Learning Complex Fault Structures from Hypocenter Distributions via Point Cloud Segmentation

Yanlan Hu[1] (hyl1225@stanford.edu); Gregory C. Beroza[1]

[1]Department of Geophysics, Stanford University

Key Takeaways

Challenge:
- Extracting fault geometry from earthquake catalogs is essential yet challenging.
- Conventional approaches often struggle with complex fault systems and require extensive manual intervention.

Approach:
- We develop a 3D segmentation framework using PointNet++ models trained on synthetic seismicity point clouds.
- The models are used to distinguish noise and identify fault structures by clustering earthquake hypocenters.

Findings:
- Our method successfully recovers coherent fault geometry from both induced and tectonic seismicity and generalizes well.
- The models also work well on regions lacking fault geometry knowledge in the SCEC Community Fault Model (CFM), suggesting a potential pathway to evolve the CFM.

Impact:
- This framework strengthens the ability to characterize blind faults and complex fault networks, providing a scalable tool for fault system analysis.

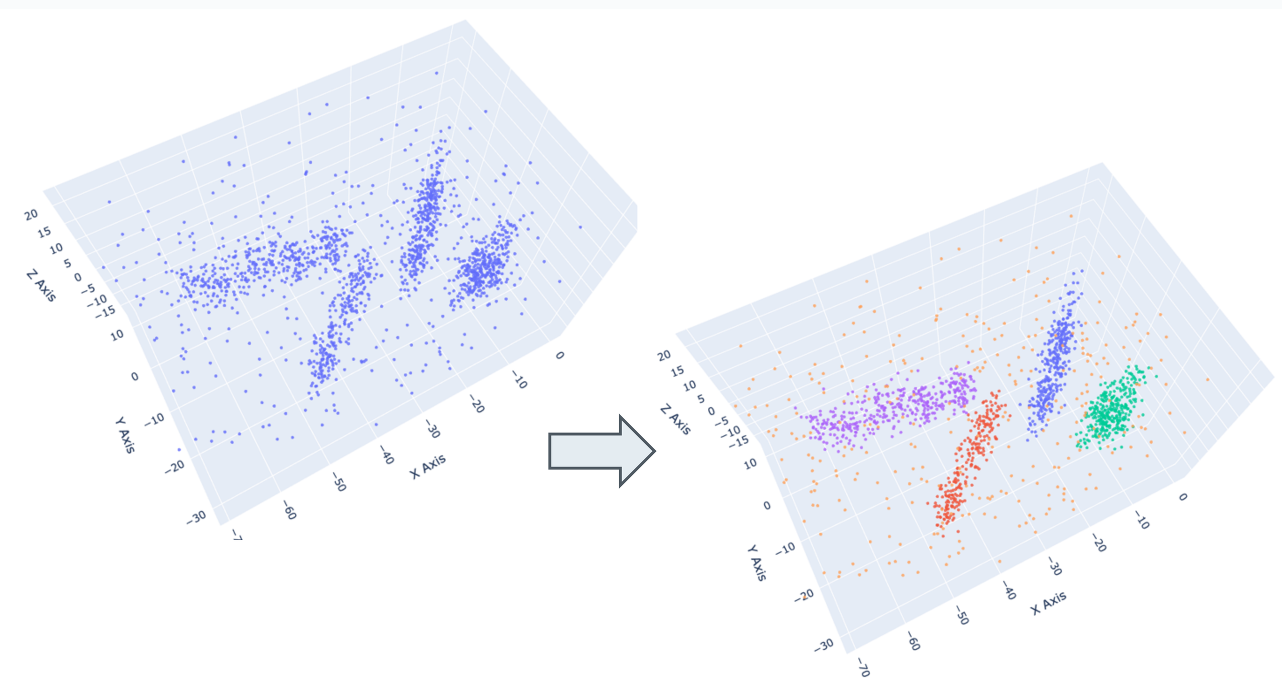

Figure 1: Conceptual illustration of the task: separating noise and clustering earthquake hypocenters into distinct fault structures.

References

1. Qi, C. R., Yi, L., Su, H., & Guibas, L. J. (2017). PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. *ArXiv*. https://arxiv.org/abs/1706.02413.

