Automated fault surface segmentation using machine learning

Canning Anat, Vizeu Fernando, Zambrini Joao, Emerson*

Summary

Automated fault interpretation of seismic data programs typically produces fault attribute volumes containing probability, strike, and dip estimates. After thresholding and thinning, this sparse data can be represented as a 3D point cloud and the challenge is to separate it into individual fault surfaces. Traditional deterministic methods such as fault tracking and region growing, as well as standard clustering algorithms struggle with this task due to the complexity of fault networks - handling intersections, discontinuities, noise, and the arbitrary nature of fault surfaces. We present a machine learning approach to solve this problem using a binary pairwise classification. A neural network learns whether any two points belong to the same fault surface - a formulation that sidesteps the label assignment problem while naturally handling variable numbers of faults. Trained on synthetically generated fault networks, the network develops understanding of fault connectivity by observing complete structural contexts during training. A custom geological distance function incorporates domain knowledge by penalizing connections perpendicular to fault strike, and graph-based clustering converts pairwise predictions into discrete fault surfaces. This combination of data-driven learning and physical constraints achieves excellent accuracy on complex fault configurations, providing a fully automated solution that requires only fault attribute input to produce segmented fault networks.

Introduction

Identifying and interpreting faults from seismic data is an important aspect of seismic interpretation and is quite a complex and labor-intensive task. Therefore, the automation of fault interpretation has recently become topical. Ideally the automated procedure will process seismic data and provide a complete fault network, where each fault is clearly identified as a unique surface. However, such technology is still not available. Fault networks can be very complex, while small faults, though important for detailed reservoir characterization, often mask the big picture of the major structural framework. Automatic fault interpretation is therefore often divided into two tasks: fault points identification, also referred to as fault attribute generation, and fault surface reconstruction.

There is a significant volume of research related to automatic fault attribute generation, among those we can mention Fault Likelihood (Hale et al., 2013) and FaultNet3D (Wu et al., 2019). They provide an automated (deterministic or ML-based) procedures that generates three 3D volume attributes from a single seismic dataset: P - probability of being a fault, θ - local dip angle of the fault, and φ - local strike angle of the fault. Fault surface reconstruction can be a secondary procedure that takes fault attribute data and constructs the fault network. In this process small faults can be discarded and the major network that contributes to the structural framework is created. Each individual fault surface is uniquely identified in this process. Methodologies published for fault surface extraction (for example, Wu and Hale, 2016; Dong et al., 2025; Liu et al., 2025) are often complex and include many (deterministic) steps.

In this paper we concentrate on fault surface extraction from 3-component fault attributes using a machine learning approach. We treat the fault attribute data as a point cloud rather than a 3D array. With the fault probability attribute, we set a probability threshold and extract only high-probability points. This makes the data very sparse, so using full 3D volumes becomes computationally inefficient. On the other hand, using 3D volume data would enable the application of established 3D CNNs or similar image processing technologies. We chose however, a point cloud data structure as it better suits the sparse nature of fault data. Before extracting the point cloud from the 3D data volumes we ensure the attribute has gone through a fault thinning process (Wu and Hale, 2016), so the attribute data can represent, as closely as possible, surfaces rather than general clusters. We then extract from the attribute volumes a point cloud of data (x, y, z, P, θ, φ) which is the input data for this process.

Conceptually, fault surface extraction can be framed as a point classification problem: assign each point a unique label representing its fault surface. We initially experimented with classic clustering methodologies to solve this classification problem, and indeed these can be tuned to do a reasonable job. However, the black-box clustering approach is limited and cannot be easily guided to the specifics of fault network like noise, discontinuities, intersections, etc. In addition, the classification of huge datasets such as seismic data is typically beyond the scope of classic classifiers. We chose the ML approach so we can train a network with examples of how our specific problem behaves.

