graph, we can easily infer the relative spatial position between any two horizon patches by checking the reachability between the corresponding two vertices. The reachability can easily be obtained through the reachability matrix  $\mathbf{R}$ , which is calculated based on the adjacency matrix of  $G$  using Algorithm 3. The term  $\mathbf{R}(i, j) = 1$ , which is the element

at the  $i$ th row and  $j$ th column of  $\mathbf{R}$ , indicates that the  $i$ th vertex can reach the  $j$ th vertex on  $G$ ; otherwise,  $\mathbf{R}(i, j) = 0$ . For example, the relative spatial position between A and E in Figure 2 is not given by scanning traces, but A can reaches E via C. It suggests that horizon patch A is located on top of horizon patch E, which is in line with the ground truth. In summary, if  $s$  can reach  $t$  it indicates that the two patches denoted by  $s$  and  $t$  cannot be grouped into one horizon because it would violate the over/under relationships of stratigraphic sequences. If not, there is a possibility

that the two patches can be grouped if they satisfy other aforementioned requirements. From this, we can see that the relative spatial positions among horizon patches have been fully encoded in the graph  $G$ .

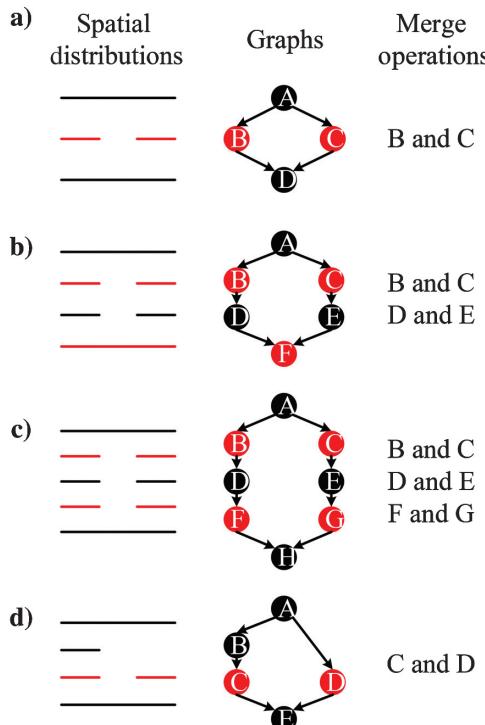
**Algorithm 2.** Vertically scanning for the addition of edges on  $G$ .

---

**Input:** horizon patches  $P$  and the threshold for filtering noise  $n_o$   
**Output:** the edge list  $E$

- 1 Initialize  $E = \{\}$ ;
- 2 **for** each trace  $t$  of  $P$  **do**
- 3   Extract a list  $l$  of indices of horizon points in increasing order of time;
- 4   **for**  $i = 1:\text{length}(l)-1$  **do**
- 5     **if**  $(l(i), l(i+1))$  exists in  $E$ ;
- 6     **then**
- 7        $E[(l(i), l(i+1))] += 1$ ;
- 8     **end**
- 9     **else**
- 10       $E[(l(i), l(i+1))] = 1$ ;
- 11     **end**
- 12   **end**
- 13 **end**
- 14 Remove all edges in  $E$  that appear smaller than  $n_o$  times;
- 15 return  $E$ ;

