Electronic Seismologist

OPEN ACCESS

SURF: An Automated Method for Building Nonplanar 3D Fault Models from Earthquake Hypocenters

Travis Alongi*[1], Austin J. Elliott[1], Robert J. Skoumal[1], David R. Shelly[2], and Alexandra E. Hatem[2]

Abstract

Accurately characterizing 3D fault geometry is vital for improving our understanding of earthquake behavior and informing the development of seismic hazard models. Despite their importance, subsurface fault structures tend to be poorly constrained because of limitations in observational data. Improvements to the seismic networks and earthquake detection algorithms have increased the precision and volume of earthquake catalogs, which help illuminate detailed subsurface fault structure and provide the most direct information available about fault geometries at depth. We present a Python package to automate generating 3D fault geometries directly from hypocentral seismicity patterns. This method begins with clustering events based on their spatial density, identifying coherent patterns. Nearby clusters are then merged based on the similarity of their orientations. We fit nonplanar surfaces using support vector regression to balance surface accuracy with minimal deviations from planarity. The fault models are output as quadrilateral meshes at user-defined resolution. In the process of generating the 3D fault surfaces, we compute the spatial density of seismicity around the surface and the planarity as quantitative metrics of the model outputs.

As a proof of concept, we apply this approach to the San Andreas–Calaveras fault junction region and the 2019 Ridgecrest earthquake sequence, both in California, which contain complex subparallel faults well defined at the Earth's surface and abundant microseismicity. These case studies demonstrate the method's ability to model complex fault structures, including long continuous fault surfaces, crossing faults, variably dipping segments, and subparallel faults. We test the method on both standard network catalogs and double-difference relocated catalogs. We find that our seismicity-based fault model results align with published 3D models that incorporate additional constraints and interpretations (Plesch *et al.*, 2020; Aagaard and Hirakawa, 2021). This workflow provides a low-user-input solution for estimating fault geometries at depth from earthquake catalogs.

Cite this article as Alongi, T., A. J. Elliott, R. J. Skoumal, D. R. Shelly, and A. E. Hatem (2025). SURF: An Automated Method for Building Nonplanar 3D Fault Models from Earthquake Hypocenters, *Seismol. Res. Lett.* **XX**, 1–17, doi: 10.1785/0220250126.

Supplemental Material

Introduction

The spatial distribution and magnitude of earthquakes depend on 3D fault structure, yet it is rarely known in detail. Earthquake hazard assessments rely on representations of fault locations and areas (Field *et al.*, 2014). Currently, the representations of faults, in many cases, are based on mapped surface traces and have limited dip information. This presents a problem because natural faults are nonplanar, are multistranded, and have variable dip both along strike and depth (Segall and Pollard, 1980; Carena *et al.*, 2002; Fletcher *et al.*, 2016; Ulrich *et al.*, 2019; Chu *et al.*, 2021; Alongi *et al.*, 2022, 2024). Furthermore, faults are often geometrically complex, featuring roughness, stepovers, and curvature (Willemse and

---

1. Earthquake Science Center, U.S. Geological Survey, Moffett Field, California, U.S.A., https://orcid.org/0000-0002-0865-8064 (TA); https://orcid.org/0000-0001-5924-7268 (AJE); https://orcid.org/0000-0002-5627-6239 (RJS); 2. Geologic Hazard Science Center, U.S. Geological Survey, Golden, Colorado, U.S.A., https://orcid.org/0000-0003-2783-5158 (DRS); https://orcid.org/0000-0001-7584-2235 (AEH)

*Corresponding author: talongi@usgs.gov

Pollard, 2000; Candela *et al.*, 2012; Candela and Brodsky, 2016) that can play an essential role in earthquake nucleation, propagation, and arrest (Poliakov *et al.*, 2002; Wesnousky, 2006; Lozos *et al.*, 2011; Hatem *et al.*, 2015; Cattania and Segall, 2021). Limited knowledge of 3D fault geometry restricts the predictive capabilities of earthquake hazard models and highlights the need for improved methods that can estimate fault geometry at depth.

Existing methods

The importance of 3D fault models has driven efforts such as the Statewide California Earthquake Center's Community Fault Model (CFM), which uses data that include surface traces, hypocentral distributions, well penetrations, seismic reflection profiles, and geologic cross sections to build 3D fault surfaces (Plesch *et al.*, 2007). However, these data-intensive models are often limited to well-studied regions such as southern California and require interpretive expertise. In less well-studied areas, seismicity trends can provide a uniform and data-driven alternative for identifying faults at depth.

Previous studies have explored various methods for determining 3D fault orientations from seismicity. Ouillon *et al.* (2008) laid the groundwork for this using clustering and principal component analysis to define planes. Subsequent work introduced Gaussian mixture models, probabilistic clustering, and Bayesian information criterion to improve fault detection accuracy (Wang *et al.*, 2013). Other studies have incorporated stochastic plane-fitting approaches to overcome shortcomings inherent to regression (e.g., L1- and L2-norm fits) as a way to better represent faults (Kaven and Pollard, 2013; Skoumal *et al.*, 2019). Other work has used Monte Carlo simulations to account for hypocentral uncertainty and focal mechanisms to assess fault orientation reliability (Truttmann *et al.*, 2023). Recently, the normal vectors of the nearest-neighbor events have been included in the clustering step (Sawaki *et al.*, 2025) as well as strain-rate-based fault reconstructions (Jourdon *et al.*, 2025). Although these methods have advanced fault characterization, they produce numerous small, disconnected planes, limiting their ability to capture larger, more continuous structures. Importantly, to our knowledge, existing methodologies assume planarity, preventing them from accurately representing the nonplanarity of typical fault surfaces. Our method directly addresses these limitations by building larger faults from small clusters of seismicity and applying a flexible nonplanar surface fitting technique.

Fault zones

Although nonplanar surfaces improve the representation of faults, it remains important to evaluate whether the modeled surfaces conform to what is known about fault zone structure. Fault zones are recognized as volumes of inelastic deformation, recorded both geologically—through features such as secondary fractures, cracks, and deformation bands—and geophysically via observations of rock properties and seismic velocity structure at depth (Caine *et al.*, 1996; Fialko, 2004; Li *et al.*, 2004; Chester *et al.*, 2005; Cochran *et al.*, 2009; Choi *et al.*, 2016; Alaei and Torabi, 2017; Scott *et al.*, 2018; Antoine *et al.*, 2021; Atterholt *et al.*, 2022; Cocco *et al.*, 2023). A consistent observation across studies is that the intensity of deformation, often approximated by fracture density or seismicity, tends to decay with distance from the principal slip zone (Shipton and Cowie, 2001; Berg and Skar, 2005; Faulkner *et al.*, 2010; Powers and Jordan, 2010; Savage and Brodsky, 2011; Perrin *et al.*, 2021; Alongi *et al.*, 2022, 2024; Rodriguez Padilla, Oskin, *et al.*, 2022). These patterns have been described using exponential or power law functions and provide a quantitative framework for assessing off-fault deformation. In this study, we use the decay of seismicity density with distance from a modeled fault surface as a proxy for deformation localization, enabling an appraisal of how well our surface-fitting method reproduces known fault zone characteristics.