Training a neural network for fault surface extraction requires a large, diverse set of labeled examples. Rather than relying on manually interpreted seismic data, we generate synthetic fault surfaces on-the-fly that capture the geometric complexity of real geological faults. The synthetic data generation rationale and methodology is described in detail in Canning et al. (2026) and produces diverse training examples with variable grid spacing, fault orientations and distributions, surface curvature, irregular boundaries, discontinuities, and 3D noise - comprehensively covering realistic geological scenarios.

Method

The objective of our method is to build a neural network that takes a 6-dimensional point cloud (x, y, z, P, θ, φ) and classifies it into individual labels (x, y, z, L), where L identifies the fault surface. Training is performed on synthetic data which is generated on-the-fly during the training process.

Neural Network Formulation: A direct multi-class classification approach—assigning each point a surface label L is unsuitable, because instance labels are arbitrary and do not represent meaningful classes. In instance segmentation tasks, including fault segmentation, label values differ between training examples with no consistent mapping (De Brabandere et al., 2017). Fault surfaces may share similar geometric characteristics, and what distinguishes them is their spatial separation and connectivity rather than intrinsic class features. Therefore, a network cannot learn to predict integer labels directly, as "surface 1" in one example has no relationship to "surface 1" in another.

A key advantage of this formulation is that during training, the neural network observes point pairs within the context of complete fault networks. Each training example contains multiple fault surfaces with varying relationships, and the network processes thousands of point pairs drawn from these networks - some from the same surface, others from different surfaces that may be nearby, parallel, intersecting, or unrelated. Through this exposure, the network learns statistical patterns of fault connectivity: what spatial separations, angular relationships, and geometric configurations characterize same-surface versus different-surface pairs in realistic geological contexts. At inference time, although the network evaluates individual point pairs, its predictions reflect this accumulated understanding of fault network structure learned from observing complete systems during training.

Network Architecture: We employ a simple yet effective Multi-Layer Perceptron (MLP) for pairwise classification, as illustrated in Figure 1. The network takes two fault points as input, each represented by six features: spatial coordinates (x, y, z) and fault attributes (P, θ, φ). A customized distance is computed between the points, and all features are concatenated into a 13-dimensional (2 points + distance) input vector. The network processes multiple point pairs simultaneously through batch processing for computational efficiency.

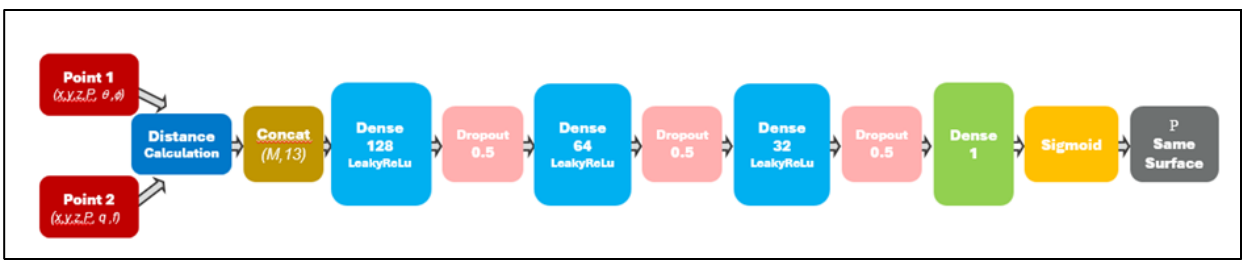

Figure 1: MLP Network Architecture

Pairwise Classification Approach**:** To avoid this permutation-invariant labeling problem, we reformulate the task as a binary pairwise classification problem: instead of predicting a label for each point, the network predicts whether two points belong to the same surface. This relational formulation is independent of how surfaces are numbered in the training data and naturally generalizes to datasets with different numbers of faults. Once pairwise connections are estimated, the final segmentation is obtained by constructing a graph of linked point pairs and extracting connected components, each representing a fault surface.

