

Seismological Research Letter

Supplementary Material for

E2F: Extract Fault Geometry from Earthquake Catalog

Zhengtao He, Ruijia Wang*

Contents of this file

Figures S1 to S30.

Text S1-S3 corresponds to the application of E2F in ToC2ME, Oklahoma, and Türkiye regions, respectively.

Additional Supporting Information (Files uploaded separately)

Table S1. Fault geometric parameters of the ToC2ME region for fault structure modeling.

Table S2. The classification of ToC2ME seismic events within each fault model.

Introduction

The supporting information provides the user operation interface for E2F (Fig. S1), the fitted k -Mw relationship (Fig. S2), detailed applications to ToC2ME (Text S1, Table S1-S2, and Fig. S3-S9), Oklahoma (Text S2 and Fig. S10-S20), and Türkiye (Text S3 and Fig. S21-S30), respectively.

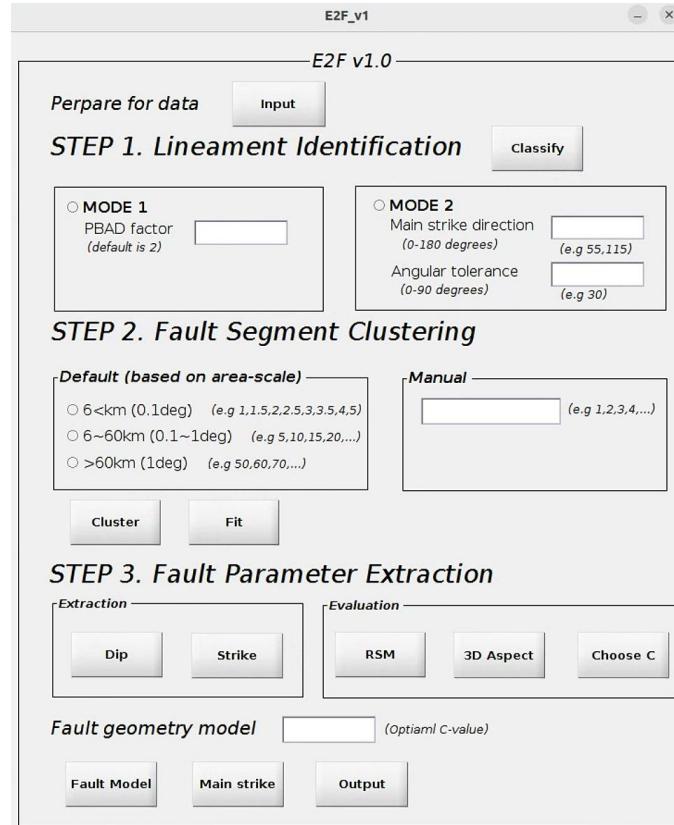


Figure S1. User operation interface for E2F (version 1.0).

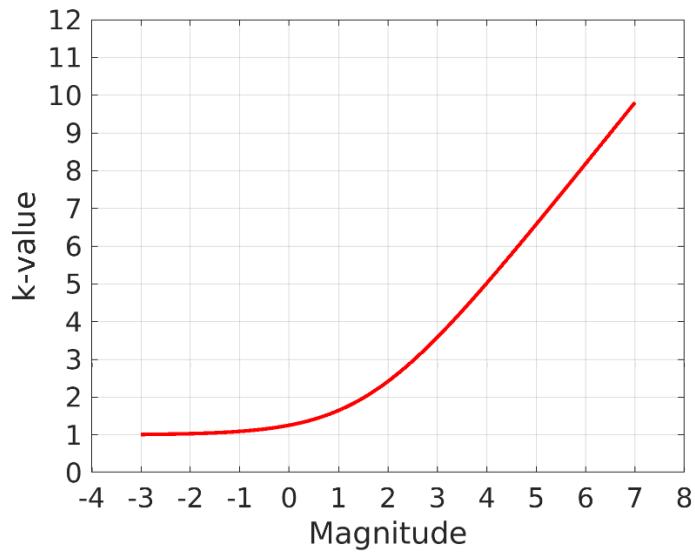


Figure S2. The k -Mw relationship.

Text S1. Applying E2F to ToC2ME: Parameter Details and Evaluations

We first compare the fault clustering results between magnitude-based and manual setting ϵ using DBSCAN (Fig. S3). The magnitude-based fault clustering results (Fig. S3d) provide more independent fault structure than the manual settings (Fig. S3a-c): a more precise local classification for SW1-SW5 and extraction of microseismic (NS2). Additionally, we provide the following detections: fault zone fitting (Fig. S4), the statistics for fault strikes (Fig. S5) and distribution of dip angles (Fig. S6), the RSM (Fig. S7) and 3-D Aspect Ratio (Fig. S8), and fault structure model based on the optimal C -value of 3.5. (Fig. S9). As C increases: 1) three parallel seismic swarms (Fig. S4e-g) were merged to a single north-south (N-S) structure (Fig. S4h); 2) although the strikes were primarily oriented along N30°E, a N-S secondary orientation was also observed (Fig. S5).

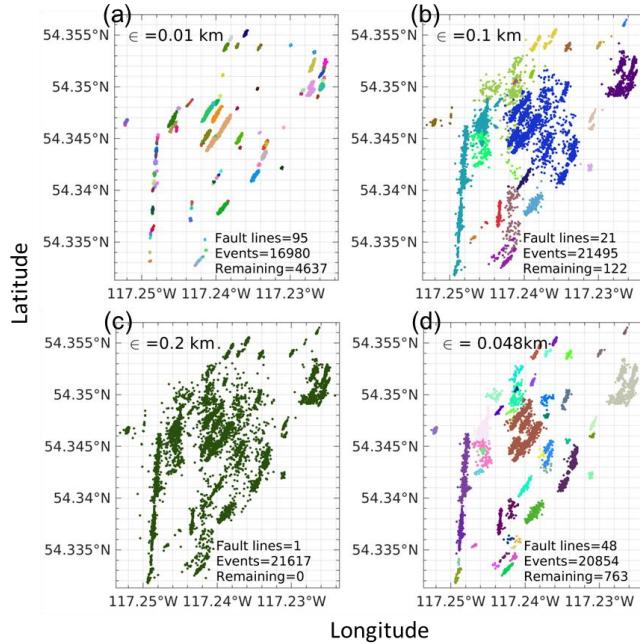


Figure S3. The comparison between magnitude-based and manual settings of DBSCAN in

ToC2ME. (a), (b) and (c) show the results with manual settings (i.e., ϵ of 0.01, 0.1, and 0.2 km, respectively). (d) illustrates the clustering result using the maximum magnitude ($M_w=3.21$).

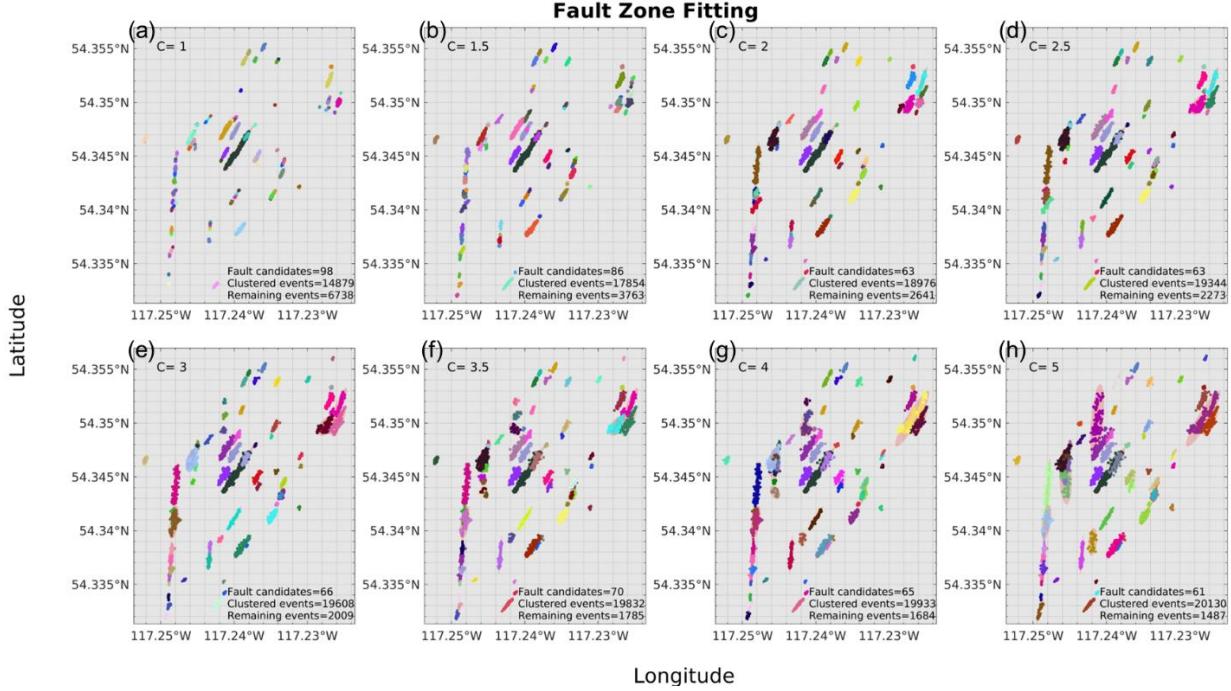


Figure S4. Fault zone fitting of ToC2ME when C -values range from 1 to 5. Each color represents an individual cluster/fault, and the red semi-transparent ellipsoid indicates the fitted fault zone.