Methodology

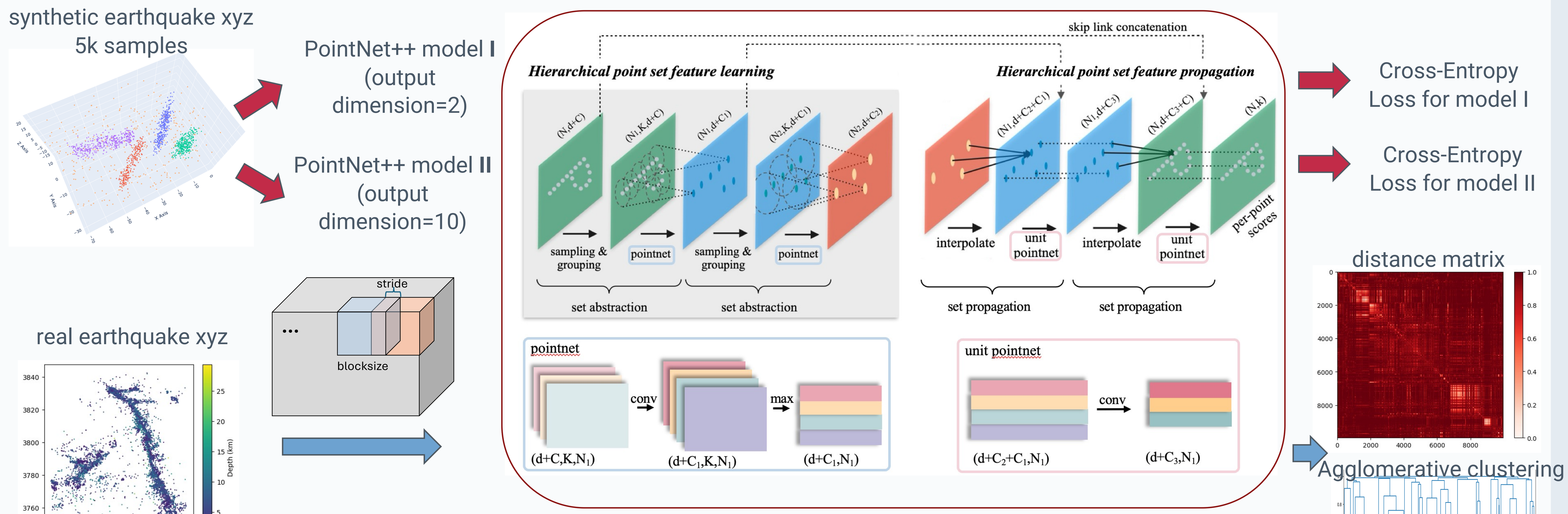

Figure 2: Workflow for training and application of our models. We trained two PointNet++ models, I to distinguish fault seismicity from a diffuse background and II to classify distinct fault structures. During application, we systematically shift a spatial block, apply both models, integrate their outputs into a distance matrix, and derive final cluster assignments through voting based on Agglomerative clustering. The model architecture illustrated in the figure is modified from Qi et al. (2017).