This study

Many previous efforts to generate 3D fault representations have largely concentrated on improving clustering techniques and addressing hypocentral aleatory uncertainty. Here, we introduce the Seismicity Utilized for Reconstructing Faults (SURF) algorithm (Alongi and Skoumal, 2025) (see Data and Resources) for determining fault geometries from seismicity patterns. We build on prior work (Ouillon *et al.*, 2008; Kaven and Pollard, 2013; Wang *et al.*, 2013; Skoumal *et al.*, 2019), introducing a low-user-input approach to estimate fault geometries from seismicity patterns. Our method focuses on generating fault surfaces from numerous small, nearly coplanar clusters, which are then merged based on nearest-neighbor coplanarity. This approach allows us to model faults that are spatially continuous with variable geometry over large distances. We fit 3D nonplanar, semiparametric surfaces to the data, and the events near the modeled surface do not impact the misfit and effectively account for a finite damage zone width and location uncertainty. The result is a set of nonplanar fault geometries generated from seismicity patterns and requiring minimal user input. The detailed fault geometries and locations can be used to model rupture scenarios and may contribute to improving hazard assessments in both well-studied and underresearched regions.

Methods

The SURF algorithm is designed to build on and overcome the limitations of previous methods (Ouillon *et al.*, 2008; Kaven and Pollard, 2013; Wang *et al.*, 2013; Skoumal *et al.*, 2019), which produce numerous small and purely planar fault surfaces. Inspired by traditional mapping approaches, our method identifies clusters of seismicity based on spatial density and merges nearby clusters with similar orientations to create longer and more continuous faults. Unlike prior techniques that simplify fault geometries as planes, our approach

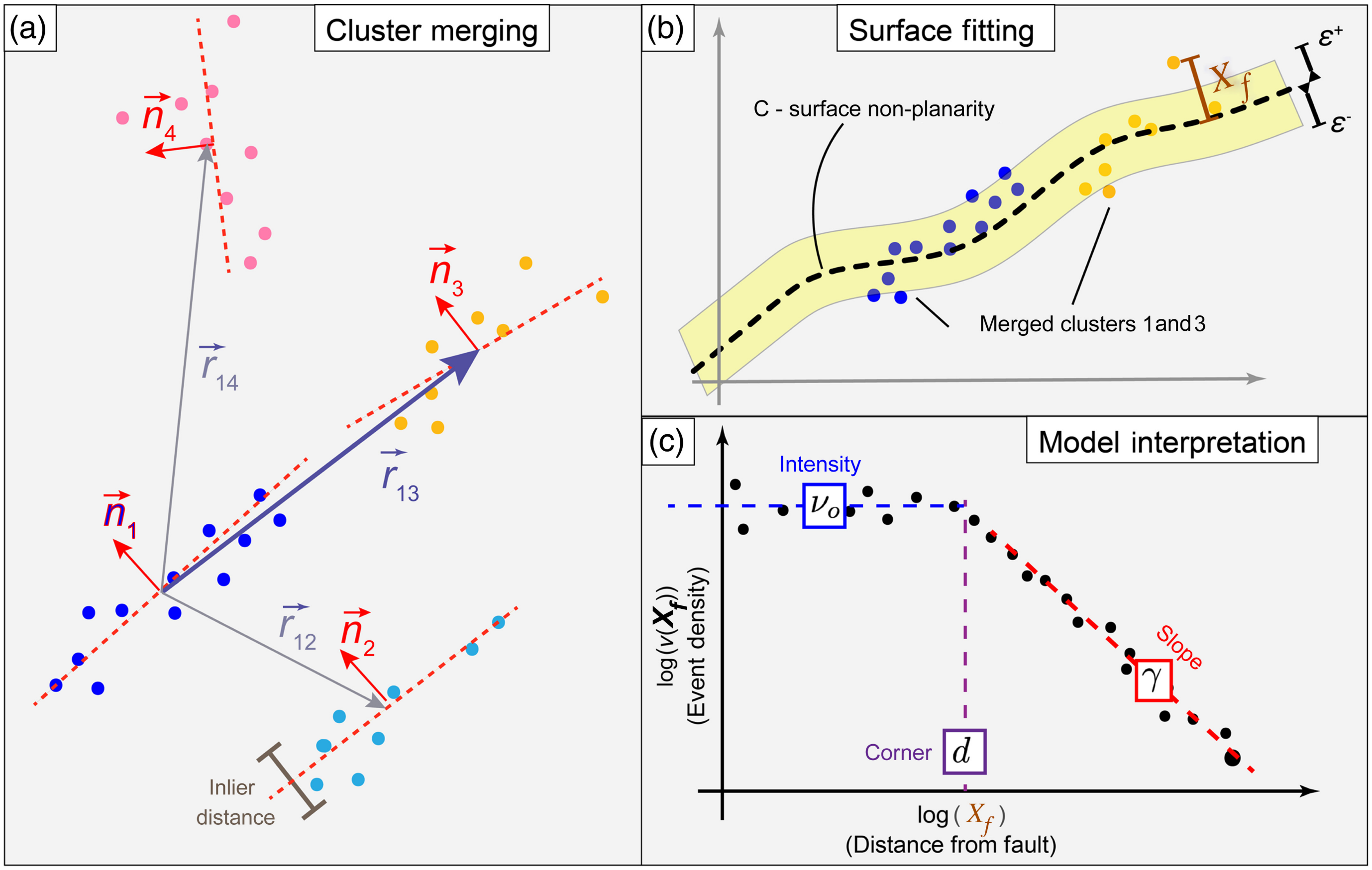


Figure 1. Schematic illustrating the key aspects of the seismicity utilized for reconstructing faults (SURF) algorithm. (a) The earthquake cluster merging process. First, each initial cluster (indicated by color) is fit with a plane using random sampling consensus to determine its normal vector $\vec{n}_i$. Next, the k, nearest-neighbor clusters ($k = 3$ in this example) are identified, and the direction vector $\vec{r}_{ik}$ to the kth cluster is calculated. Finally, normal vectors and direction vectors are used to construct the measurement matrix **M** (from equation 3), which quantifies the coplanarity of neighboring clusters and informs the merging process; see the main text for more details. (b) An illustration to help build intuition about the support vector regression (SVR) parameters C and ϵ. C describes how nonplanar the modeled fault surface is, with higher values indicating greater nonplanarity. Here, ϵ describes the length scale near the surface where points do not impact the fit. (c) A schematic to build intuition about the quality indicator parameters in equation (4). The equation fits a modified power law to the earthquake density as a function of distance from the fault surface (in log space). The color version of this figure is available only in the electronic edition.

independently fits each cluster with a nonplanar surface, allowing for more realistic fault geometries without imposing prior assumptions about their shape (Fig. 1).

Density-based clustering

The first step is to generate a spatial framework to build the faults around that is tied to active structures that are identified by where seismicity is spatially localized and clustered. To identify clusters, we use the hierarchical density-based spatial clustering of applications with noise (HDBSCAN) algorithm (Campello *et al.*, 2013) from the scikit-learn Python package (Pedregosa *et al.*, 2011). HDBSCAN is chosen for its ability to identify diverse, fault-like shapes, and it requires only one parameter, the minimum number of events to be considered a unique cluster. We find that setting the minimum cluster size to between 20 and 40 events ensures that clusters can be oriented while remaining small enough to merge in the next step. Note that clusters typically contain many more events than the minimum number, and this depends on the dataset. In our tests, the median number of events per cluster in is typically around 70–90.

Cluster merging

To extend fault lengths, we merge clusters with similar orientation that are nearly coplanar. Refer to Figure 1 for a schematic of the following steps.

Orientation assessment. To assess the coplanarity of clusters, we determine their orientations using a planar approximation. Here, we use a 3D implementation of the random sampling consensus algorithm (Fischler and Bolles, 1981). This stochastic method iteratively selects three earthquake hypocenters to create two vectors and compute their cross product, providing trial plane-normal vectors. Next, to assess which iteration provides the best fit, points are classified as inliers or outliers based on the distance from the trial plane. In this study, we use a distance threshold of 500 m, which is a typical absolute hypocentral depth uncertainty in California (typically greater than the horizontal location uncertainty) (Thurber *et al.*, 2003). The iteration with the greatest number of inliers is selected as the best fit. These best-fit orientations and the direction vectors to its nearest neighbors (calculated using k-dimensional [KD] tree; Bentley, 1975) provide a measure of coplanarity.