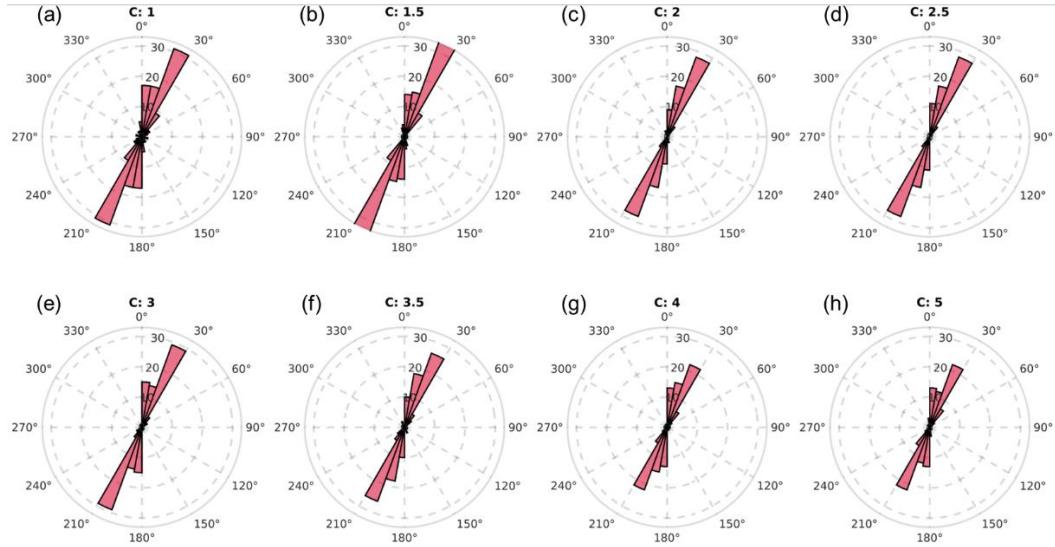


Figure S5. The statistics of fault strikes for C -values ranging from 1 to 5 indicate a dominant

orientation near N30°E, with a secondary trend in the N–S direction.

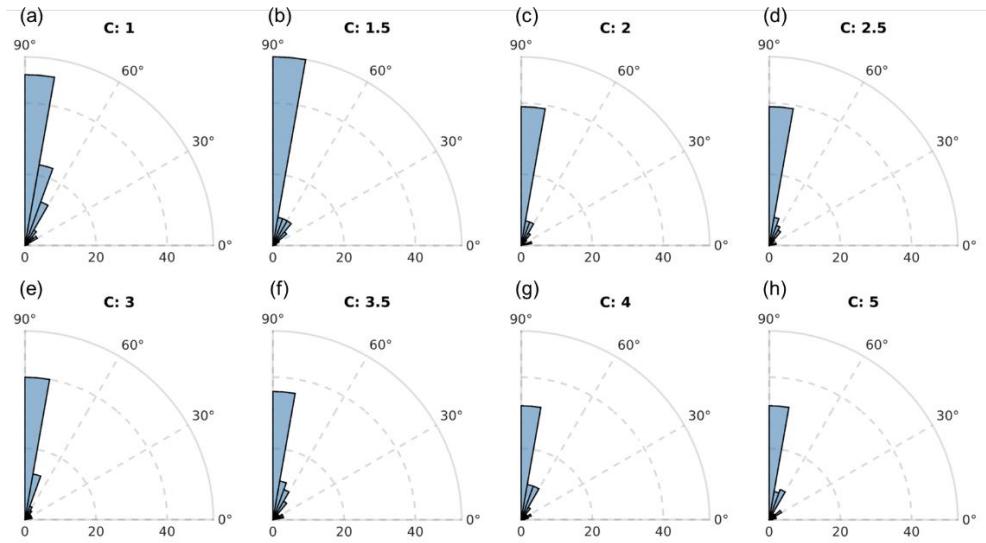


Figure S6. Fault dip angles for C -values ranging from 1 to 5. The faults are predominantly vertical.

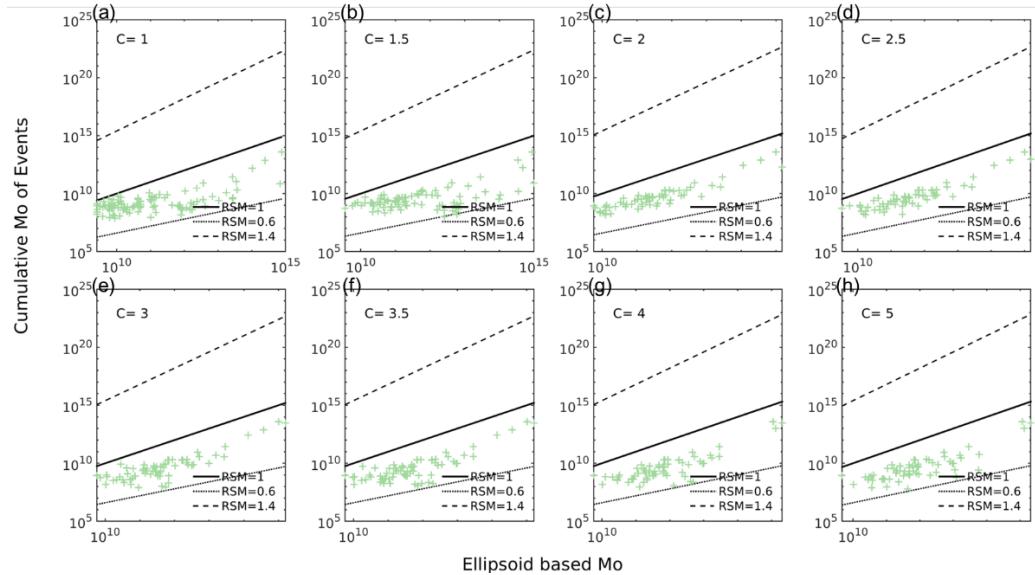


Figure S7. RSM values for C -values between 1 and 5 are shown, with each green cross representing an individual RSM between the events and the fitted fault. The majority of RSM values lie between 0.6 and 1.0.

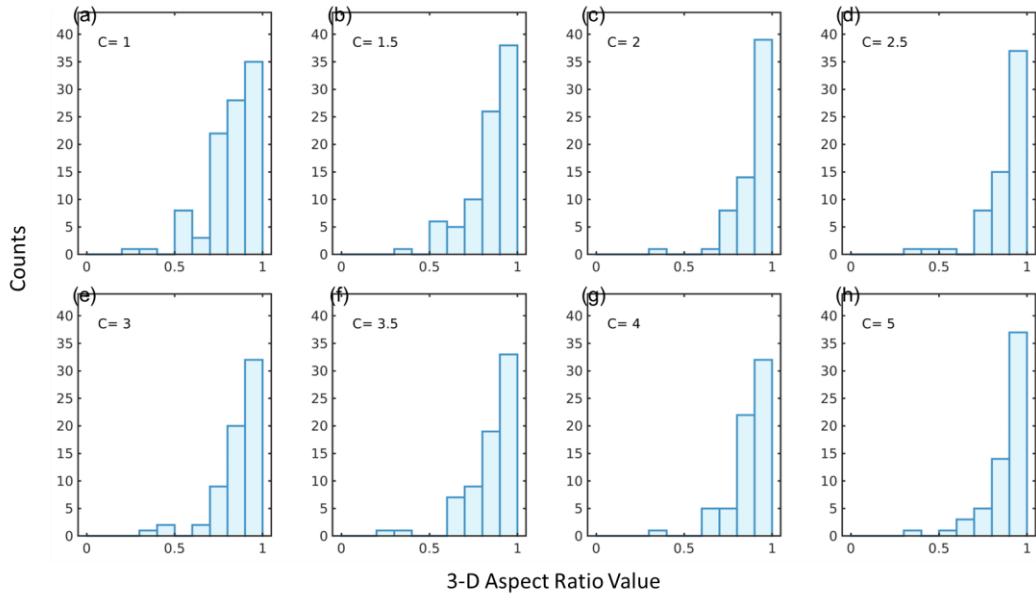


Figure S8. The 3-D Aspect Ratio varies from 0 (a sphere) to 1 (a simple line or plane), corresponding to C -values ranging from 1 to 5.