Applications

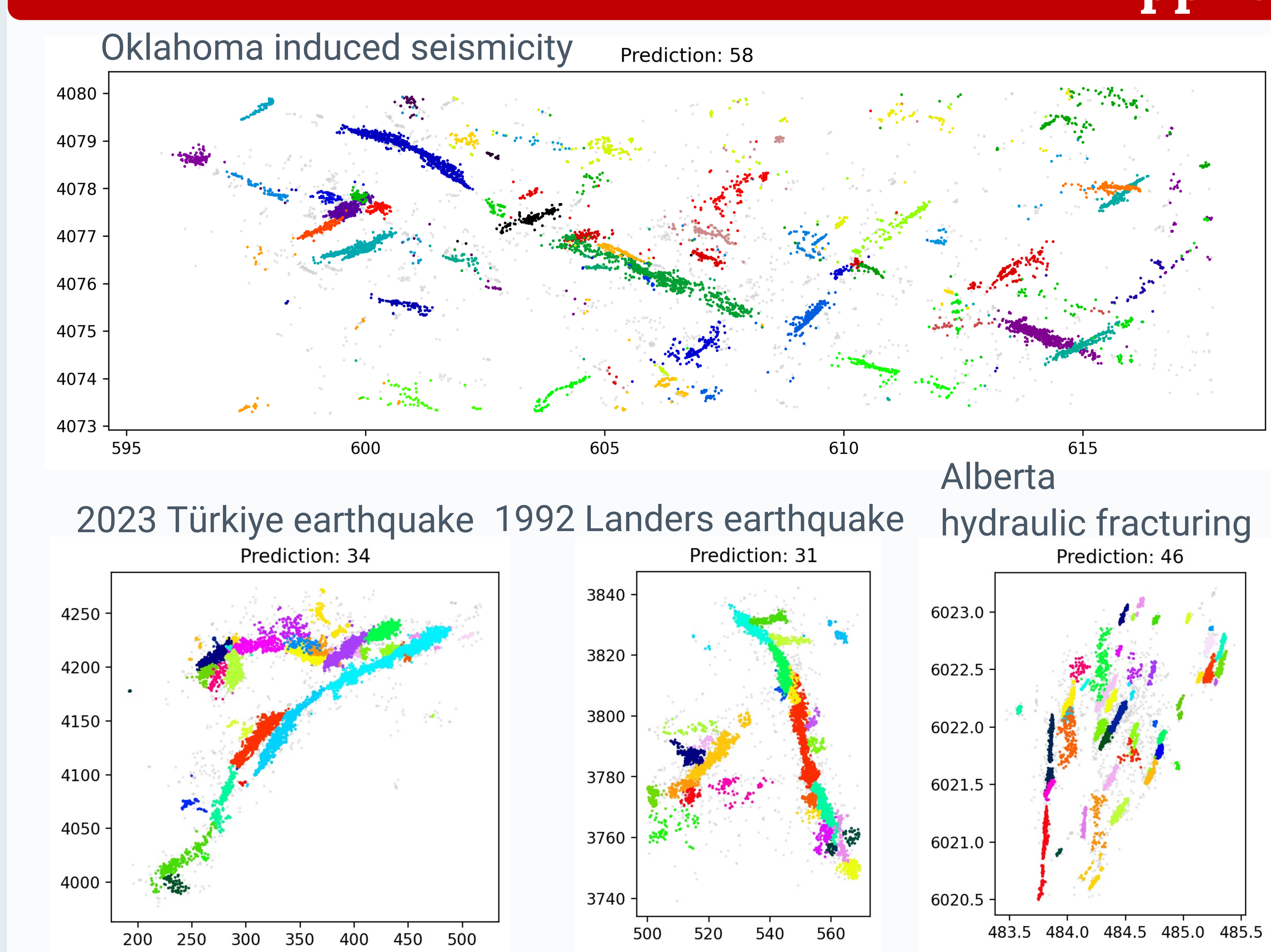

Figure 3: Clustering results of our workflow on the induced seismicity and mainshock-aftershock sequences. The number of clusters is indicated in the titles. Light gray dots denote noise earthquakes.