Coplanarity determination. To quantitatively assess coplanarity, we use the cluster's nearest-neighbor's normal vectors and direction vectors to calculate the following:

$$\vec{N} = \vec{n}_i \cdot \vec{n}_{jk}, \tag{1}$$

$$\vec{D} = \vec{n}_i \cdot \vec{r}_{jk}. \tag{2}$$

Here, $\vec{N}$ is the scalar dot product of the normal vector for the ith cluster with the normal vector j of k nearest neighbors. The value of $\vec{N}$ describes how similar the orientations of each cluster are with its nearest neighbors. $\vec{D}$ is the scalar dot product of the ith normal vector with the direction vector to k nearest neighbors; this quantity describes the coplanarity of pairs of clusters. Combining equations (1) and (2) provides a measure of coplanarity given by

$$\mathbf{M} = |\vec{N}| \times (1 - |\vec{D}|), \tag{3}$$

M ranges from 0 to 1, in which a value of 1 means that the pair i, k clusters are coplanar and 0 indicates that clusters are orthogonal, are on an adjacent parallel feature, or a combination of the two. We put a threshold on **M** to determine which clusters to combine. Clusters that do not meet the threshold are retained but are neither discussed nor displayed in this study.

Surface fitting

After merging clusters, we again use scikit-learn (Pedregosa *et al.*, 2011), this time to fit nonparametric surfaces to the data using support vector regression (SVR), a machine learning method based on support vector machines (SVMs) (Smola and Schölkopf, 2004; Bishop and Nasrabadi, 2006). Originally developed for classification, SVMs identify optimal decision boundaries by maximizing the margin between data classes. SVR extends this approach to regression tasks, predicting continuous values from inputs while allowing a margin of tolerance epsilon (ϵ) around the true values. The goal is to find a function that is as flat as possible while allowing deviations from the data within a specified distance ϵ and points outside this margin are penalized. The trade-off between flatness and tolerance for deviation is controlled by the parameter C. A small ϵ allows tighter fits to the data, whereas a larger C permits more complex surfaces that better capture outliers at the cost of smoothness.

We chose SVR for several reasons. It is nonparametric, meaning the surface is not restricted to a predefined shape and is instead guided directly by the data, avoiding assumptions about fault geometry. In addition, SVR's objective of maximizing flatness leads to smooth and simple surfaces that reflect the minimum complexity required to fit the data. During fitting, earthquakes that fall within the distance ϵ do not affect the regression, reducing sensitivity to location uncertainty and to diffuse seismicity within the broader damage zone. Moreover, the parameters C and ϵ can be selected through cross validation—a procedure in which the data are partitioned into training and validation sets to identify the values that generalize best to unseen data—minimizing the need for user tuning and reducing the risk of overfitting. Finally, the resulting surface can be discretized at any desired resolution and extrapolated beyond the data, making it a flexible tool for modeling fault geometry.

Model interpretation and hypocentral distributions

To help interpret the SURF algorithm's outputs, we measure the distribution of seismicity around the modeled fault surfaces. Measuring the hypocentral event density as a function of distance from the modeled surface may provide insights into the suitability or appropriateness of representing a volume of seismicity as a discrete fault surface. Specifically, we fit the distributions of seismicity around the modeled surfaces with a modified inverse power law relationship used in previous studies (Powers and Jordan, 2010; Perrin *et al.*, 2021; Rodriguez Padilla, Oskin, *et al.*, 2022; Rodriguez Padilla and Oskin, 2023). For each fault surface, we fit the equation first introduced in Powers and Jordan (2010) given by

$$v(x) = v_o\left(\frac{d^m}{|x|^m + d^m}\right)^{\frac{\gamma}{m}}. \tag{4}$$

Following Rodriguez Padilla, Oskin, *et al.* (2022), we use logarithmically spaced bins for x, in which $v(x)$ is the number of earthquakes in each bin, normalized by the bin width. The parameter d represents the distance from the fault where the inverse power law applies. The parameter m, controlling the sharpness of the distribution's corner at d, is set to 2, which is consistent with previous studies (Powers and Jordan, 2010; Rodriguez Padilla, Oskin, *et al.*, 2022). The power law exponent gamma (γ) describes the nonlinear decrease in event density with increasing distance from the fault to build some intuition (see the illustration in Figure 1). We fit equation (4) using maximum-likelihood estimation (MLE), assuming that $v(x)$ follows a Poisson distribution (Boettcher and Jordan, 2004; Powers and Jordan, 2010).

We evaluate the goodness of fit using mean relative error (MRE), which normalizes the misfit at each point by its estimated value,

$$\mathrm{MRE} = \frac{1}{n}\sum_{i=1}^{N}\frac{|v_{\mathrm{obs},i} - v_{\mathrm{fit},i}|}{v_{\mathrm{fit},i}}, \quad (5)$$

in which v_{obs} represents the data and v_{fit} is the MLE fit of the power law to the data. This provides a quantitative measure of the fit quality and is one component in discrimination between modeled faults with seismicity patterns that adhere to the power law scaling from those that do not. In addition to using the error in the fit to assess the modeled fault quality, we compare the parameters of the fit γ and d with what has been documented previously. Furthermore, we use the parameter C and ϵ from the SVR fit to provide information about the modeled surface as additional way to assess the quality. For more details about how modeled faults may be assessed, please refer to the Interpretation and evaluation section.

Results and Discussion

We applied the SURF algorithm to two different locations in California to test the robustness and versatility of our method. We find that the algorithm successfully identified long and continuous plate-boundary faults where epicenters are closely aligned with mapped surface traces. In addition, we find that documented subparallel fault strands are identified with the algorithm. The modeled results also feature several previously undocumented smaller secondary and variably oriented faults that are delineated by the seismicity patterns at depth.

San Juan Bautista, California

Our first test case focuses on the region south of the San Francisco Bay Area near the San Andreas fault (SAF) and Calaveras fault (CF) junction (Fig. 2). These two major faults creep variably within the bounds of the study region (e.g., Lienkaemper *et al.*, 2014). Two other faults are documented as active within the past 15,000 yr in the U.S. Geological Survey (USGS) Quaternary Fault and Fold Database (see Data and Resources), the Sargent fault and the Quien Sabe fault, and have seismicity associated with them (Fig. 2). This location features abundant well-located seismicity, and we used ~80,000 double-difference relocated hypocenters over 1984–2021 the time period provided by the Waldhauser and Schaff (2021) catalog. (The geographic boundary encloses about 10% of the total catalog.)

Running the catalog through the SURF workflow, described in the Methods section, we set the minimum cluster size to 40, which results in 357 unique clusters. We then merge clusters by comparing the plane orientation and directions to nine nearest neighbors using equation (3) and merging clusters that have an **M** value >0.75. This orientation merging step results in 13 consolidated clusters that are displayed as red surfaces in Figure 2 and Video S1, available in the supplemental material to this article. Only 34 of the initial clusters were not coplanar with their nearest neighbors and therefore not merged or shown.