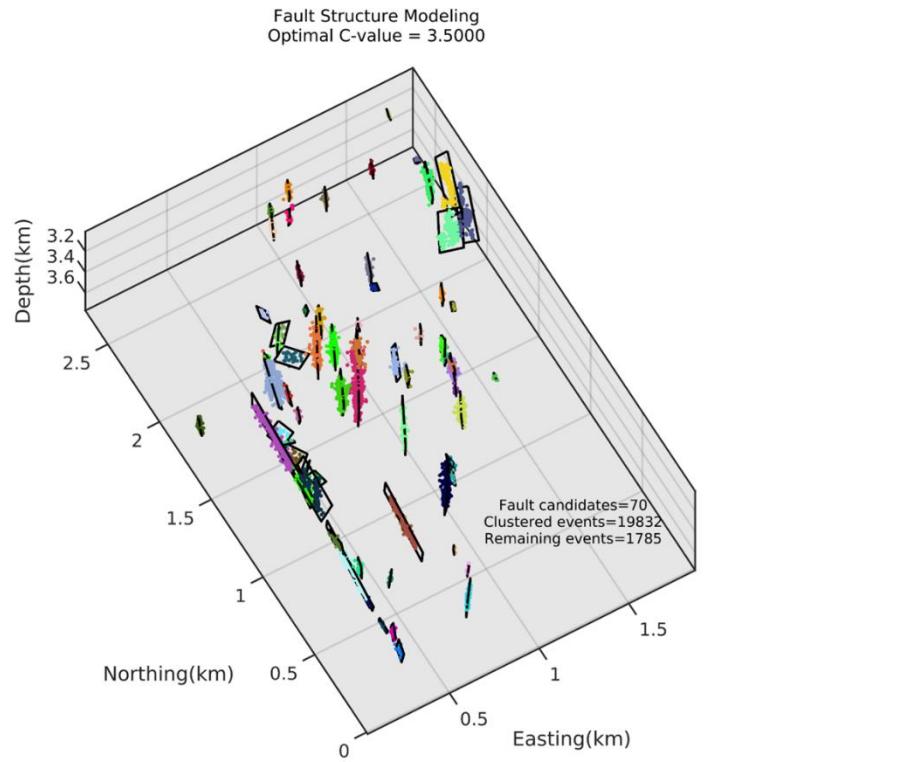


Figure S9. Fault structure modeling in the absolute distance coordinate system, using fault

geometric parameters derived with an optimal C of 3.5. Each color represents a fault/cluster,

while black rectangles represent the fault planes.

Text S2. Applying E2F to Oklahoma: Parameter Details and Evaluations

In such a complex conjugate fault system, due to the lack of prior information on fault strike

orientations, we used Mode 1 to provide the preferred strike orientation as input for Mode 2.

Specifically, we retained our initial run with Mode 1 for the Oklahoma case, as the 3-D Hough

transform didn't produce a bimodal distribution of fault orientations (i.e., orthogonal, Fig. S10g),

and some faults were categorized as "Unaccepted" (Fig. S10b). After running E2F once and

obtaining 55° and 115° as the preferred fault development directions in this area, we refined the

Oklahoma example using two preferred orientations (55° and 115°) with an uncertainty range of

$\pm 30^\circ$ (Fig. S10h) in Mode 2. Based on these inputs, we updated the lineament classifications

(Fig. S11g).

This section also provides additional analysis, including fault segment clustering (Fig. S12), fault

zone fitting via ellipsoid model (Fig. S13), the statistics of fault strikes (Fig. S14), the

distribution of dip angles (Fig. S15), the RSM (Fig. S16), and 3-D Aspect Ratio (Fig. S17), and

fault structure model based on the optimal C value of 35. (Fig. S18). As C increases: 1) the

conjugate faults-oriented NE–SW and SE–NW become more clearly (Fig. S12-S14); 2) event

utilization gradually increases, while the number of high-quality 3-D Aspect Ratio and

reasonable RSM values decreases since C of 20 (Fig. S18). We selected C of 35 and marked 14

sub-faults (A1-A14), corresponding to the faults identified by Park *et al.* (2022), as shown in Fig. S19. Specifically, E2F subdivided A9 into three more detailed fault segments (A9a-c).

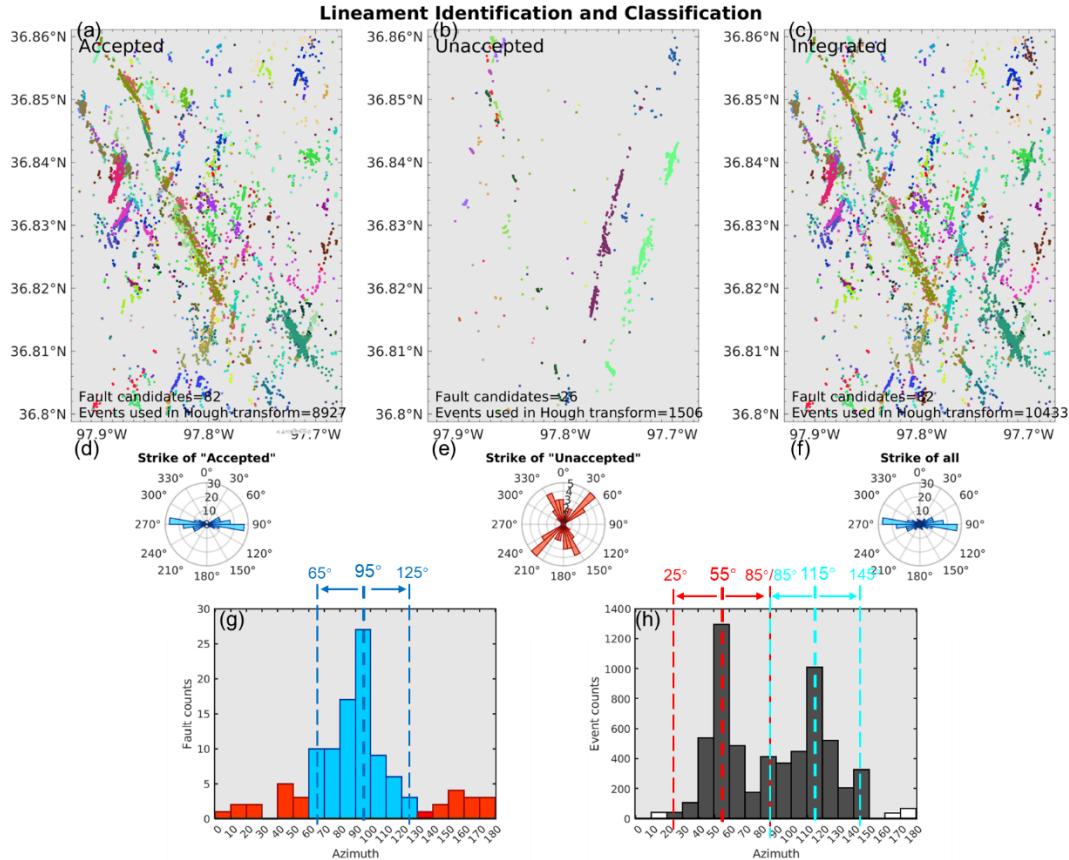
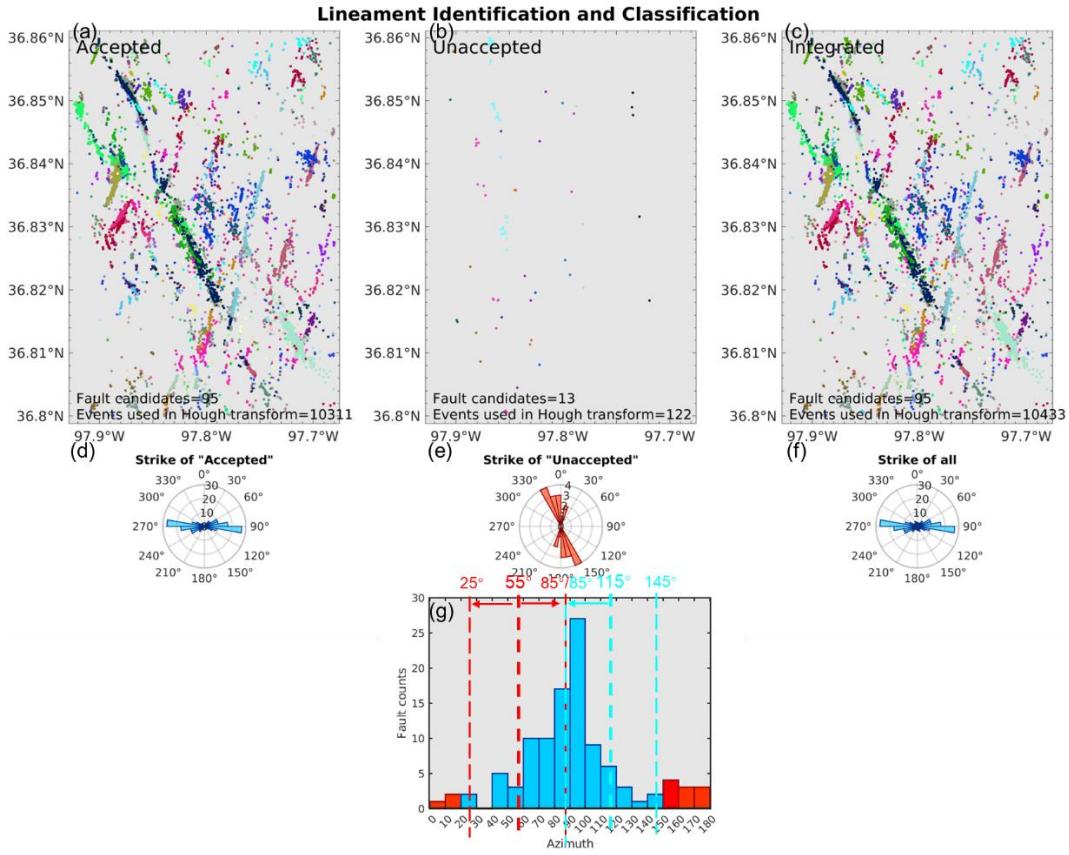