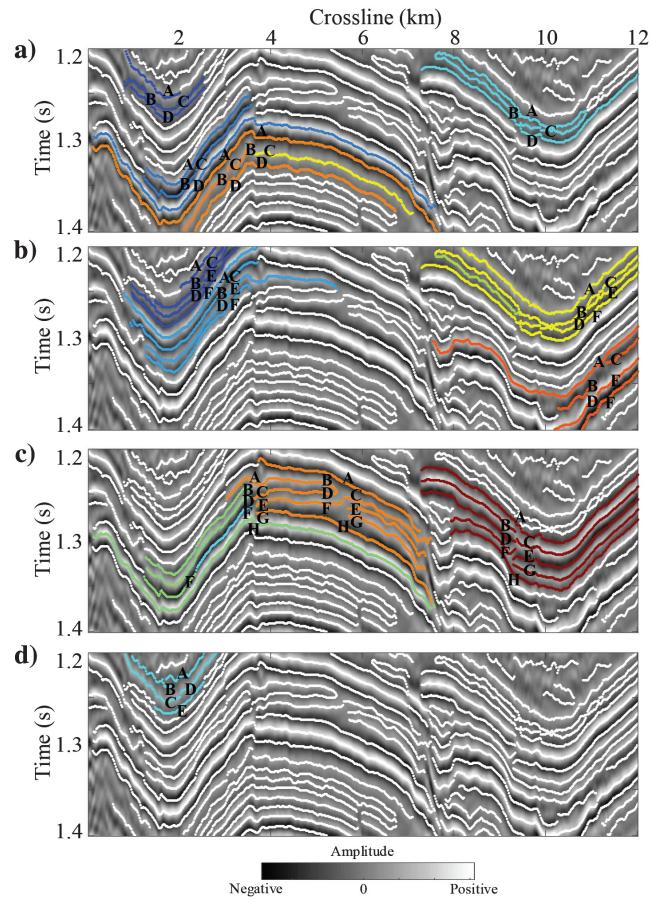
---



**Figure 3.** Structures that need to be premerged.

### Premerging

Based on the graph built in the last step, we first premerge some horizon patches through matching one of the four predefined subgraph patterns, which are shown in Figure 3. All cases are designed on the condition of parallel sets of consecutive horizons bounded by two laterally continuous horizons on the top and bottom, respectively. Under the constraint of the principle of superposition, these horizon patches are grouped in a fixed way. Take the horizon patches shown in Figure 3a as an example. There is a gap between two horizon patches  $B$  and  $C$  that are bounded by another two laterally continuous horizon patches  $A$  and  $D$ . Merging horizon patches  $B$  and  $C$  is the only option because they share the same color and no other horizon patch with the same color is in the vicinity. Similar scenarios are shown in the remaining three subplots. Furthermore, we give the matched subgraphs on the real data shown in Figure 4. Each subplot corresponds to one case shown in Figure 3, and each capital letter gives the label of the horizon patch, which indicates its role in the



**Figure 4.** Matched subgraphs on real data, which correspond to Figure 3. Each capital letter gives the label of the horizon patch, which indicates its role in the matched subgraph. For the matched subgraphs, we would directly merge patches according to the corresponding way shown in the rightmost column in Figure 3 without the calculation of waveform similarity. Some matched subgraphs may be overlapped.

matched subgraph. From these matched subgraphs, we can easily check the effectiveness of premerging. One thing to note is that these four cases are discussed due to their representativeness. Premerging is not limited to these four cases, and more subgraph patterns can be introduced into this procedure based on domain knowledge. In addition, interpreters can select the combination of subgraph patterns according to their understanding of the real data. These special spatial distributions actually suggest that there is only one way to merge these horizon patches under scenarios illustrated by these subgraphs. Merging horizon patches in this way usually is more effective than calculating the similarity between each of the two horizon patches because the similarity would not be so reliable when multiple horizon patches share similar characteristics of waveforms. Once two patches are grouped into one, the corresponding two vertices are replaced with one vertex and all of the edges that are linked with the original two vertices are set to connect with the new vertex.

A similar strategy is discussed by Forte et al. (2016). Five possible horizon grouping scenarios are discussed, and each one involves four parallel sets of vertically consecutive horizons. But without the directed and colored graph, complex subgraph patterns cannot be easily defined under the framework proposed by Forte et al. (2016). In addition, one advantage of premerging is that subgraph matching has been well-studied for efficient implementation, which can be supported in one billion node graphs (Sun et al., 2012). In summary, premerging on a directed and colored graph is easier to implement and enjoys much wider applicability compared to the method developed by Forte et al. (2016).

#### *Ordered hierarchical clustering*

After premerging, all possible spatial combinations of horizon patches have been properly grouped. Therefore, only similarities of horizon patches are available for further patch clustering. But one thing to note is that the process of clustering should follow the constraints of order, which is to ensure the geologic reasonability. As mentioned before, as long as one vertex is reachable

---

#### **Algorithm 3. Compute the reachability matrix $\mathbf{R}$ based on the adjacency matrix $A$ .**

---

**Input:** adjacency matrix  $A$  of the graph  $G$

**Output:** reachability matrix  $\mathbf{R}$

```

1 Set  $n =$  the number of the row of  $A$ ;
2 Initialize  $\mathbf{R} = I_{n \times n}$  ( $I$  denotes the identity matrix);
3 for  $i = 1:n$  do
4    $\mathbf{R} = \mathbf{R} + \mathbf{R} \times A$  (for the matrix related operations, the
     logical or is used in place of '+', and the logical and is
     used in place of '×');
5 end
6 return  $\mathbf{R}$ ;
```