Comparisons. SURF successfully identifies the four faults that have been active within the past 15,000 yr, each of which has 3D representations in two existing fault models. One model is the USGS's Bay Area Velocity Model (BAVM) (Aagaard and Hirakawa, 2021), which features detailed 3D faults created by integrating surface geology, seismicity, and gravity data. This model serves as the basis for the current iteration of the Statewide California Earthquake Center's (SCEC) CFM 7.0 (Plesch *et al.*, 2007). The second model is the third Uniform California Earthquake Rupture Forecast (UCERF3), which provides simplified representations of USGS quaternary fault trace points that are extruded to depth at fixed dip and dip direction (Field *et al.*, 2014). Although UCERF3 offers general agreement with our findings, its level of detail is lower than that of the BAVM/CFM.

We evaluate the consistency between our modeled fault surfaces and the CFM because it provides a comparable level of detail and a meaningful basis for evaluation. To accomplish this comparison, we resample the SAF and CF modeled fault surfaces from this study at 50 m increments. We apply the same gridding process to the CFM's CF and the composite San Andreas sections "Creeping," "Loma Prieta," and "Southern Peninsula." We then calculate the distance between points on our modeled fault surface and the points on the existing fault model. These distances are then binned, and we subsequently fit using the equation

$$N = \kappa 10^{-x/x_0}, \quad (6)$$

in which x is the interfault distance, x_0 is the decay distance that characterizes the distribution, and κ is a constant that is related to the count at the fault location. The decay distance x_0 represents the characteristic length scale over which the counts decrease, such that ~90% of all points are less than x_0 from the fault. Figure 3 highlights the distribution of distances between these results and SCEC CFM. What stands out is that our results closely match those generated by the SCEC CFM experts, even though we rely solely on seismicity. These figures show the distribution of distances between our results and the CFM's SAF and CF. In both cases, the models agree within 1 km, a value substantially lower than what might be expected given the lack of shallow (<4 km depth) seismic constraints.

Ridgecrest, California

Our next test case focuses on the 2019 Ridgecrest earthquake sequence, which included an M_{w} 6.4 earthquake followed by an M_{w} 7.1 event that occurred the next day. Previous studies have generated relocated earthquake catalogs that have informed

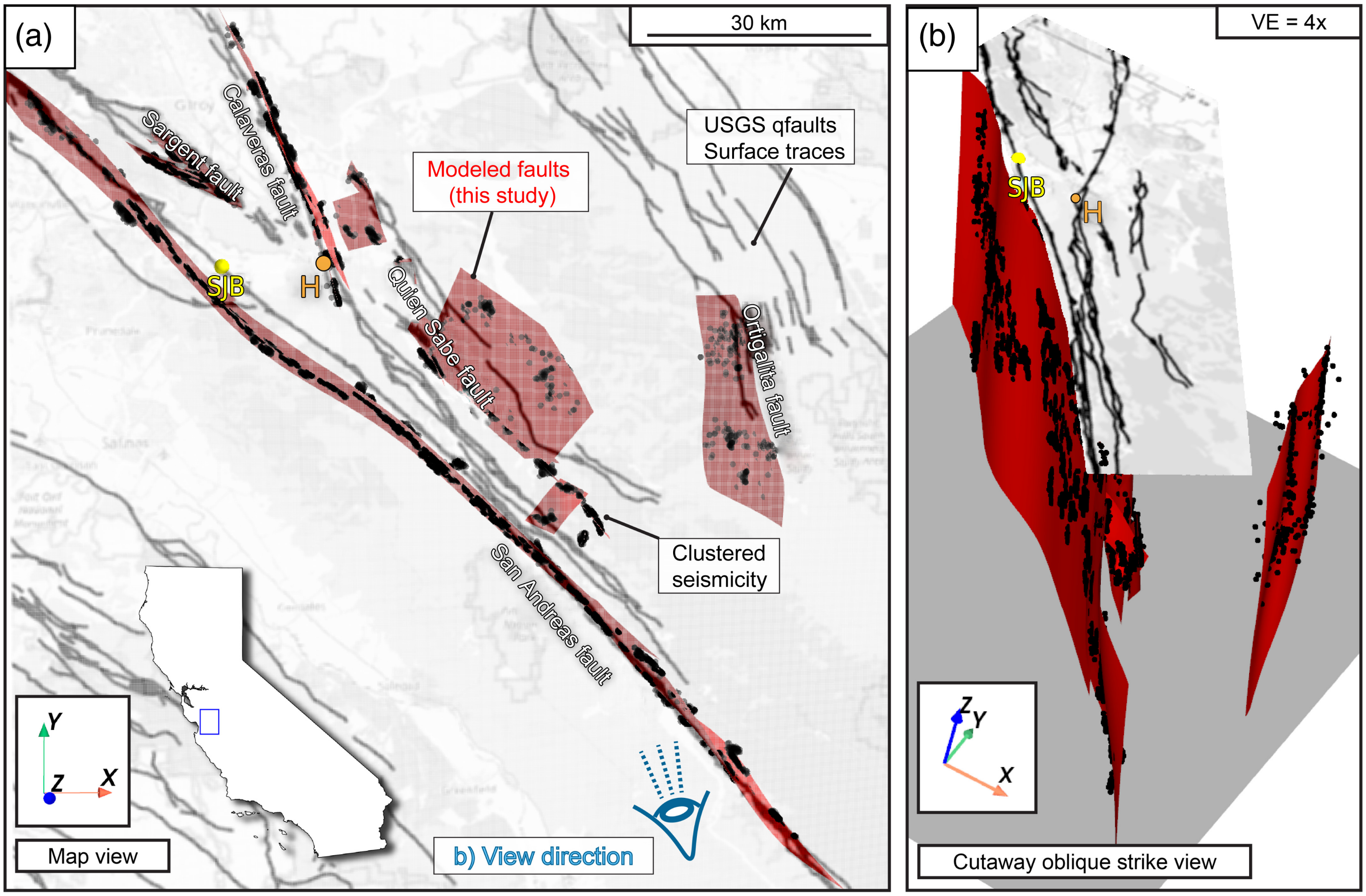


Figure 2. 3D rendering of the modeled fault surfaces in the San Juan Bautista region, California. The red surfaces represent modeled faults, and the black lines indicate the U. S. Geological Survey's Quaternary Fault and Fold Database, where the labeled fault zones are active in the latest Quaternary (<15,000 yr) and are included as fault sources in U.S. National Seismic Hazard Model (Petersen *et al.*, 2024). The black dots mark epicentral locations used to develop the fault planes. Towns mentioned in the main text, San Juan Bautista (SJB) and Hollister (H), are denoted with yellow and orange, respectively. (a) Map view of fault surfaces, where the semitransparent base map is overlain to highlight the agreement between modeled faults and the location of the mapped surface traces. The inset map shows the study region in California. The blue eye indicates the viewing direction for (b). (b) An oblique cutaway view looking northwest with 4× vertical exaggeration (VE) to highlight the geometry of the San Andreas fault. See Video S1 for a rotating animation of this model, which provides a complete 360° perspective on fault geometry and seismicity. The color version of this figure is available only in the electronic edition.

detailed interpretations of fault geometries at depth. A unique aspect of this sequence is the near-orthogonal orientation of the activated fault structures (Liu *et al.*, 2019; DuRoss *et al.*, 2020; Pollitz *et al.*, 2020; Fialko and Jin, 2021; Nevitt *et al.*, 2023), which poses challenges for distinguishing intersecting faults using focal mechanisms because the nodal planes are also orthogonal. The spatial distribution of aftershock hypocenters also reflects orthogonal features (Ross *et al.*, 2019; Lomax, 2020; Shelly, 2020b).