Figure S10. “Lineament identification and classification” using Mode 1. (a-c) The “Accepted”, “Unaccepted” and “Integrated” fault candidates, respectively, and (d-f) correspond to the distributions of strike. Each color represents an individual fault/cluster in (a-c). (g) Within 2 times PBAD ($\pm 30^\circ$) in Mode 1, the “Accepted” (labeled as blue) and “Unaccepted” (marked as red) fault candidates were automatically obtained by E2F. (h) The preferred fault orientations, 55° and 115° , were derived from Mode1 with C of 35. The two orientations, within a range of $\pm 30^\circ$ (i.e., 25° - 145° , marked as dark), were then used as input for Mode 2 (Fig. S11).



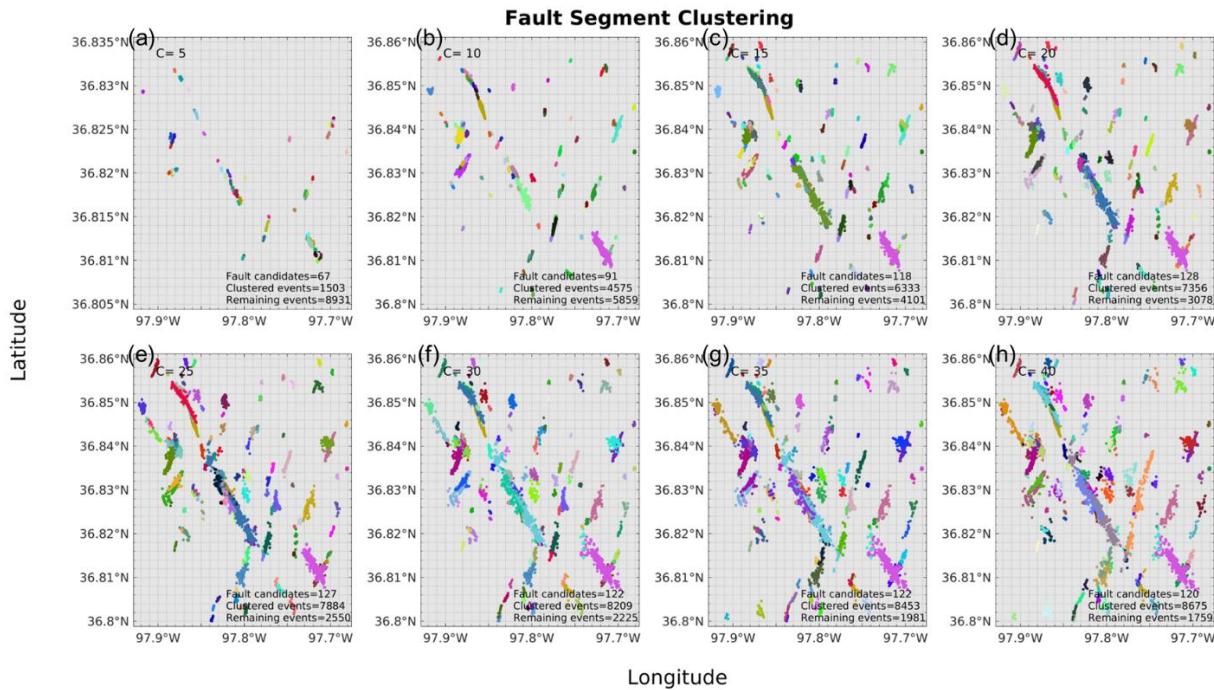


Figure S12. “Fault segment clustering” in the Oklahoma case. The C ranged from 5 to 40. Each color represents a cluster/fault. The number of faults increased from 67 to 128, while the number of events increased from 1503 to 8675.

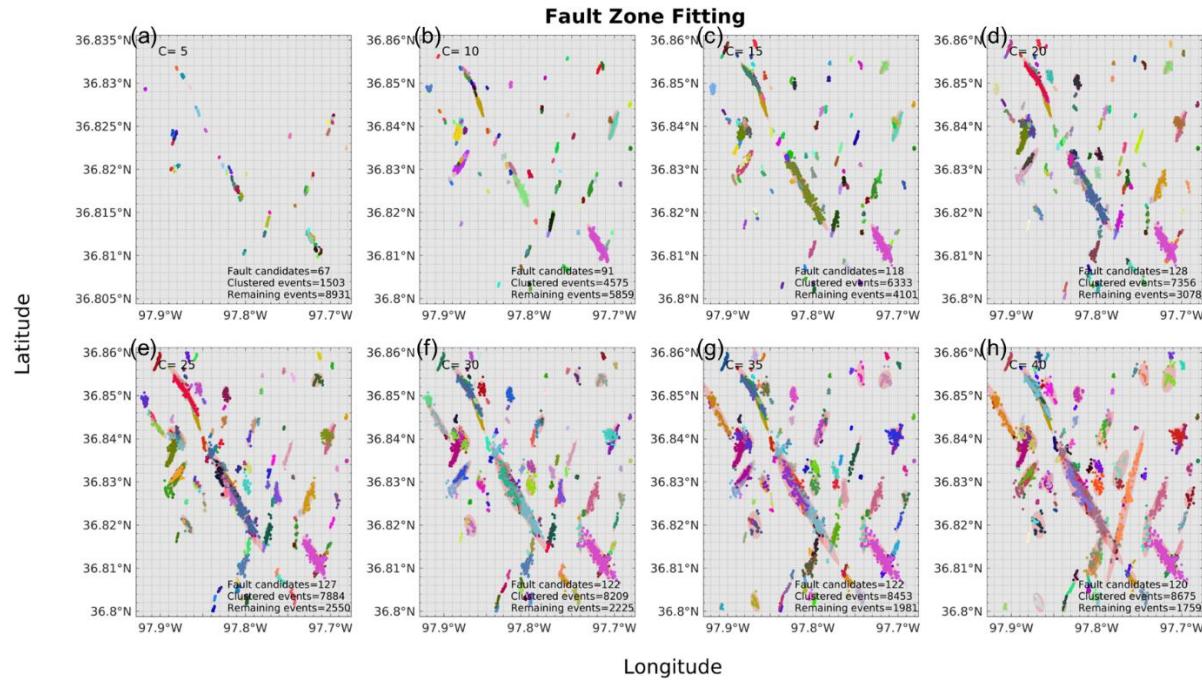


Figure S13. Fault zone fitting in the Oklahoma case when C -values range from 5 to 40. Each

color represents an individual cluster/fault, and the red semi-transparent ellipsoid indicates the fitted fault zone.

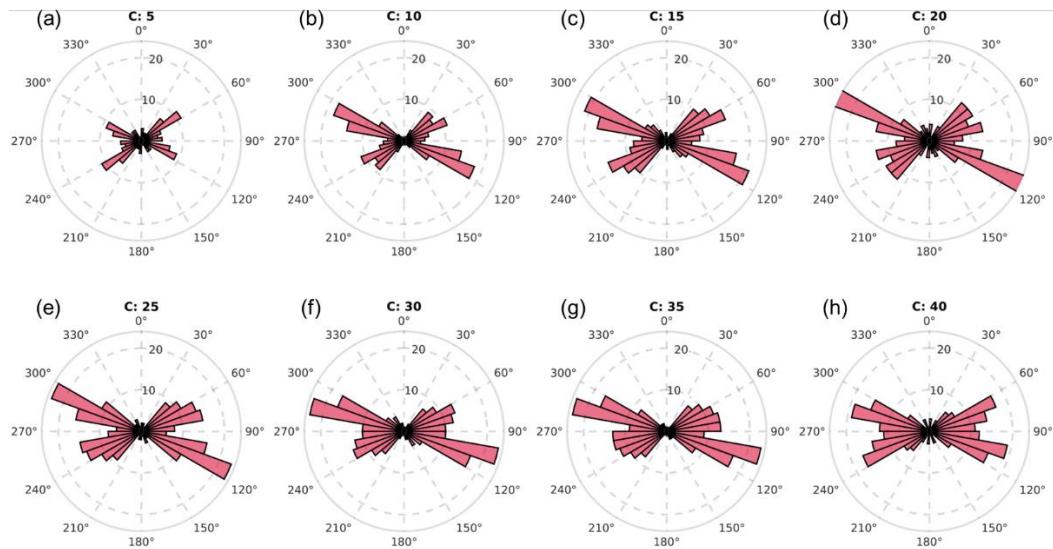


Figure S14. For C -values ranging from 5 to 40, the statistics of fault strike reveal a set of conjugate faults oriented approximately $N60^\circ E$ and $N60^\circ W$.