---

by another vertex, the two corresponding horizon patches cannot be grouped and the information of order is given by the reachability matrix (Algorithm 3). The input to Algorithm 3 is an  $n \times n$  adjacency matrix (obtained in the way shown in Figure 2) of the graph  $G$ , where  $n$  is the number of horizon patches. In addition, because the grouping is conducted in a sequential way, the order should also be updated along with clustering for the changes of the graph  $G$ . Here, to ensure that the clustering method works well with the order, we propose a novel ordered hierarchical clustering method to group horizon patches in a hierarchical way. Classic hierarchical clustering (Rokach and Maimon, 2005) begins with each observation in its own cluster, and then it repeatedly executes the following steps: Identify the two clusters that are most similar and merge them into one cluster. This continues until all clusters are merged together. To decide which clusters should be merged, a measure of similarity between sets of observations is required. Common choices are the Euclidean distance or Hamming distance. In ordered hierarchical clustering, we follow the main idea of classic hierarchical clustering and recursively find a pair of horizon patches to merge until there is not a pair of horizon patches that meets the requirement for the emergence. But the difference is that the similarity between two observations is not only the factor that is taken into consideration for mergence but also the order between the two observations (inferred from Algorithm 3) is included. The order specifies the relative geologic age of the horizon patch, and two horizon patches with different relative geologic age would not be allowed to be merged. In addition, in

---

#### **Algorithm 4. Ordered hierarchical clustering.**

---

**Input:** Graph  $G$  and the mean of windowed waveform of each horizon patch  $W$

**Output:** Clustering result  $G_c$

```

1 Initialize  $G_c = G$ ;
2 while True do
3   Calculate the reachability matrix  $\mathbf{R}$  using Algorithm 3 by
     inputting the adjacency matrix of  $G_c$ ;
4   Calculate the pairwise distance matrix  $S$  based on  $W$ 
     between each two horizon patches represented by  $G_c$ ;
5    $D = \delta_\infty(\mathbf{R}) + S$ , where  $\delta_\infty(x) = \infty$  if  $x = 1$ , else 0;
6   Find the minimum  $m_d$  on  $D$ , of which index is  $(i, j)$ ;
7   if  $m_d < \infty$  then
8     Group node  $i$  and  $j$  into one new node and update  $G_c$ 
     and  $W$ ;
9   end
10  else
11    Stop;
12  end
13 end
14 return  $G_c$ ;
```

---

the process of clustering, we would dynamically update the order as the change of the graph topology caused by the mergence. Specifically, we first find the most similar two patches that are not reachable to each other, which can be checked based on the reachability matrix  $\mathbf{R}$ . Here, each horizon patch consists of waveforms windowed on each horizon point and the similarity between two horizon patches is defined as the cosine similarity between the means of waveforms of the two horizon patches, which can be represented as

$$\text{similarity} = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|}, \quad (1)$$

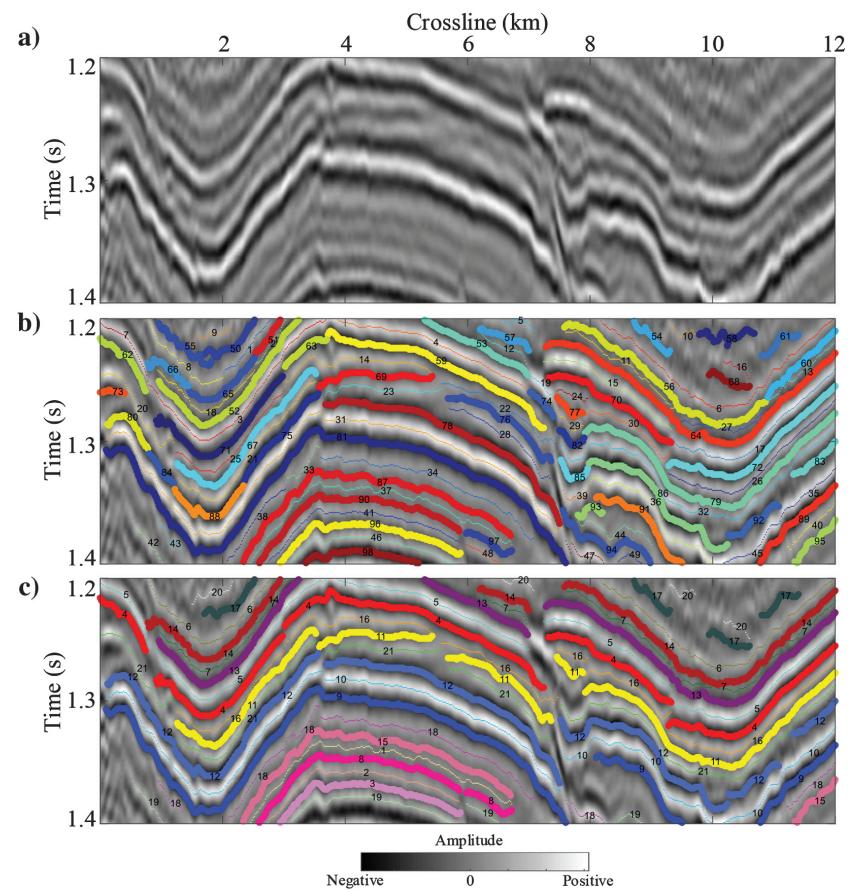
where  $\cdot$  denotes the dot product and  $\|\cdot\|$  represents the magnitude. Then, the selected pair of horizon patches is grouped into one horizon patch by replacing the two nodes with a new node, and all incoming and outgoing edges to the original two nodes are transferred to the new node. Because the topology of the graph  $G$  has been updated, the reachability matrix would be updated using Algorithm 3. By repeating the above procedure until we cannot find any two horizon patches that are not reachable to each other, the clustering is considered to be completed. Algorithm 4 provides the details of ordered hierarchical clustering.