For this study, we use the double-difference relocated earthquake catalog by Shelly (2020a), which includes the foreshock sequence and the first 10+ days of the aftershocks following the M 7.1 mainshock. This catalog was generated using ~13,500 cataloged events as waveform templates, detecting and locating >34,000 events that provide insight into this complex fault system. We apply the SURF algorithm to the catalog and use similar parameters as those in the San Juan Bautista example. However, in this case, we set the minimum cluster size to 35, resulting in 100 unique clusters. These initial clusters are then combined and merged by comparing orientations and direction to the five nearest neighbors as described by equation (3). Clusters with **M** >0.75 were merged. This process resulted in 15 merged clusters; 34 remain unmerged. The merged results are then fit and are displayed as a warm color map (indicating depth) in Figure 4a and red surfaces in Figure 4b,c.

Comparison. The resulting fault model shown in Figure 4a demonstrates general agreement with the surface rupture trace locations from DuRoss *et al.* (2020). The modeled fault surfaces that correspond to the foreshock surface rupture traces (southwest section in Fig. 4a) exhibit steep and variable dips. In this region, the modeled faults highlight the method's

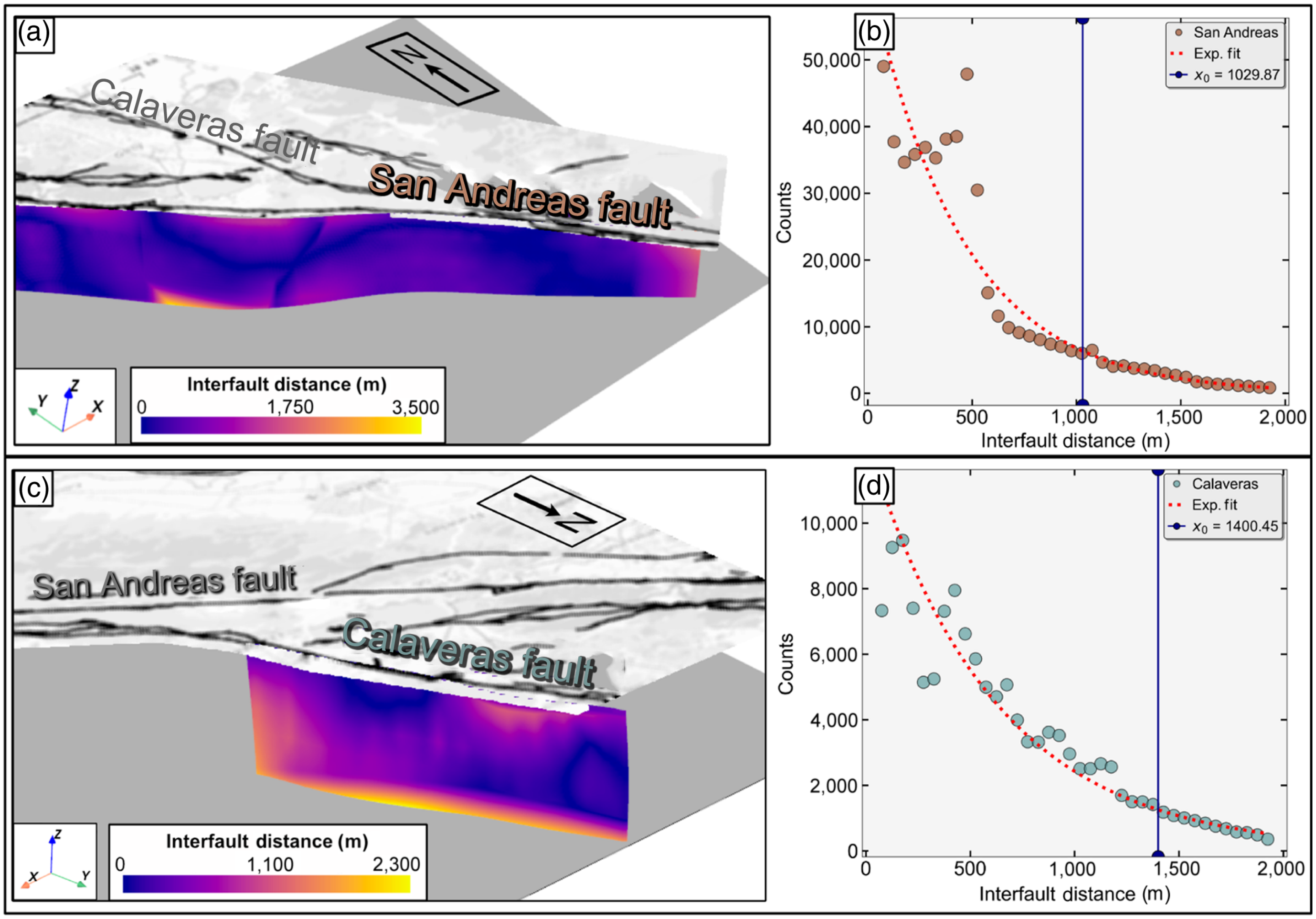


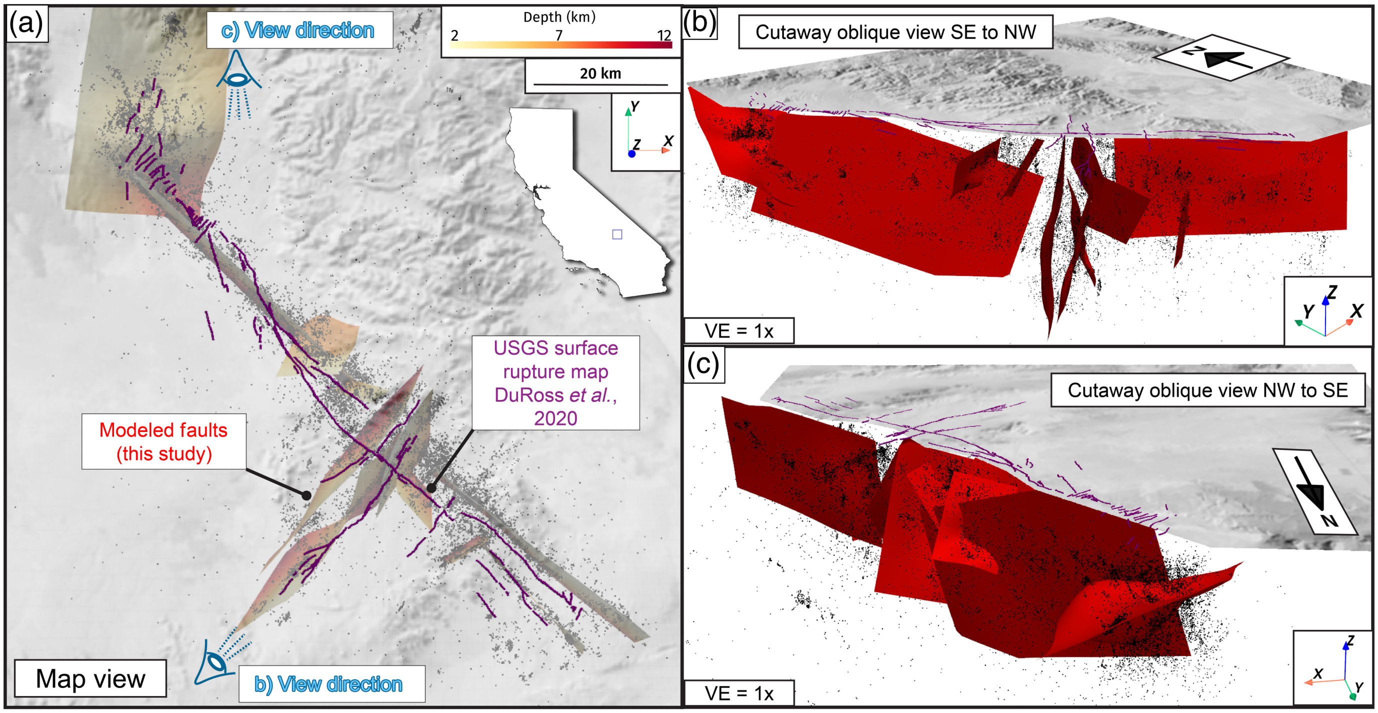
Figure 3. Comparison of the modeled fault surfaces from this study with the Statewide California Earthquake Center's Community Fault Model (CFM) v.7.0 (Plesch *et al.*, 2007), which approximates the Bay Area Velocity Model's fault surfaces (Aagaard and Hirakawa, 2021). Both datasets were resampled to a uniform 50 m node spacing, and the minimum distance between the faults was calculated. (a,b) Results for the San Andreas fault. (c,d) Results for the Calaveras fault. (a,c) The modeled fault result rendered, where the faults are colored from purple to yellow, indicating the distance between the modeled result and the CFM. (b,d) The distribution of interfault distances these results versus CFM and fit with exponential from equation (6) in which x_0 indicates the decay distance. The color version of this figure is available only in the electronic edition.