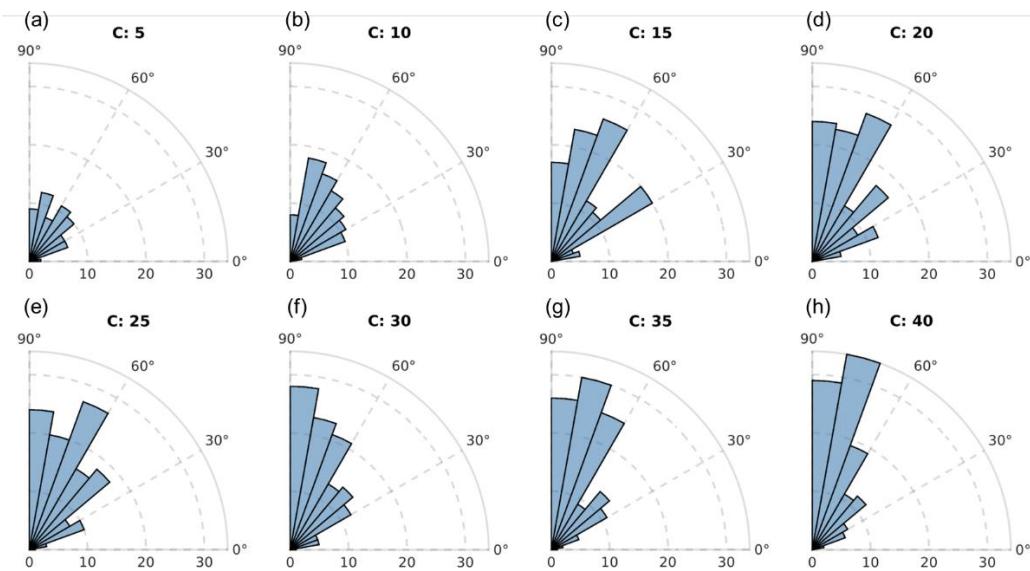


Figure S15. Fault dip angles for C -values ranging from 5 to 40. The dominant dip orientations are inclined at approximately 75° to the horizontal plane.

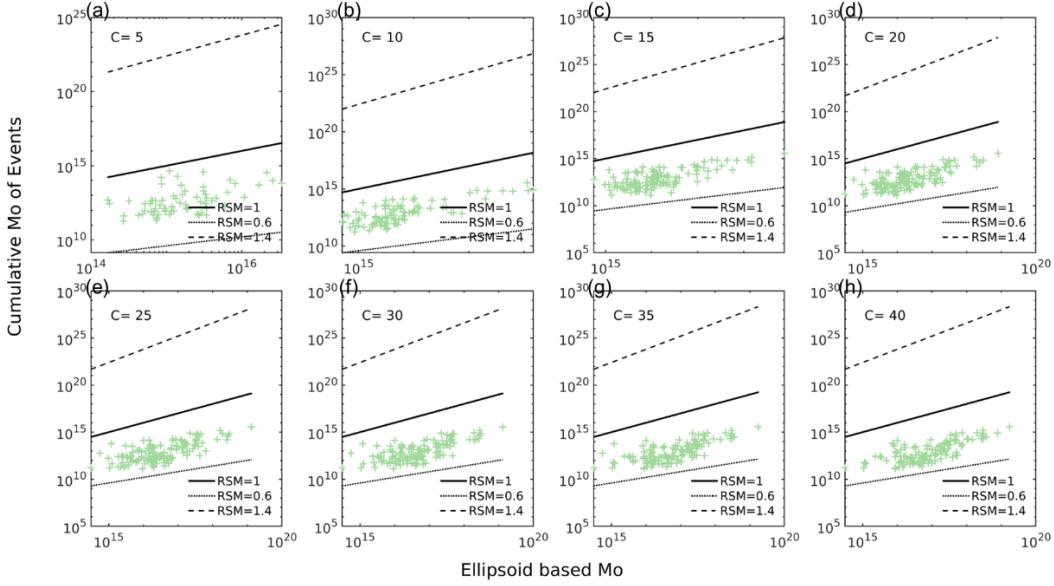


Figure S16. RSM values for C -values between 5 and 40 are shown, with each green cross representing an individual RSM between the events and the fitted fault. The majority of RSM values lie between 0.6 and 1.0.

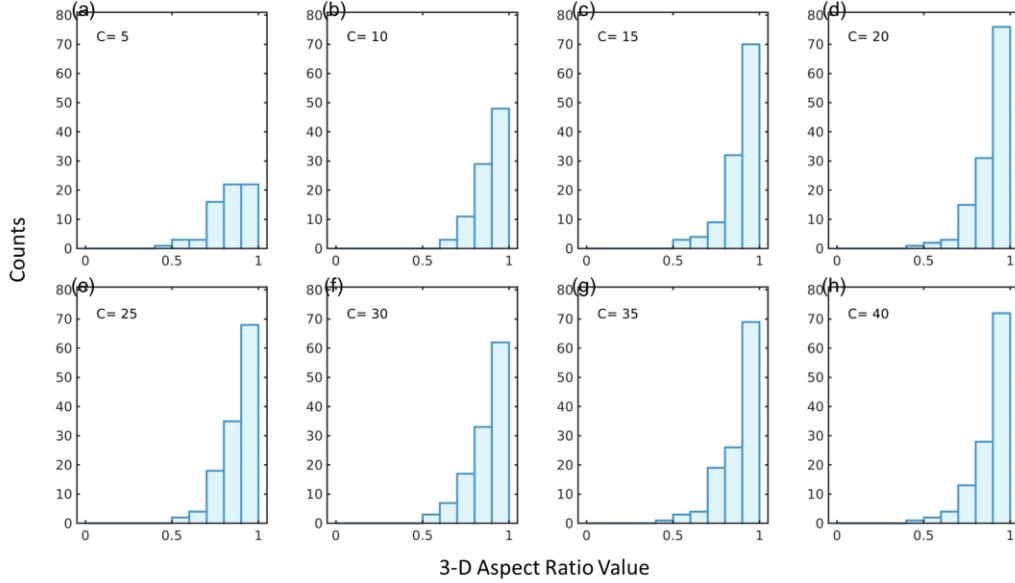


Figure S17. The 3-D Aspect Ratio varies from 0 (a sphere) to 1 (a simple line or plane), corresponding to C -values ranging from 5 to 40.

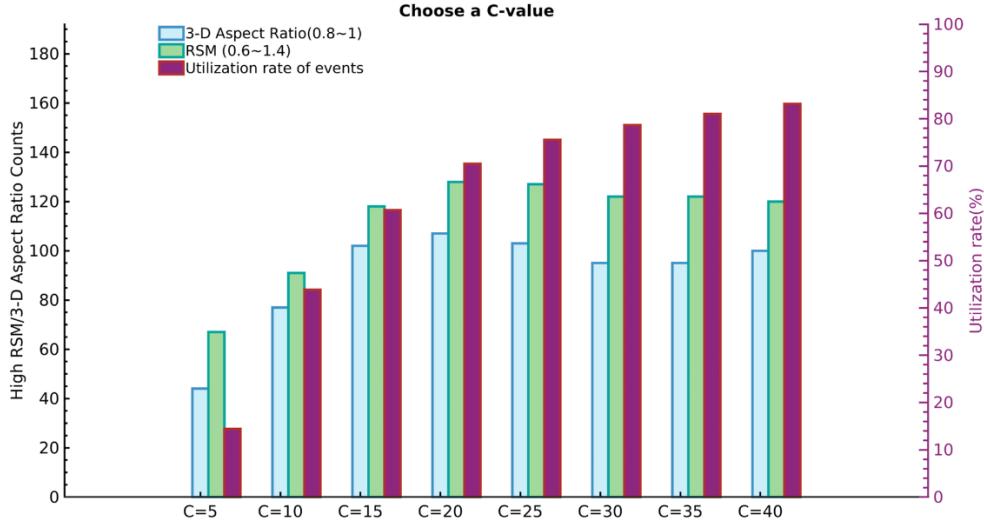


Figure S18. Variation in the number of high-quality 3-D Aspect Ratio (0.8–1), reasonable RSM (0.6–1.4), and the corresponding utilization rate of events.