## Experiment

We apply our method to a real seismic data set obtained from the Sichuan Basin in China. This survey has a strong regional tectonic setting and a characteristic set of developed faults, complex structural forms, and large-scale traps, which collectively facilitate the migration and accumulation of oil and gas.

Figure 5a shows a poststack seismic section mainly involving a fault in the middle. The fault breaks the continuities of multiple horizons, which makes the extraction of horizons quite difficult. Figure 5b provides the result of patching. Nearly 100 horizon patches are extracted. Some short horizon patches are filtered because it would possibly introduce errors. Because the two hyperparameters, which are the size of lateral window  $w$  and the threshold of maximum time difference for growing  $\delta$ , are set with relatively small values (both are three), we can see that those horizons are easily broken into several patches for tiny discontinuities. Even though the behavior of oversegmentation would lead to extensive merging operations, it reduces the possibility of wrongly merging two horizon patches belonging to two different horizons into one (e.g., patch 23 and

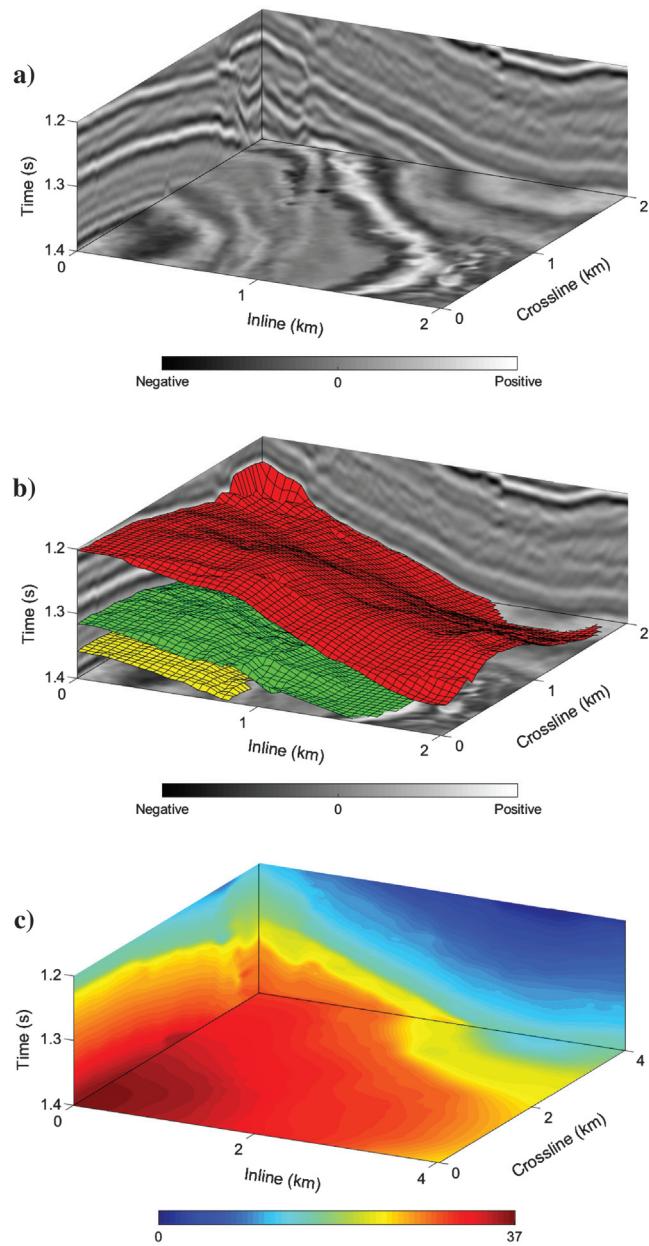
patch 22, which is located near [6, 1.25] in Figure 5b), which would possibly lead to incorrect constraints of order and cause a cascade of errors. Figure 5c shows the result of extracted horizons using our method. We can see that (1) no crossed horizons are observed, which is mainly because of the constraints of order and (2) the occurrence of the fault does not affect the extraction of horizons. The horizons patches located on each side of the fault are precisely matched and grouped into one horizon. Taking the grouping of patch 85 (located on the right side of the fault and near [8, 1.32] in Figure 5b) as an example, there are two horizon patches that share similar characteristics of the waveform on the left side of the fault, which are patch 59 (located near [6, 1.24]) and patch 78 (located near [6, 1.29]). But once patch 28 and patch 32 are grouped or patch 22 and patch 24 are grouped together, constrained by the order inferred from the graph, patch 59 and patch 85 cannot be merged because one horizon is located between them. Therefore, in the scenario that horizon patches are not aligned vertically due to the occurrence of the fault, the ordered hierarchical clustering actually simplifies the problem by selecting the pair of horizon patches that are most similar