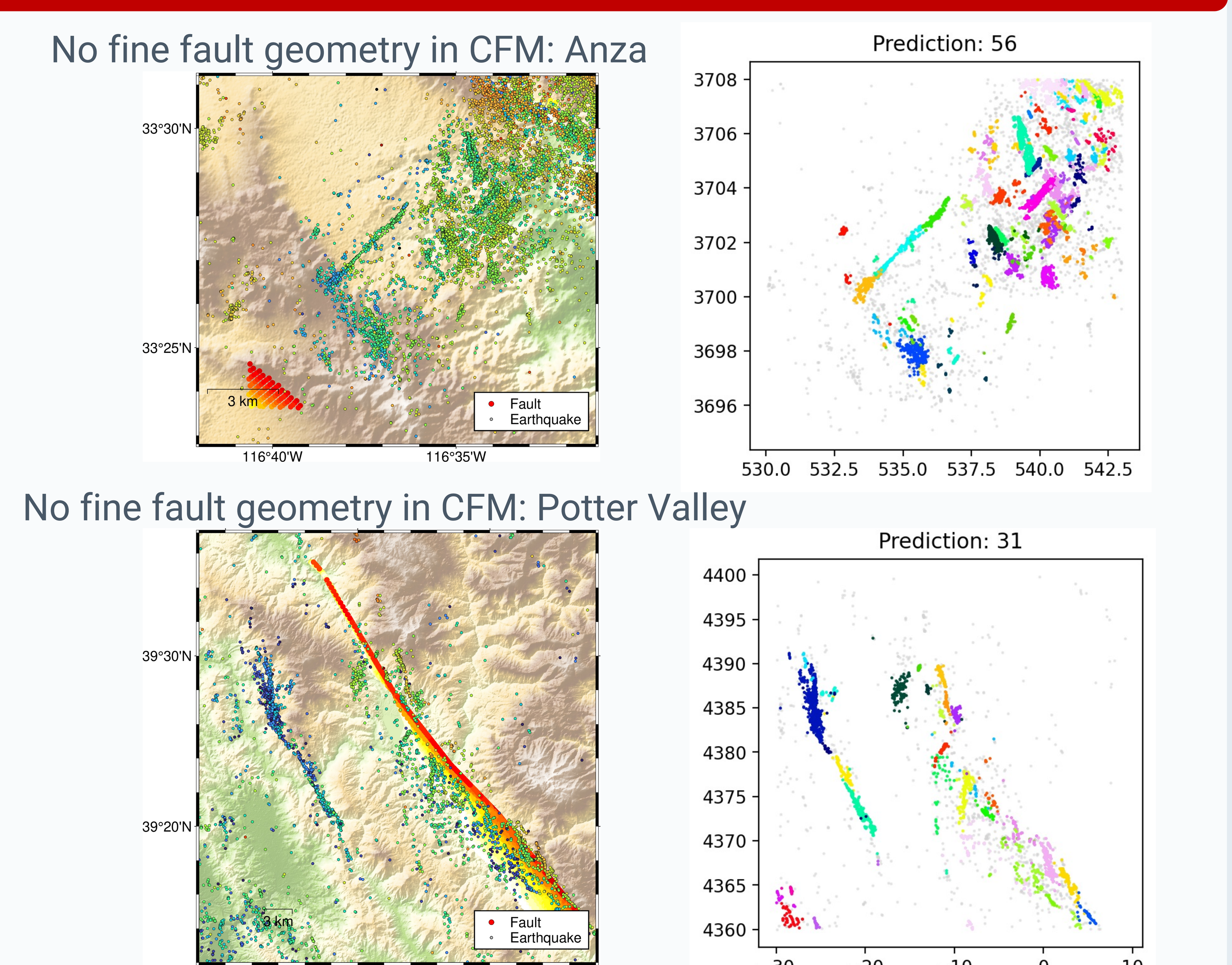

Figure 4: Clustering results of our workflow on northern and southern California regions where the CFM lacks fine fault models. The number of clusters is indicated in the titles. Light gray dots denote noise earthquakes.