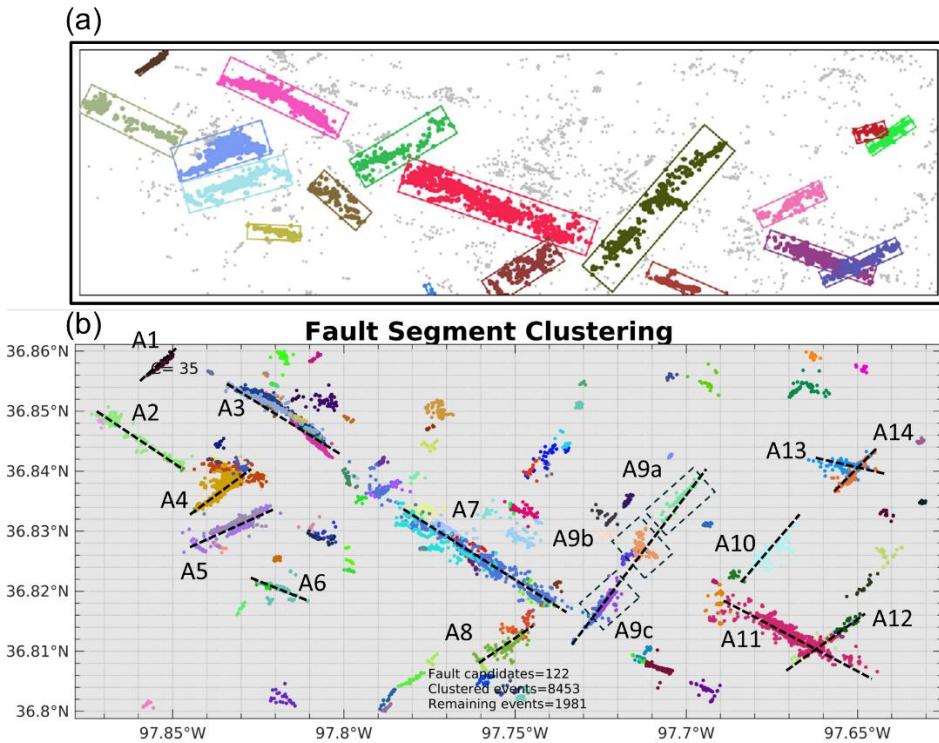


Figure S19. The fault classification comparison in Oklahoma. Each color represents an individual cluster/fault. (a) Faults in Oklahoma were extracted from the study by Park *et al.* (2022). (b) E2F identified 122 faults in Oklahoma, with A1-A14 corresponding to the different

fault colors illustrated in (a).

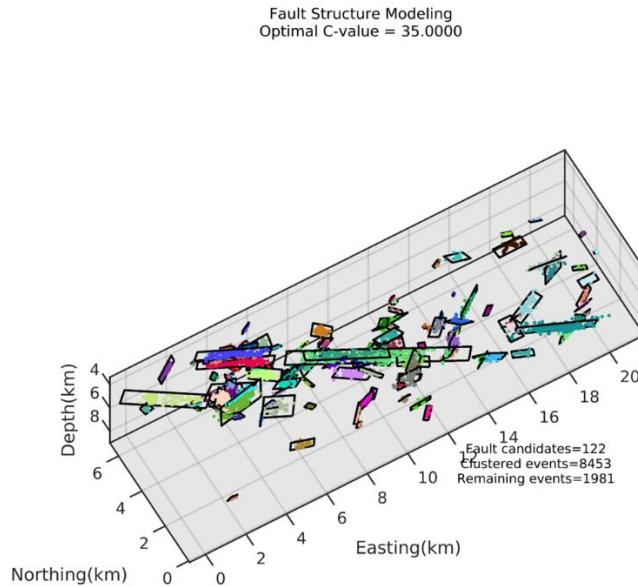


Figure S20. Fault structure modeling in the absolute distance coordinate system, using fault parameters derived with an optimal C of 35. Each color represents a fault/cluster, while black rectangles represent the fault planes.

Text S3. Applying E2F to Türkiye: Parameter Details and Evaluations

We tested the 2023 Türkiye earthquake sequence using Mode 1. This section provides the “Accepted” and “Unaccepted” fault candidates (Fig. S21), fault segment clustering (Fig. S22), fault zone fitting using ellipsoid model (Fig. S23), the statistics of fault strikes (Fig. S24), the distribution of dip angles (Fig. S25), the RSM (Fig. S26), and 3-D Aspect Ratio (Fig. S27), and fault structure model based on the optimal C value of 100. (Fig. S30). As C increases: 1) the number of clustered events increases (Fig. S22-S23); 2) event utilization gradually increases,

while the number of high-quality 3-D Aspect Ratio and RSM values decreases (Fig. S28). Notably, the change of C results in a limited difference in the total number of faults for the Türkiye case, as the catastrophic earthquake sequences usually involve several apparent faults that cover minor faults. In such relatively simple scenarios, adjusting for an optimal C is easier and less critical.

Based on event utilization, high-quality 3-D Aspect Ratio, and reasonable RSM values, we selected C of 100 compared with the previous geological survey of fault traces (Fig. S29). For the faults of Türkiye, E2F classified 8 planar sub-faults (B1-B8) that are consistent with previous studies. For instance, B1 belongs to the Amanos segment, and B2 corresponds to the Narlı segment of the Dead Sea Fault zone. B3-B8 correspond with the Pazarcık segment and Erkenek segment of the East Anatolian Fault zone, Savrun Fault, Çökak Fault, Çardak Fault, and Doğanşehir Fault zone of the Malatya Fault, respectively (Emre *et al.*, 2018).

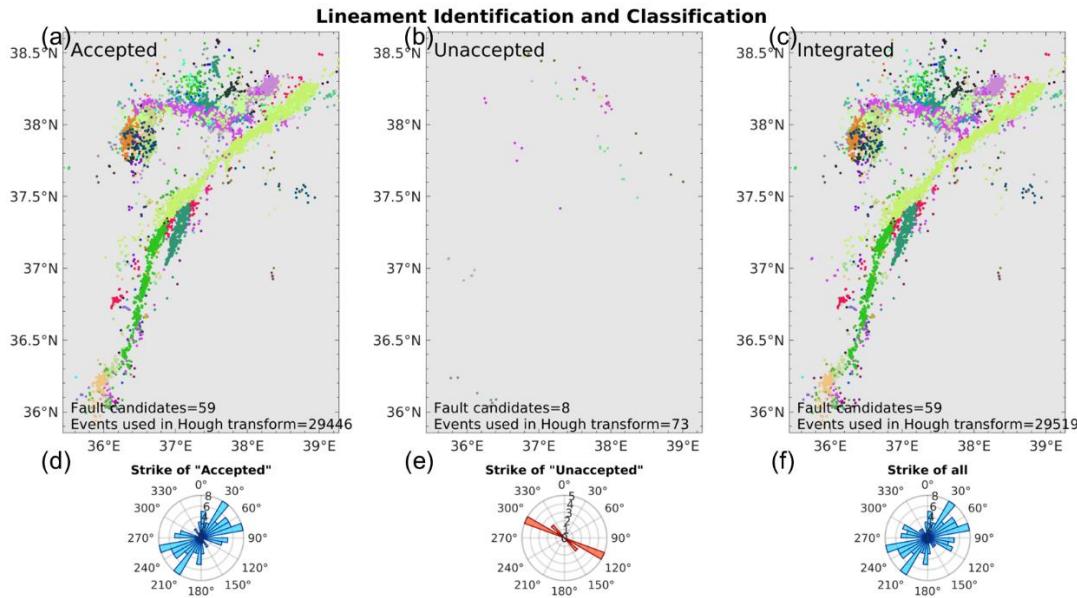


Figure S21. “Lineament identification and classification” in the Türkiye case using Mode 1. (a-c)

The “Accepted”, “Unaccepted” and “Integrated” fault candidates, respectively, and (d-f) correspond to the distributions of strike. Each color represents an individual fault/cluster in (a-c).

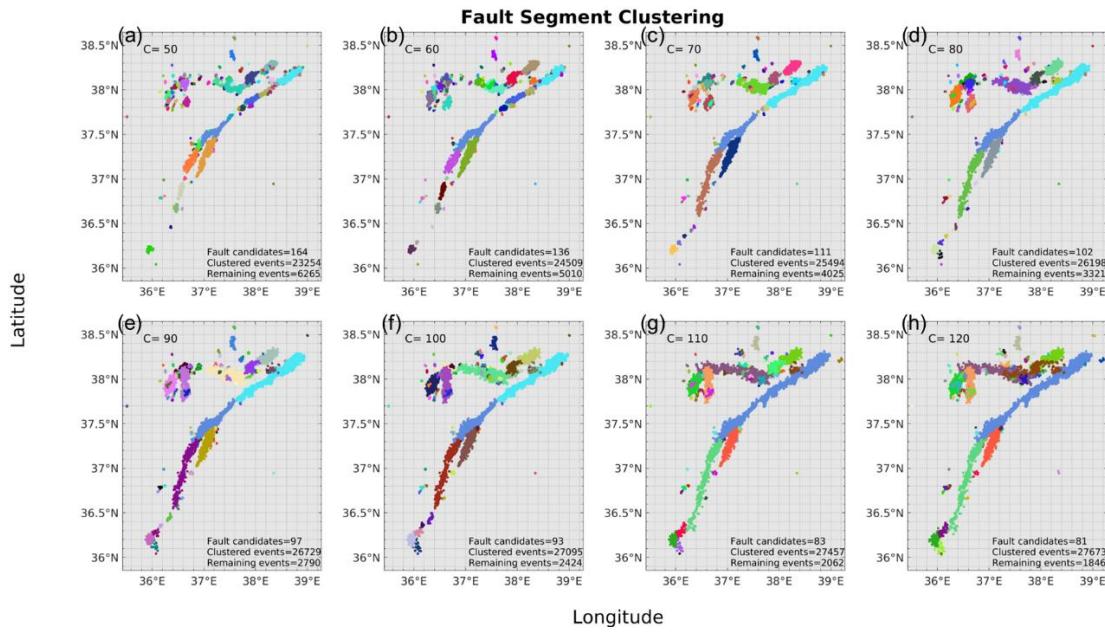


Figure S22. “Fault segment clustering” in the Türkiye case. The C ranged from 50 to 120. Each color represents a fault. The number of faults changed from 164 to 81, and the number of events varied from 23254 to 27673.