**Figure 5.** (a) Two-dimensional seismic profile that is used for testing the proposed method and (b) extracted horizon patches. Different patches are labeled with different colors and patch indices in the middle of patches and (c) extracted horizons. Thick lines represent troughs, and thin lines represent peaks. Horizons with a relatively small length are filtered for better visualization.

to the group and reconstructing the relative order among these horizon patches located on the two sides of the fault step by step. Even in a quite complex scenario that our method cannot precisely handle, a reasonable grouping can be obtained because a lot of ambiguities are clarified with only a few manual interactions.

Moreover, we test our method on a 3D data set in the same work area, which is shown in Figure 6. Based on the extracted horizons, we also show the corresponding RGT volume (Stark, 2003, 2004; Labrunye and Carn, 2015) shown in a chair display (Figure 6c), in which the RGT value for each horizon is assigned by the interpreter based on the topological sorting of the graph after all groupings. We can observe that the extracted



**Figure 6.** (a) Three-dimensional seismic volume, (b) three extracted horizons, and (c) chair display of the corresponding RGT volume.

horizons coincide well with seismic reflectors. As for the running time, it takes 15.4 s to generate the result on the 2D data set with size of  $601 \times 201$  and 292.3 s on the 3D data set with the size of  $201 \times 201 \times 200$  (the experiment environment is an Intel E3-1230 V2 3.30 GHz CPU, 16 GB RAM). With effective and parallel implementation and support of powerful hardware, the method can easily process large seismic data within a reasonable time.

To demonstrate the effectiveness of ordered clustering and premerging, we also give the results of removing the constraints of order and premerging, respectively, which are shown in Figure 7. Because removing the constraints of order would lead to merging all horizon patches into one horizon (step 11 in Algorithm 4 would never be reached), we would stop clustering when the number of clusters is 32, which is the same as the result shown in Figure 5c (note that not all horizons are shown and horizons with a relatively small length are filtered for better visualization). From Figure 7a, we can see that the generated horizons are not geologically reasonable without the constraint of order. One case can be observed by looking at horizon 19 marked by the black circles in Figure 7a. Furthermore, we provide the results generated by removing the premerging shown in Figure 7b and the matched subgraphs shown in Figure 4. One error that can be observed in Figure 7b is the relationship between patch 13 and patch 21 (located near [6, 1.28] in Figure 7b), which should be grouped into one horizon. This error can be corrected by premerging because these two horizon patches and another four form the special spatial distribution shown in Figure 3c. With premerging, patch 13 and patch 21 would be grouped before conducting the ordered clustering. In addition, from the matched subgraphs shown in Figure 4, we can see that all of the matched subgraphs lead to reasonable groupings of horizon patches. In summary, we conclude that these two modules are necessary for precise horizon extraction.

## Discussion

The most closely related work to ours is Monsen et al. (2007), in which a directed graph is also used to estimate the relative geologic ages based on the results given by the topological sorting of the graph. Two key differences are (1) the directed graph is only designed to generate topological sorting in Monsen et al. (2007), which determines the RGT of seismic events. However, the topological sorting is prone to error if the topological relationships among horizon patches are not fully observed. For example, given a directed graph with five vertices A, B, C, D, and E and five edges AB, BC, CE, AD, and DE, the linear ordering is A:1, B:2, C:3, E:4, and D:2, which suggests that patch B and patch D should be merged together. However, it is also possible that patch C and patch D should be merged (just like the subplot in Figure 3d, C and D should be merged, which is inferred from the polarity). In this example, the topological sorting possibly generates errors because it does not fully observe the relationship among B, C, and D. Therefore, the information that two

patches in the same order of topological sorting should be merged is prone to error. Moreover, in a graph with many vertices that represent horizon patches, more aforementioned errors would occur. But, the relative over/under relationships between two patches (e.g., A precedes B and A precedes C) are much more reliable, which is also the basic assumption behind constructing the directed graph in our paper. Therefore, instead of using the linear ordering given by topological sorting, we mainly focus on the relative order between each two vertices. That is, we care that A is in front of B rather than A being the first one in the linear ordering and (2) other than the ordering information inferred from the directed graph, we also introduce more effective auxiliary information (the polarity, spatial distributions, and waveform similarity) into the process of grouping and two new and effective techniques, subgraph matching-based premerging and ordered hierarchical clustering, are proposed and experiments demonstrate the effectiveness of the two methods.