Supplementary

Synthetic earthquake xyz

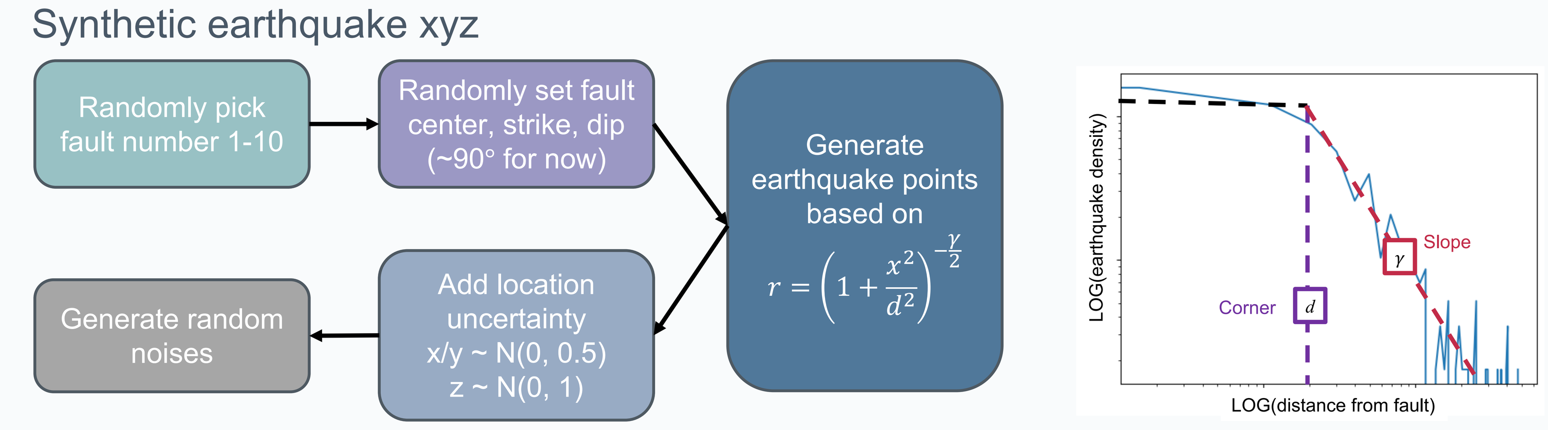

Figure 5: Workflow for synthesizing training samples. The schematic illustration of earthquake distribution as a function of distance from the fault is adapted from Travis Alongi (personal communication).

Loss computing for model II

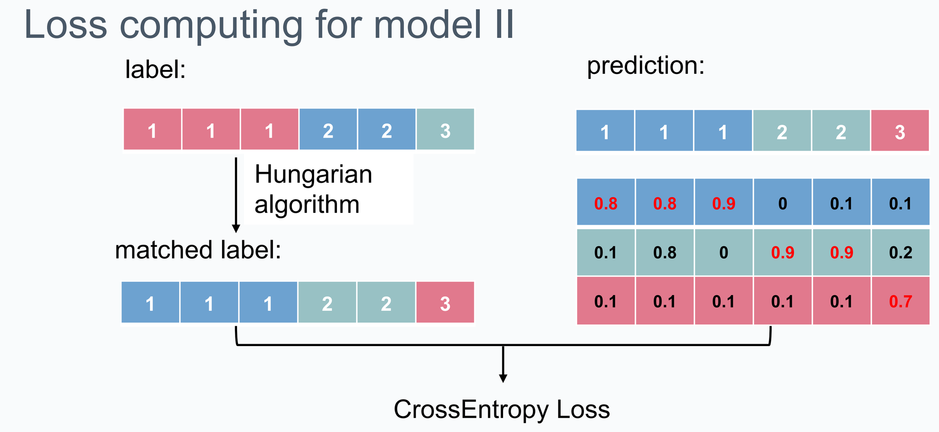

Figure 6: The labels are permuted using the Hungarian algorithm to align with the predictions prior to computing the cross-entropy loss.

How to determine blocksize

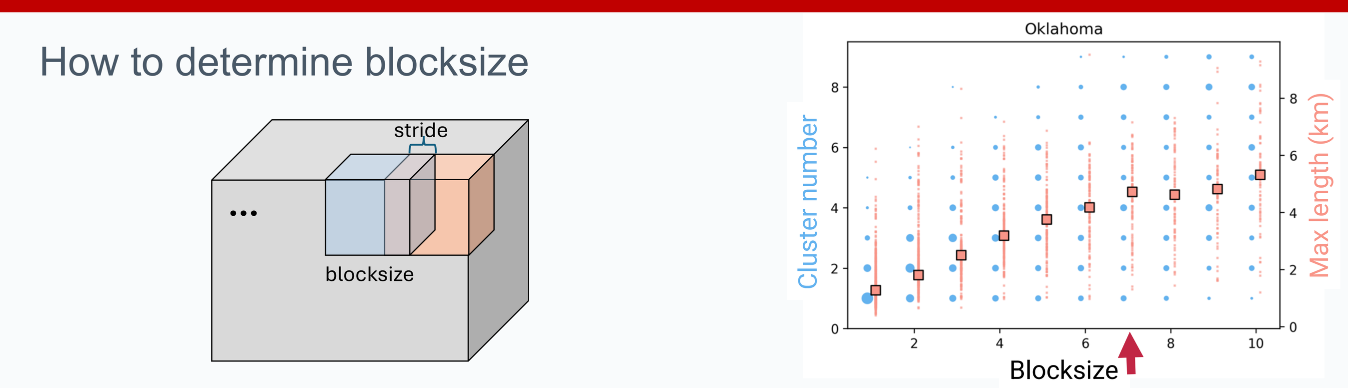

Figure 7: For each blocksize, we compute the number of clusters (blue dots, with marker size proportional to block number) and the maximum cluster lengths (orange squares). The larger squares denote the median values. A suitable blocksize is determined based on two criteria: (1) the presence of 6–8 clusters in most blocks, and (2) the identification of a plateau in maximum length. For the Oklahoma case, we select a block size of 7 km.