The architecture consists of three hidden layers with dropout regularization, progressively reducing dimensionality to extract discriminative features. The final output is a probability [P] representing the likelihood that the two input points belong to the same fault surface. This relatively simple architecture proved sufficient to learn complex geological relationships between points, as the discriminative power comes not from architectural complexity but from the pairwise formulation itself and the incorporation of domain knowledge through custom distance-based pair selection as discussed below. The network is trained using binary cross-entropy loss with positive class weighting and Adam optimizer.

Custom Distance**:** Our input attribute data contains, in addition to probability of each point being on a fault, also local estimates of strike and dip angles. We leverage this rich geometric information to define connectivity. Standard Euclidean distance is insufficient for fault segmentation because geological connectivity depends not only on spatial proximity but also on alignment along the fault surface. Two points may be spatially close but belong to different faults if they don't align properly along the strike direction (Figure 2).

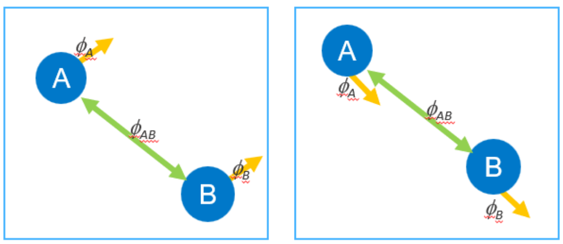

Figure 2: Yellow arrows show the local strike angle at each point. Green arrow shows the direction of the connection between the two points. Left: when the connection direction does not align with the local strike angles, points A and B belong to different faults. Right: when the connection direction aligns with the local strike angles, points A and B are on the same fault surface.

We developed a custom distance function that incorporates a directional penalty to encode this knowledge. The base distance combines spatial proximity (Euclidean distance in x, y, z) with strike angle compatibility. The key point is a directional penalty that considers *how* two points are connected relative to their input fault orientations. The penalty works as follows: we compute the azimuthal angle ϕ_{AB} of the vector connecting points *A* and *B*, we then compare this direction to the strike angles of both points (ϕ_A and ϕ_B). The angular mismatch

$$\Delta = \max(|\phi_{AB} - \phi_A|, |\phi_{AB} - \phi_B|)$$

measures whether the connection aligns with the fault strike. If points lie along the strike direction, Δ is small and minimal penalty is applied. If points lie perpendicular to the strike (across different faults), Δ is large and the distance is heavily penalized by a multiplicative scaling factor $c = 1 + k\sqrt{\sin(\Delta)}$, where k is a large constant chosen empirically (we use k=1000) to ensure strong separation of cross-strike connections. The square root function provides gradual penalty growth that balances sensitivity near parallel alignment with strong penalties for perpendicular connections.

This encodes the intuition that points on the same fault surface should be aligned along the strike direction, not across it. Figure 3 demonstrates the effectiveness of this approach on a 2D depth cross-section where faults appear as lines. Using standard Euclidean distance incorrectly merges nearby points from different faults. The custom distance correctly separates faults by recognizing that connections perpendicular to the strike direction are implausible, even when points are spatially close. Note for example the yellow points that correctly align along one line despite the large spatial separation, enabling the algorithm to bridge gaps.

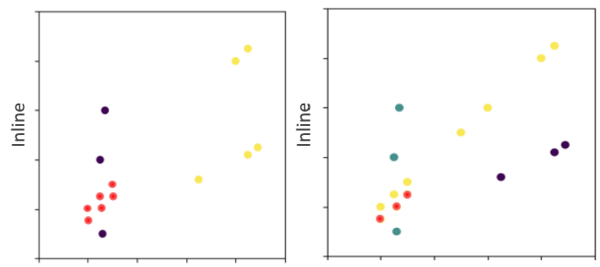

Figure 3: Illustration of fault point segmentation using a 2D depth cross-section. Colors represent different identified fault segments. Left: standard Euclidean distance incorrectly merges points from different parallel faults into single clusters. Right: our method correctly separates the faults.