ability to delineate subparallel faults, as illustrated in Figure 4b and Video S2. Further evidence of these subparallel fault delineations can also be seen in the southeasternmost extent of the mainshock rupture, where modeled faults are offset from the surface mapping by 1–2 km. In addition, it is notable that the modeled fault of the mainshock is not continuous at the intersection with the foreshock, as can be seen in Figure 4b.

The most significant deviation in the fault model compared to previous mapping efforts (DuRoss *et al.*, 2020; Hudnut *et al.*, 2020; Pierce *et al.*, 2020; Rodriguez Padilla, Quintana, *et al.*, 2022; Young *et al.*, 2024) occurs at the northwestern tip of the mainshock rupture near the Coso volcanic field. In our model, the fault is characterized by a shallow dip to the southeast, contrasting with the steeply dipping horsetail-like features mapped at the surface in previous studies. This discrepancy may stem from fitting a continuous surface through a diffuse cloud of seismicity that lacks indications of fault network structure. Although the horsetail-like lineations are difficult to distinguish visually at depth, focal mechanisms suggesting vertical strike-slip features are abundant (Plesch *et al.*, 2020; Wang and Zhan, 2020; Fialko, 2021). Further discussion of the distribution of seismicity and the modeled fault results is provided in the Interpretation and evaluation section.

Interpretation and evaluation

We present SURF, a method for semiautomated 3D mapping based on hypocentral locations. The two test cases described in the Results and Discussion sections demonstrate the potential for applying this method in regions with existing 3D fault models and in areas where such models are lacking. These results raise several fundamental questions: (1) What are the limitations of using this novel yet simplistic approach? (2) How will the method perform in regions without relocated catalogs? (3) When is it appropriate to fit a plane through a diffuse cloud



of seismicity? (4) What are our suggested best practices for implementation, and what do the model outputs mean? All of these questions are explored in the following subsections.

Limitations. The SURF algorithm proposed in this study takes a simplistic and flexible approach to modeling 3D faults. To minimize user-defined parameter choices, the code does not impose strict assumptions about the distances that nearest-neighbor clusters should or should not be merged. However, this could be problematic if clusters are farther apart than the user would like to allow merging to occur. In such scenarios, the clusters must be coplanar and have similar orientations. Although our tests have not encountered issues, we acknowledge that the current approach could produce unrealistic results in some scenarios. If needed, adding a distance cutoff parameter would be a straightforward modification.

The model outcomes are inherently dependent on the density and accuracy of the input earthquake catalog in the region of interest. Reliable results cannot be obtained in areas with sparse seismicity or where the events lack a discernible trend. Both case studies presented here have abundant relocated seismicity where near-linear trends are observable. We find that in some cases, modeled surfaces extrapolated to Earth's surface align with the mapped surface traces, but others do not (Figs. 2 and 4). For example, the modeled SAF surface near San Juan Bautista (SJB) is offset from the surface trace by up to 2 km, whereas farther south, in the creeping section, the discrepancy is reduced to a few hundred meters. Notably, these deviations persist across all tested earthquake catalogs, including double-difference, standard-network, and double-difference tomographic results (Zhang and Thurber, 2006). This suggests that variable offset of seismicity from the surface trace is a feature of the earthquake catalog rather than an error in the method. One possible explanation is an abrupt change in dip near Earth's surface (Fig. 2). Alternatively, the offsets may stem from errors in the absolute locations, an issue that is debated in the literature (e.g., Thurber *et al.*, 1997; McGuire and Ben-Zion, 2005; Zaliapin and Ben-Zion, 2011). The CFM features the same offset of the fault at depth as in our results, but the fault is constructed to connect to the surface trace. This requires that the CFM here feature a slightly shallower dip than our results, as well as a shallowing of the dip of the fault near the surface.

Figure 4. 3D fault model for the 2019 Ridgecrest earthquake sequences. The modeled fault surfaces from this study are shown in 3D renderings. The purple lines represent surface traces mapped by DuRoss *et al.* (2020). (a) Map view of the fault surfaces, colored by depth (yellow indicates shallow, and red indicates deep). The light blue eye icons indicate view directions for (panels b and c) and labeled accordingly. (b) Southern region of the sequence, showing the M_w 6.4 foreshock and near-orthogonal surfaces with subparallel strands extending to depth. (c) Northern region near the Coso volcanic field, where the fault model diverges from the existing model. No VE is applied to any 3D renderings. See Video S2 for a rotating animation of this model, which provides a complete 360° perspective on fault geometry and seismicity The color version of this figure is available only in the electronic edition.

Network versus relocated catalogs. The results from the test cases presented in this study use double-difference catalogs from highly instrumented regions of California. Although this represents an ideal scenario, it raises the question: how do the