There are three hyperparameters in our method, which are the lateral window size  $w$ , vertical window size  $\delta$ , and threshold for filtering noise  $n_0$ . The terms  $w$  and  $\delta$  specify the size of the analysis window of the region growing for patching, and the region growing method considers that all points of interest within the window belong to one cluster. Therefore, large values of  $w$  and  $\delta$  usually lead to fewer horizon patches and each of them is larger, which would simplify the following procedure of grouping. But this setting is prone to incorrectly collecting horizon points into one horizon patch. We recommend to increase the two hyperparameters from a small value and stop once incorrect

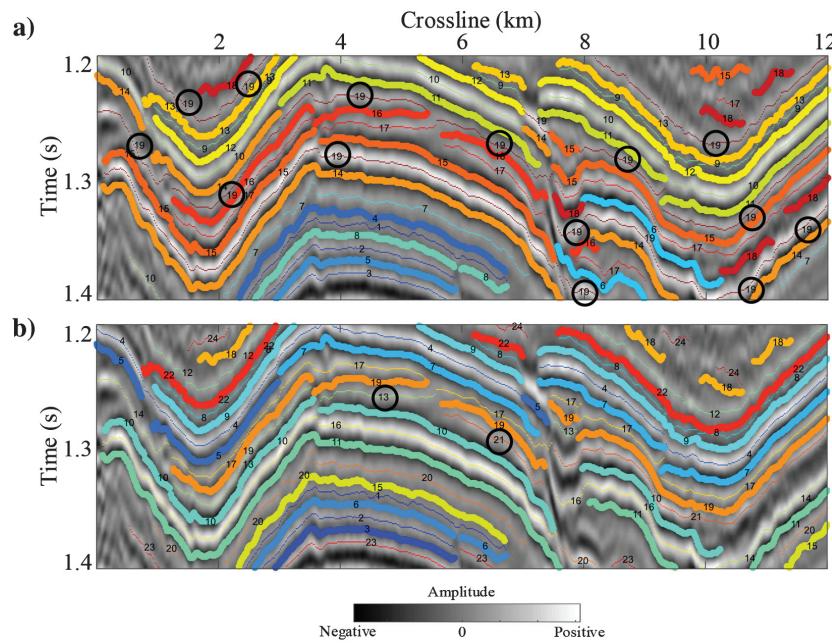
patching is presented. In our implementation, we conservatively set both of them to three because of the poor lateral continuity caused by the occurrence of faults. The third hyperparameter  $n_0$  is to filter these noisy horizon points that would introduce errors while conducting the vertical scanning for building the graph. The basic requirement for  $n_0$  is less than the minimum overlap of any two horizon patches. But the minimum overlap is difficult to estimate; therefore, we empirically find that the empirical setting  $n_0 = 5$  (two dimensions) or  $n_0 = 20$  (three dimensions) can handle most cases.

One minor defect of our method is that errors raised in the previous step would greatly deteriorate the results given by the following steps due to the cascaded workflow. For example, when different horizons are observed at a similar depth due to the complex geology, region growing-based patching would easily collect extreme points from different horizons into one horizon patch. If that is the case, then wrongly merged horizon patches would introduce wrong over/under relationships into the process of grouping. Therefore, we would recommend using a relatively small window size and including more criteria (such as waveform similarity) for the region growing method. Even though many more horizon patches would be generated with small window sizes, the extreme points would be collected rightly into horizon patches with a higher possibility. A similar problem would also be observed in the procedure of ordered hierarchical clustering because it is implemented in an incremental way. Once a pair of horizon patches is wrongly merged, the order information given by the directed and colored graph would be

wrong in some local areas. One way to solve this kind of problem is to integrate manual guidance or introduce a mechanism of self-correction (e.g., abandon some operations of mergence once they lead to the violation of some rules [defined by some domain knowledge] in the process of incremental grouping).

## Conclusion

In this paper, we present a novel method on a directed and colored graph for horizon tracking. Following the framework of patching and grouping, we first generate horizon patches by applying the region growing method on the detected extreme points. Then, a directed and colored graph is built to encode the effective context information that delineates the relatively over/under spatial positions of the horizon patches. Based on the graph, two techniques, which are subgraph matching-based premerging and ordered hierarchical clustering, are introduced to group horizon patches into complete horizons. The subgraph matching-based premerging provides an effective



**Figure 7.** Extract horizons using the proposed method by removing the constraints of order or premerging. The figure shows the results (a) without the constraints of order and (b) without premerging. Thick lines represent troughs, and thin lines represent peaks. Horizons with a relatively small length are filtered for better visualization.

way to use the spatial distributions of horizon patches, and some ambiguities brought by grouping based on the waveform similarity between patches are solved. On the other side, ordered hierarchical clustering groups horizon patches in an incremental way constrained by the RGT inferred from the directed and colored graph. Overall, the proposed method fully explores the spatial information about horizon patches and generates promising results.