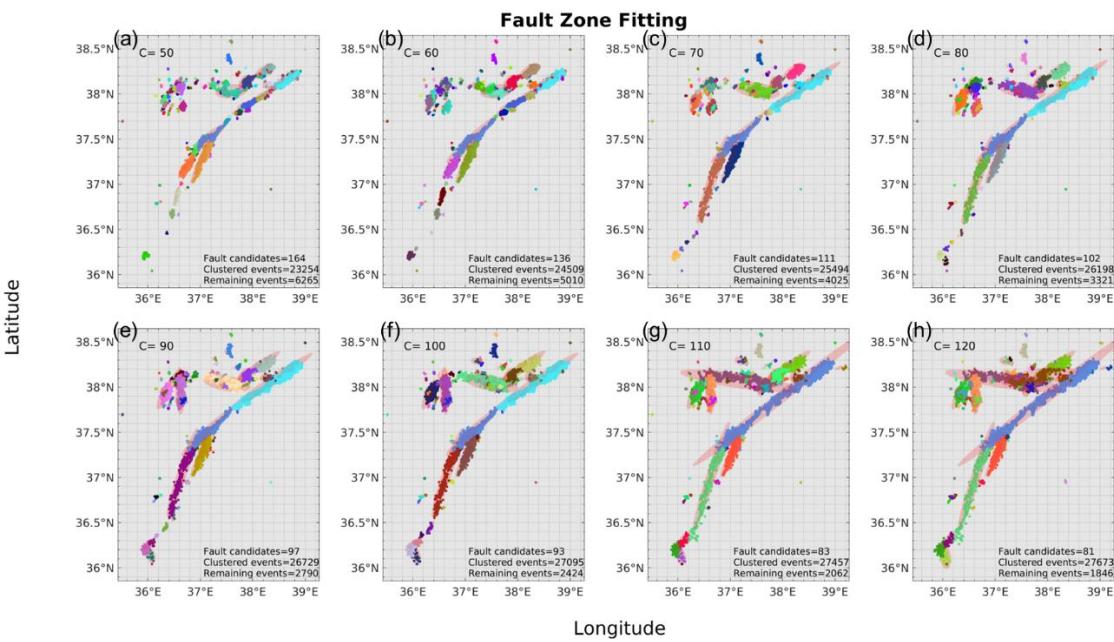


Figure S23. Fault zone fitting of the 2023 Türkiye case when C -values range from 50 to 120.

Each color represents an individual cluster/fault, and the red semi-transparent ellipsoid indicates the fitted fault zone.

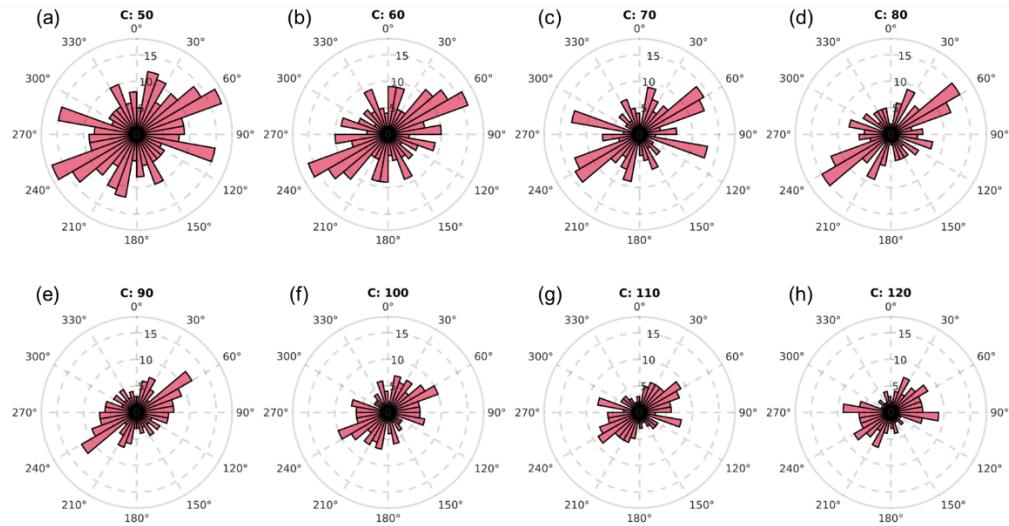


Figure S24. The statistics of fault strikes for C -values ranging from 50 to 120 indicate a dominant strike orientation near N60°E.

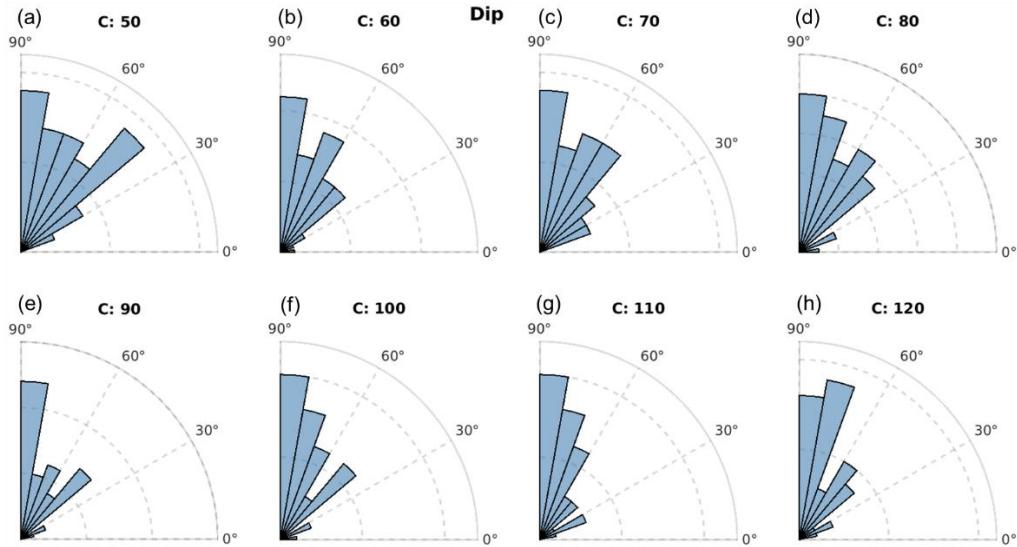


Figure S25. Fault dip angles for C -values ranging from 50 to 120. The dominant directions are almost perpendicular to the horizontal plane.

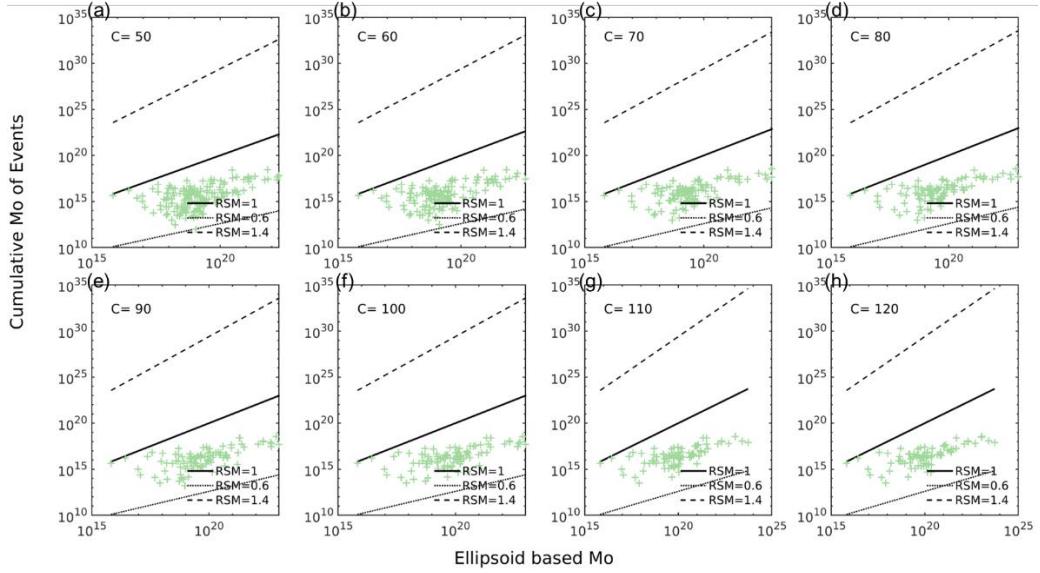


Figure S26. RSM values for C -values between 50 and 120 are shown, with each green cross representing an individual RSM between the events and the fitted fault. The majority of RSM

values lie between 0.6 and 1.0.