Graph-Based Clustering: The final stage converts pairwise predictions into discrete fault surface segments through graph-based clustering. The approach uses dual filtering: custom distance selects candidate pairs within a threshold, then the neural network evaluates these candidates. This two-stage filtering keeps computation scalable by restricting the number of pairs sent to the neural network, which is critical for large seismic datasets. Point pairs are accepted as connections only if both the neural network probability and the custom geological distance indicate they belong to the same surface.

Approved point pairs define edges in an undirected graph where nodes represent points. We construct a sparse adjacency matrix from these edges and apply connected components analysis to identify clusters - each connected component represents a distinct fault surface. This graph-based approach naturally handles complex fault geometries including branching, gaps, and irregular shapes. Clusters with small number of points are labeled as noise (L=-1).

Examples

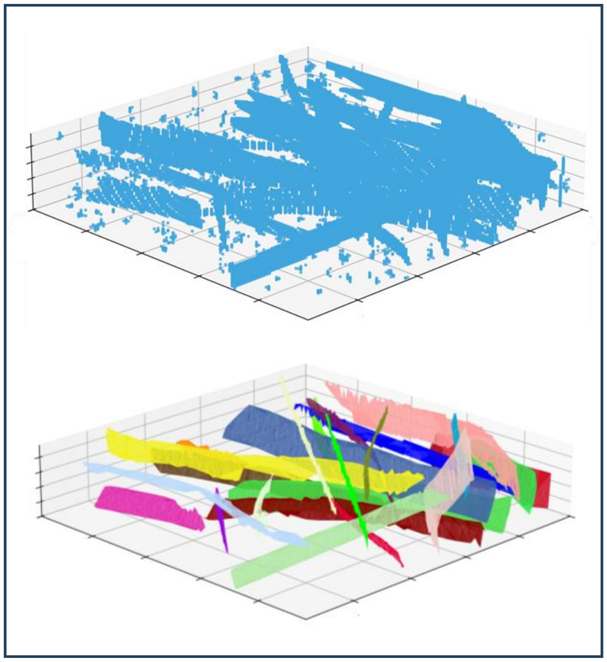

Figure 4: Synthetic example of fault surface extraction: input point cloud data (top), segmented fault surfaces with noise removed (bottom).

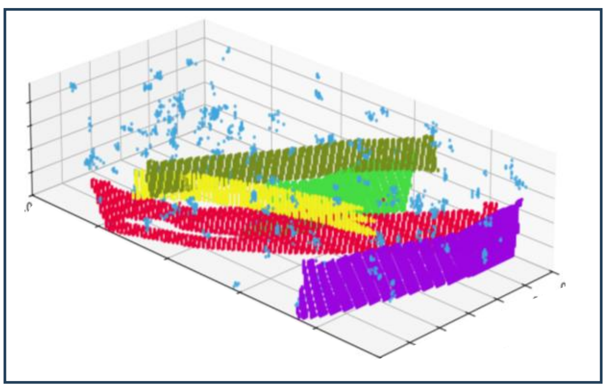

Figure 5: Synthetic example of automatically extracted faults surfaces and noise identification

Figures 4-5 demonstrate the method on synthetic data. Figure 4-top shows the input point cloud containing fault points and noise. Figure 4-bottom shows the extracted surfaces, each with a distinct color, and noise labeled L=-1 (light blue). The method achieves excellent accuracy, cleanly handling intersections, discontinuities, and irregular boundaries. Figure 5 shows another example with complex geometry including multiple intersecting faults. Note how fault intersections are cleanly defined and noise is correctly identified.

Conclusions

The pairwise classification approach effectively segments fault point clouds into individual surfaces without requiring manual labeling of training data. The combination of neural network predictions with geological distance constraints provides robust handling of fault-specific challenges including noise, discontinuities, and intersections. From the user perspective, the method provides a one-step automated procedure using a pretrained neural network model, requiring only fault attribute input to produce segmented fault surfaces. The method naturally generalizes to datasets with variable numbers of faults and handles complex fault geometries and varying dataset configurations. Training on diverse synthetic data enables the network to learn fault connectivity patterns while avoiding overfitting to specific seismic surveys.