Future works will focus on (1) how to correct wrong relative spatial positions among patches. Although the relative position plays an important role in the process of clustering, one error would lead to a cascade of errors. Therefore, a mechanism of self-correction would be necessary. (2) How to deal with reverse faults because vertically scanning would take two horizon patches belonging to one horizon but separated by a reverse fault as two patches that cannot be merged by the constraints of order (a case can be observed on patch 19 and patch 15, which are located near [7.2, 1.25] in Figure 5b).

## Acknowledgments

The research was supported by the National Natural Science Foundation of China (nos. 41604107, 41804112, and U156220066). The data were provided by the Geophysical Exploration Company, Chuanqing Drilling Engineering Co. Ltd., CNPC. We also thank the reviewers for their constructive comments and suggestions. The authors gratefully acknowledge the financial support from the China Scholarship Council.

## Data and materials availability

Data associated with this research are confidential and cannot be released.

## References

- Adams, R., and L. Bischof, 1994, Seeded region growing: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**, 641–647, doi: [10.1109/34.295913](https://doi.org/10.1109/34.295913).
- Borgos, H. G., T. Skov, T. Randen, and L. Sonneland, 2003, Automated geometry extraction from 3D seismic data: 73rd Annual International Meeting, SEG, Expanded Abstracts, 1541–1544, doi: [10.1190/1.1817590](https://doi.org/10.1190/1.1817590).
- Chopra, S., and K. J. Marfurt, 2007, Seismic attributes for prospect identification and reservoir characterization: SEG and EAGE.
- Di, H., 2018, Developing a seismic pattern interpretation network (SpiNet) for automated seismic interpretation: arXiv preprint arXiv:1810.08517.
- Di, H., M. Alfarraj, and G. AlRegib, 2017, 3D curvature analysis of seismic waveform and its interpretational implications: 87th Annual International Meeting, SEG, Expanded Abstracts, 2255–2259, doi: [10.1190/segam2017-17747188.1](https://doi.org/10.1190/segam2017-17747188.1).
- Di, H., D. Gao, and G. AlRegib, 2018, 3D structural-orientation vector guided autotracking for weak seismic reflections: A new tool for shale reservoir visualization and interpretation: Interpretation, **6**, no. 4, SN47–SN56, doi: [10.1190/INT-2018-0053.1](https://doi.org/10.1190/INT-2018-0053.1).
- Dossi, M., E. Forte, and M. Pipan, 2015, Automated reflection picking and polarity assessment through attribute analysis: Theory and application to synthetic and real ground-penetrating radar data: *Geophysics*, **80**, no. 5, H23–H35, doi: [10.1190/geo2015-0098.1](https://doi.org/10.1190/geo2015-0098.1).
- Fomel, S., 2010, Predictive painting of 3D seismic volumes: *Geophysics*, **75**, no. 4, A25–A30, doi: [10.1190/1.3453847](https://doi.org/10.1190/1.3453847).
- Forte, E., M. Dossi, M. Pipan, and A. Del Ben, 2016, Automated phase attribute-based picking applied to reflection seismics: *Geophysics*, **81**, no. 2, V141–V150, doi: [10.1190/geo2015-0333.1](https://doi.org/10.1190/geo2015-0333.1).
- Harrigan, E., J. Kroh, W. Sandham, and T. Durrani, 1992, Seismic horizon picking using an artificial neural network: IEEE International Conference on Acoustics, Speech, and Signal Processing, IEEE ICASSP-92, 105–108, doi: [10.1109/ICASSP.1992.226265](https://doi.org/10.1109/ICASSP.1992.226265).
- Hoyes, J., and T. Cheret, 2011, A review of “global” interpretation methods for automated 3D horizon picking: *The Leading Edge*, **30**, 38–47, doi: [10.1190/1.3535431](https://doi.org/10.1190/1.3535431).
- Huang, K.-Y., 1997, Hopfield neural network for seismic horizon picking: 67th Annual International Meeting, SEG, Expanded Abstracts, 562–565, doi: [10.1190/1.1885963](https://doi.org/10.1190/1.1885963).
- Khan, M. M., and A. Alam, 2012, High resolution adaptively steered horizon tracker: Istanbul International Geophysical Conference and Oil & Gas Exhibition, SEG and The Chamber of Geophysical Engineers of Turkey, 1–4, doi: [10.1190/IST092012-001.95](https://doi.org/10.1190/IST092012-001.95).
- Labrunye, E., and C. Carn, 2015, Merging chronostratigraphic modeling and global horizon tracking: Interpretation, **3**, no. 2, SN59–SN67, doi: [10.1190/INT-2014-0130.1](https://doi.org/10.1190/INT-2014-0130.1).
- Leggett, M., W. A. Sandham, and T. S. Durrani, 2003, Automated 3-D horizon tracking and seismic classification using artificial neural networks, in W. A. Sandham and M. Leggett, eds., *Geophysical applications of artificial neural networks and fuzzy logic*: Springer, 31–44.
- Liu, Z., C. Song, H. Cai, X. Yao, and G. Hu, 2017, Enhanced coherence using principal component analysis: Interpretation, **5**, no. 3, T351–T359, doi: [10.1190/INT-2016-0194.1](https://doi.org/10.1190/INT-2016-0194.1).
- Lou, Y., and B. Zhang, 2018, Automatic horizon picking using multiple seismic attributes: 88th Annual International Meeting, SEG, Expanded Abstracts, 1683–1687, doi: [10.1190/segam2018-2997698.1](https://doi.org/10.1190/segam2018-2997698.1).
- Lowell, J., and G. Paton, 2018a, Application of deep learning for seismic horizon interpretation: 88th Annual International Meeting, SEG, Expanded Abstracts, 1976–1980, doi: [10.1190/segam2018-2998176.1](https://doi.org/10.1190/segam2018-2998176.1).
- Lowell, J., and G. Paton, 2018b, Deep learning based horizon interpretation: 80th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: [10.3997/2214-4609.201800923](https://doi.org/10.3997/2214-4609.201800923).
- Monsen, E., H. G. Borgos, P. Le Guern, and L. Sonneland, 2007, Geologic-process-controlled interpretation based on 3D Wheeler diagram generation: 77th Annual