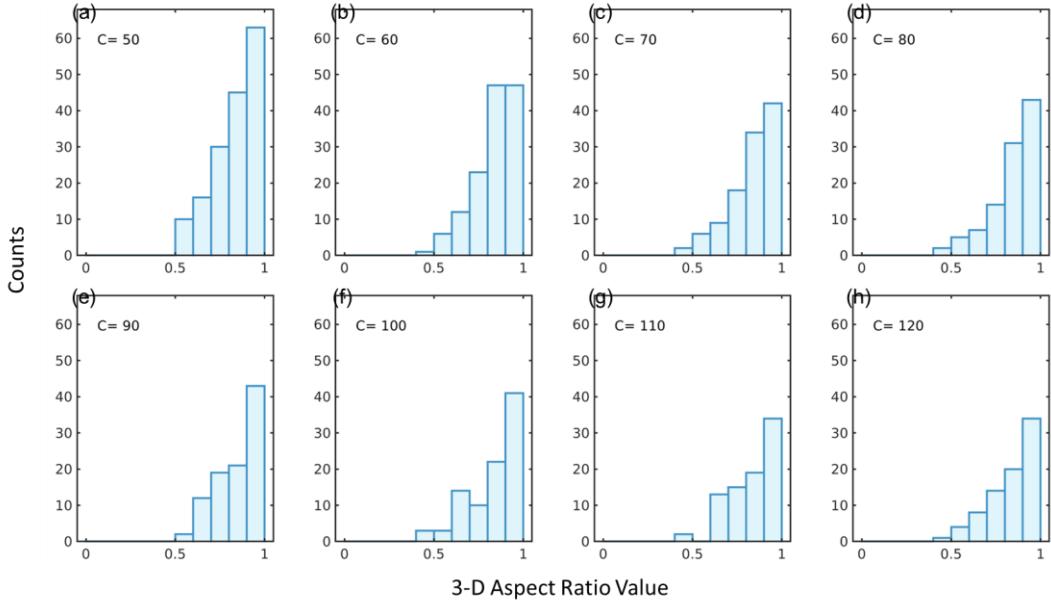


Figure S27. The 3-D Aspect Ratio varies from 0 (a sphere) to 1 (a simple line or plane), corresponding to C -values ranging from 50 to 120.

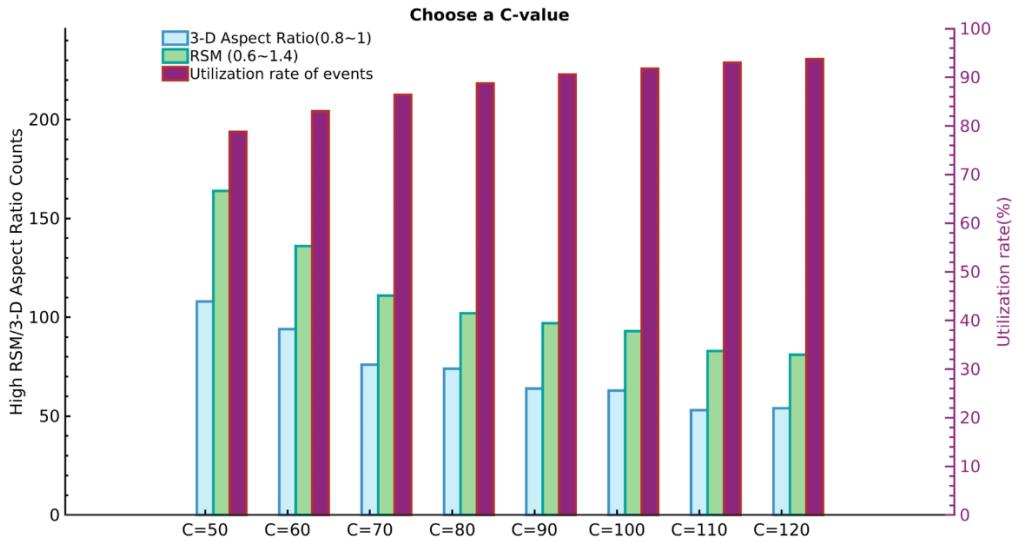


Figure S28. Variation in the number of high-quality 3-D Aspect Ratio (0.8–1), reasonable RSM (0.6–1.4), and the corresponding utilization rate of events.

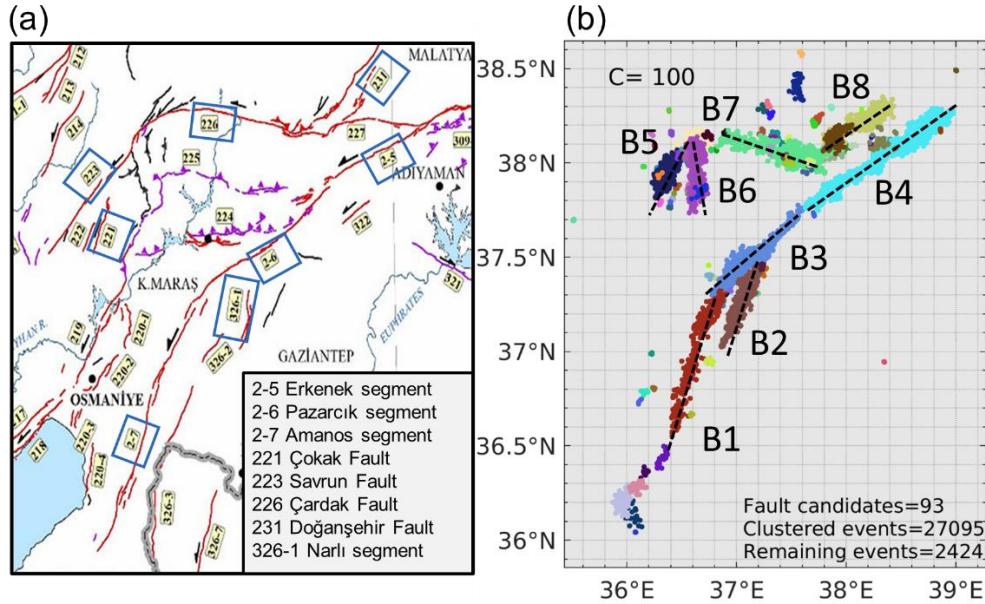


Figure S29. The fault classification comparison. (a) Spatial and temporal distribution of active faults in Türkiye, with each fault or segment assigned a unique identifier (marked in the right corner). This figure is cited from Emre *et al.* (2018). (b) E2F identified 93 faults in Türkiye, including B1–B8. Each color represents an individual cluster/fault.

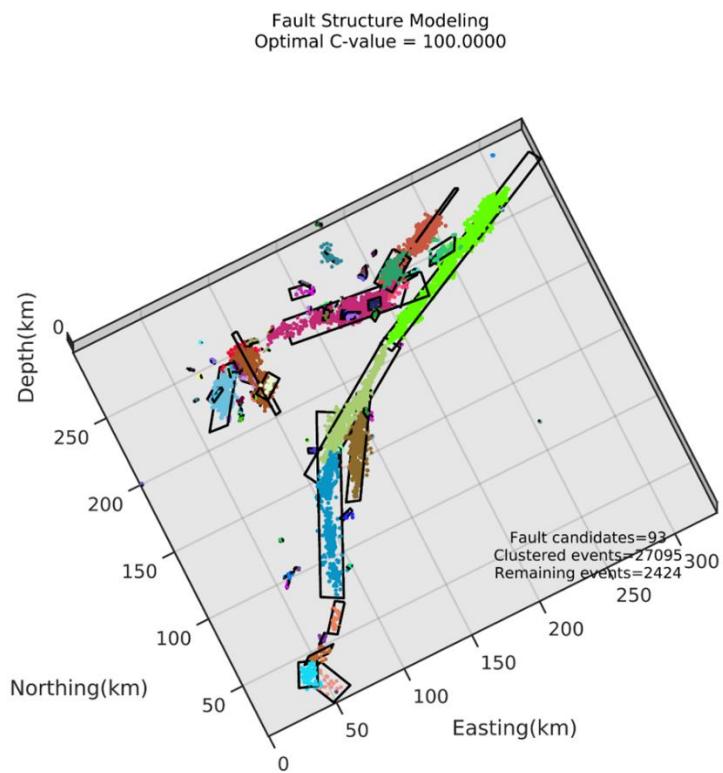


Figure S30. Fault structure modeling in the absolute distance coordinate system, using fault parameters derived with an optimal C of 100. Each color represents a fault/cluster, while black rectangles represent the fault planes.

Reference

Park, Y., G. C. Beroza, and W. L. Ellsworth (2022). Basement fault activation before larger earthquakes in Oklahoma and Kansas, *The Seismic Record*, 2, 197–206.

<https://doi.org/10.1785/0320220020>.

Emre, Ö., T. Y. Duman, S. Özalp, et al. (2018). Active fault database of Turkey, *Bulletin of Earthquake Engineering*, 16, 3229–3275. <https://doi.org/10.1007/s10518-016-0041-2>.