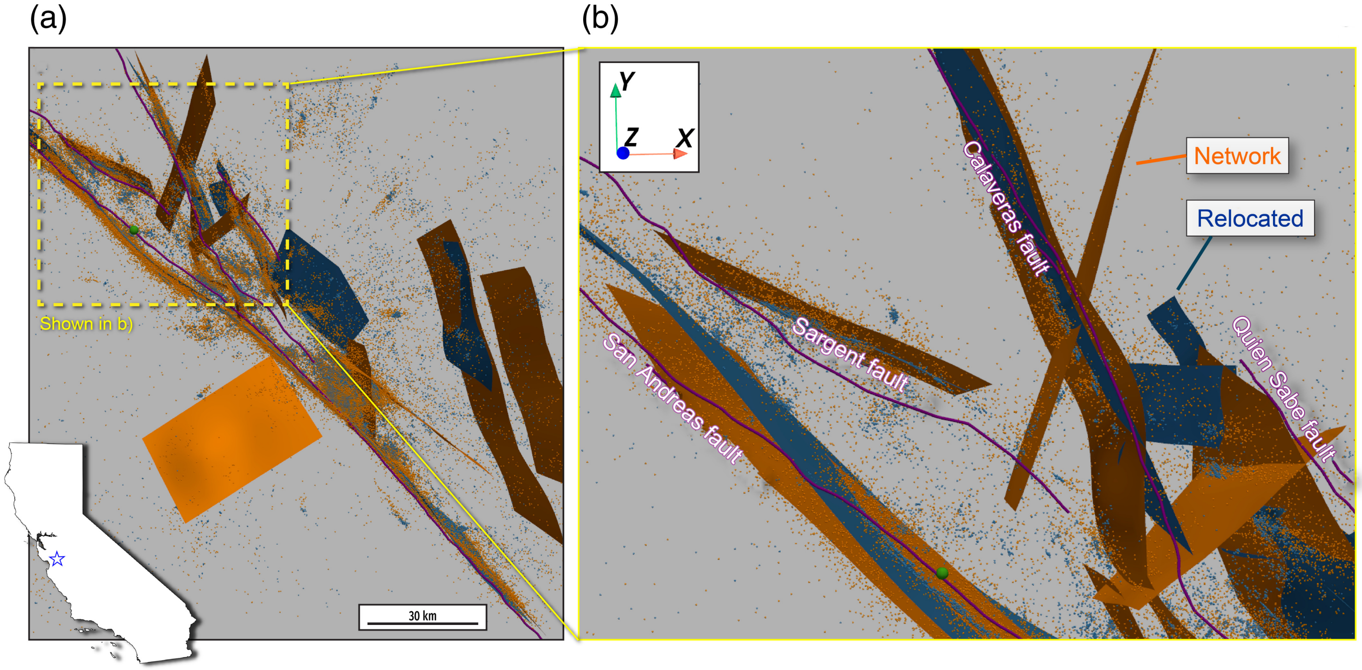


Figure 5. Fault model results for the San Juan Bautista region of central California. The orange surfaces and spheres are the model results using "network" catalog using Northern California Earthquake Data Center's (NCEDC's) Northern California Seismic System (1984–2021). The blue surfaces and spheres are the results using the "relocated" double-differenced catalog during the same timeframe (Waldhauser and Schaff, 2021). Note that these surfaces are the same as what is shown in Figure 2. (a) The region of interest where the relocated and network catalogs broadly agree along the San Andreas and Quien Sabe faults. (b) A zoom in of the junction region of the highlighting how Calaveras and Sargent faults (labeled) where the model results differ; the network results include rougher surfaces and some results that may be erroneous. See Video S3 for a rotating animation of this model, which provides a complete 360° perspective on fault geometry and seismicity. The color version of this figure is available only in the electronic edition.

results change when a routine network catalog is used that has not been relocated? To address this, we modeled the fault system near SJB (discussed in San Juan Bautista, California) using the standard Northern California Earthquake Data Center's Northern California Seismic System catalog, which covers the same coordinates and time frame (1984–2021) as the double-difference catalog results discussed in San Juan Bautista, California. These two different datasets (double difference and network) have relative errors that vary by roughly an order of magnitude, tens of meters for the double-difference versus hundreds of meters for the network catalog ($e_m \approx \mathrm{rmse}_t \cdot v_{\mathrm{avg}}$), in which rmse is root mean square error; see Figure S1 for the error distributions. For a visual comparison of the algorithm's results for the different datasets, refer to Figure 5 and Video S3. Qualitatively, the general model outputs for the double-difference catalog and the network catalog are somewhat consistent; however, there are second-order differences. We find that model results corresponding to the location of the SAF and Sargent fault are very similar for both network and relocated catalog results. However, results diverge near the SAF and CF junction, close to SJB and Hollister. Here, the network catalog yields a greater number and more complex fault shapes and morphologies that we attribute to the diffuse distribution of hypocenters. In addition, the network results for the CF are more nonplanar and have more intersecting faults than the double-difference results. These features of the CF have not been documented in other studies. In our interpretation, the double-difference catalog results better agree with existing models and are more physically realistic. When double-different catalogs are available, they will likely produce the best results, which is an intuitive result given the structural clarity that more precise hypocentral locations generally provide.

Quality assessment, diffuse seismicity, and interpretation. A key challenge in fault modeling is determining when a surface approximately represents the underlying seismicity. In both the SJB and Ridgecrest examples, we observe regions where seismicity is broadly distributed and lacks indicators of continuous throughgoing fault structures. These regions include the SAF–CF junction near SJB, where seismicity is diffused and the algorithm generates variably oriented surfaces (Fig. 2), and the northwestern tip of the Ridgecrest mainshock rupture, near Coso volcanic field, where seismicity is confined to a narrow depth range but lacks clear fault structure (Fig. 4). These examples highlight the fundamental questions of when is it justifiable to fit planes or surfaces through diffuse clouds of seismicity, and how should model results be interpreted?

One approach for assessing model quality is comparing the spatial distribution of seismicity around the modeled fault surfaces to the known tendency for the number of earthquakes to decrease as a function of distance from the fault in an inverse power law relationship (introduction). By fitting equation (4) to the observed hypocentral distribution around the modeled fault surfaces and evaluating the misfit using the MRE (equation 5), we identify regions where the distribution of seismicity aligns well with the modeled faults.

Faults such as the San Andreas, Quien Sabe, and portions of the CF exhibit low MRE misfit (<0.5), indicating that the seismicity distribution follows the expected power law decay. However, in regions with diffuse seismicity, the power law deteriorates, suggesting that a single fault surface may not fully capture the underlying fault structure (Figs. S2 and S3).

Interpreting fit parameters *d* and γ in power law distributions. Beyond overall misfit, the fit parameters d and γ in equation (4) offer additional insight into how well the modeled fault surfaces represent seismicity. The parameter d, which defines the characteristic corner in the power law distribution of earthquake density with distance, can be particularly informative when also considering hypocentral location uncertainty. In the SJB double-difference catalog, location uncertainties are lognormally distributed and range from ≈5 to 50 m (Fig. S1). In this study, where the fits for d significantly exceed the location uncertainty (Figs. S2 and S3), it may indicate real unmodeled fault complexity or orientation misfit, as shown in Figure 6b,c.