- International Meeting, SEG, Expanded Abstracts, 885–889, doi: [10.1190/1.2792549](https://doi.org/10.1190/1.2792549).
- Qian, F., C. Luo, Z. Su, X. Yao, G. Hu, and X. Cheng, 2014, A two-step mechanism for automated 3D horizon picking: 84th Annual International Meeting, SEG, Expanded Abstracts, 1548–1553, doi: [10.1190/segam2014-0984.1](https://doi.org/10.1190/segam2014-0984.1).
- Rokach, L., and O. Maimon, 2005, Clustering methods, in O. Maimon and L. Rokach, eds., Data mining and knowledge discovery handbook: Springer, 321–352.
- Song, C., Z. Liu, Y. Wang, X. Li, and G. Hu, 2017, Multi-waveform classification for seismic facies analysis: Computers & Geosciences, **101**, 1–9, doi: [10.1016/j.cageo.2016.12.014](https://doi.org/10.1016/j.cageo.2016.12.014).
- Stark, T. J., 2003, Unwrapping instantaneous phase to generate a relative geologic time volume: 73rd Annual International Meeting, SEG, Expanded Abstracts, 1707–1710, doi: [10.1190/1.1844072](https://doi.org/10.1190/1.1844072).
- Stark, T. J., 2004, Relative geologic time (age) volumes — Relating every seismic sample to a geologically reasonable horizon: The Leading Edge, **23**, 928–932, doi: [10.1190/1.1803505](https://doi.org/10.1190/1.1803505).
- Sun, Z., H. Wang, H. Wang, B. Shao, and J. Li, 2012, Efficient subgraph matching on billion node graphs: Proceedings of the VLDB Endowment, **5**, 788–799, doi: [10.14778/2311906.2311907](https://doi.org/10.14778/2311906.2311907).
- Wu, H., and B. Zhang, 2018, A deep convolutional encoder-decoder neural network in assisting seismic horizon tracking: arXiv preprint arXiv:1804.06814.
- Wu, X., 2017, Structure-, stratigraphy- and fault-guided regularization in geophysical inversion: Geophysical Journal International, **210**, 184–195, doi: [10.1093/gji/ggx150](https://doi.org/10.1093/gji/ggx150).
- Wu, X., and S. Fomel, 2018, Least-squares horizons with local slopes and multigrid correlations: Geophysics, **83**, no. 4, IM29–IM40, doi: [10.1190/geo2017-0830.1](https://doi.org/10.1190/geo2017-0830.1).
- Yu, Y., C. Kelley, and I. Mardanova, 2011, Automatic horizon picking in 3D seismic data using optical filters and minimum spanning tree (patent pending): 81st Annual International Meeting, SEG, Expanded Abstracts, 965–969, doi: [10.1190/1.3628233](https://doi.org/10.1190/1.3628233).
- Zhao, T., F. Li, and K. J. Marfurt, 2018, Seismic attribute selection for unsupervised seismic facies analysis using user-guided data-adaptive weights: Geophysics, **83**, no. 2, O31–O44, doi: [10.1190/geo2017-0192.1](https://doi.org/10.1190/geo2017-0192.1).

---

Biographies and photographs of the authors are not available.