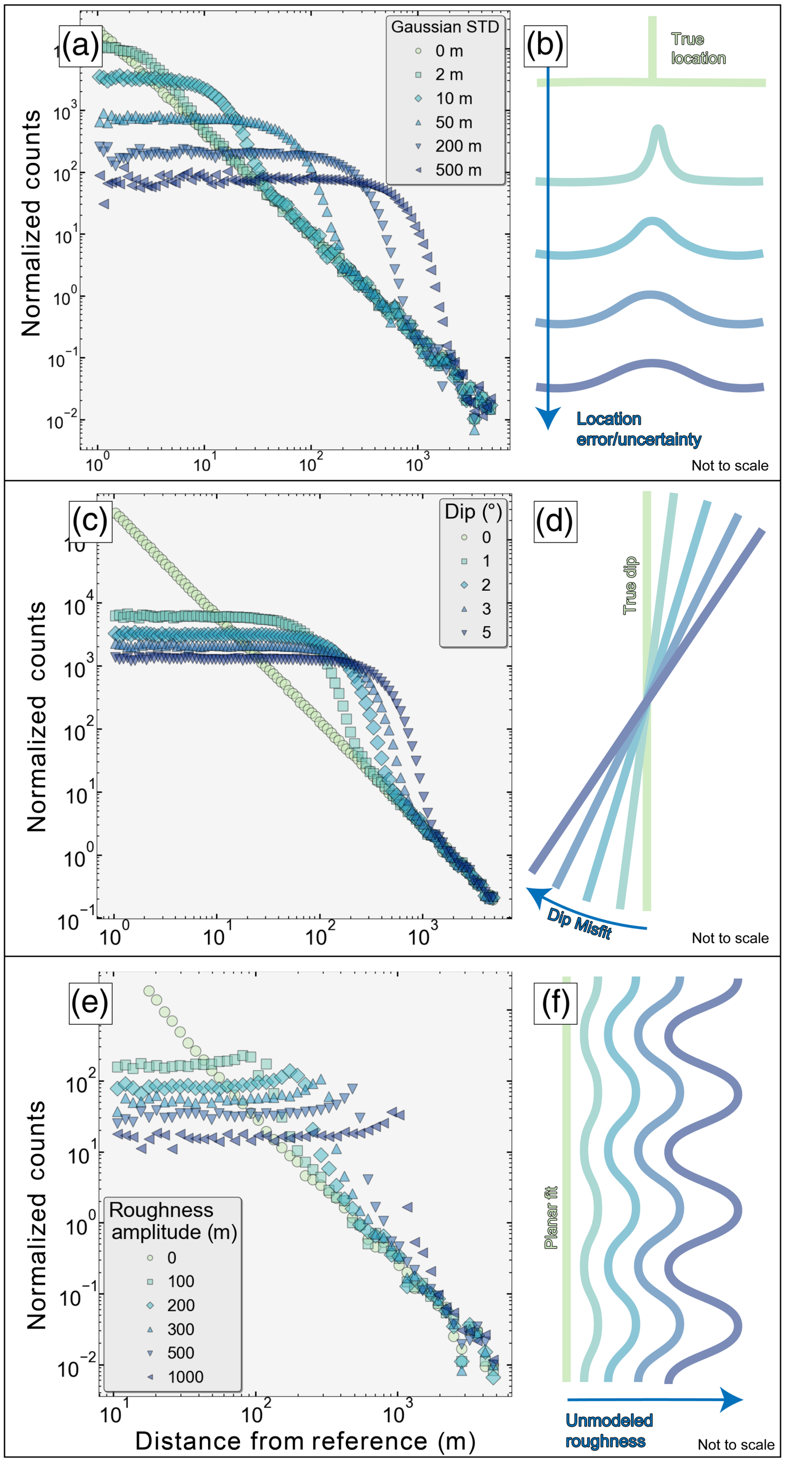


Figure 6. Impact of misfit and unmodeled complexity on the corner d and slope γ on the power law relationship equation (4). The points (green circles) are distributed in an idealized pattern around the fault, with lateral distances following a power-law distribution from a central line, and positions along the fault length and depth following a Gaussian distribution. (a) The figure shows how earthquake location uncertainty impacts the power law distribution. (b) The impact on the power law relationship when modifying the dip of the reference measured surface while holding the points fixed and measuring distances to the rotated reference surface. (c) The effect of unmodeled surface roughness modeled by superimposing a sine wave to the prescribed distribution of points and then measuring the distances with respect to the original planar fit. (d–f) Schematics illustrating the synthetic perturbations that are plotted to the left. Note that examples feature a similar result of moving the corner of the power law farther from the fault with increasing unaccounted for fault complexity or fit errors. The color version of this figure is available only in the electronic edition.

Previous studies have proposed a link between d and hypocentral uncertainty in power law observations (Goebel *et al.*, 2014; Rodriguez Padilla and Oskin, 2023). However, our analysis finds little to no correlation between the mean hypocentral location uncertainty and the parameters discussed in this subsection (Fig. S4), reinforcing the idea that d may capture structural complexity beyond location errors.

To further explore the interpretation of d, we conducted synthetic tests simulating fault-like seismicity distributions while systematically perturbing the fault geometry. In these simulations, we generate a simplified 1D linear and vertical strike-slip fault by distributing points (representing hypocenters) along its length with a Gaussian depth profile, a reasonable assumption for California (Fig. S5). We then prescribe an idealized power law decay of distances from the line, after which we introduced controlled geometric perturbations. To model fault misfit, we incrementally rotate the line. To model surface roughness, we superimpose a sine wave on the "fault" line, representing unresolved surface complexity (Fig. 6).

These results demonstrate that even slight deviations, such as a rotation of just a few degrees or adding a low-amplitude sine wave, are sufficient to introduce a measurable corner d where none initially existed. We interpret d as a characteristic length scale that reflects the distance between the modeled fault and the true underlying geometry. For example, the magnitude of d scales with the amplitude of the imposed sine wave, as shown in Figure 6, and we observe that $d \approx \sigma * \tan \delta$ in which σ is the standard deviation of the normal distribution along the synthetic fault length, and δ is the dip misfit. These synthetic examples suggest that d captures the effect of unmodeled complexity or geometric misfit and may be used as a diagnostic measure of the relative accuracy of the modeled fault (Fig. 7).

In some cases, we find anomalously low values of the power law exponent γ, with values <0.8 (Powers and Jordan, 2010; Rodriguez Padilla, Oskin, *et al.*, 2022; Rodriguez Padilla and Oskin, 2023) (Fig. 7). Such low γ values indicate a more gradual decrease in seismicity density with distance from the fault, implying a broader spatial distribution of earthquakes than typically observed for well-defined faults. We interpret these low γ values as evidence that the seismicity is diffusely distributed around the modeled fault rather than tightly localized along a discrete structure. This suggests that in some regions, a single fault surface may not fully capture the true underlying fault structure. Consequently, low γ values serve as an additional metric for evaluating whether a modeled fault surface is a meaningful representation of the observed seismicity.

Machine learning surface fit parameters. A complementary approach to interpreting the modeled faults is to examine the SVR fit parameters, which provide additional insight into the structural characteristics of the modeled surfaces, as discussed in the Surface fitting section. The SVR parameters C and ϵ describe the nonplanarity of the surface and the spatial distribution of seismicity, respectively.

We find that low C values, indicating nearly planar surfaces, sometimes coincide with regions where modeled surface is fit to diffuse seismicity clouds (Fig. 8). In such cases, the low C values suggest that incorporating additional curvature does not improve the SVR fit. This may imply that either the seismicity is uniformly distributed around a genuinely planar fault or that the spatial trends of seismicity do not provide meaningful structural constraints.

The distribution of C values appears bimodal, and we use Gaussian mixture modeling to determine a separation threshold as $C = 0.35$ between the two distributions (see Figure S6). In our test cases, model results from the low C population generally correspond to regions with diffuse seismicity. However, before using the C value as a quality indicator, the expected fault geometry should be considered. In this study, we have prior knowledge that the fault geometries are nonplanar (Plesch *et al.*, 2020; Aagaard and Hirakawa, 2021). In contrast, some studies, particularly in regions with induced earthquakes, feature nearly planar features, making the C value less informative as a quality indicator (Skoumal *et al.*, 2019; Cochran *et al.*, 2020; Park *et al.*, 2022).

In addition, we find that the SVR parameter ϵ is moderately correlated with d from equation (4) with a correlation coefficient of $R^2 = 0.61$ (Fig. S7). This is unsurprising because both metrics provide information about the distribution of near-fault seismicity. Although ϵ lacks a clear physical basis as a quality indicator, its correlation with d suggests that it offers complementary information.

Synthesis: A multimetric approach to fault model evaluation. No single metric fully captures the modeled fault surface uncertainty, but by integrating multiple measures, d, γ, C, ϵ, and MRE, model reliability may be cautiously interpreted. Whereas the goodness-of-fit to the inverse power law distribution (MRE) highlights faults that exhibit expected seismicity trends, deviations, particularly in γ, suggest cases in which a single surface may not fully represent the fault system. In cases in which hypocentral location uncertainty is low, the corner parameter d captures unmodeled fault complexity, with higher values indicating either systematic misfits or genuine small-scale structural variations. Similarly, the SVR parameter C helps distinguish between well-constrained and poorly constrained fault surfaces, in which low values often coincide with diffuse seismicity clouds, potentially indicating a lack of clear structural trends. By considering these parameters together, we can develop a more comprehensive perspective on when and where fault surfaces can be confidently interpreted and where model-based structural interpretations should be treated with caution. This framework provides a systematic way of interpreting fault models and highlights anomalous areas that may justify further investigation.

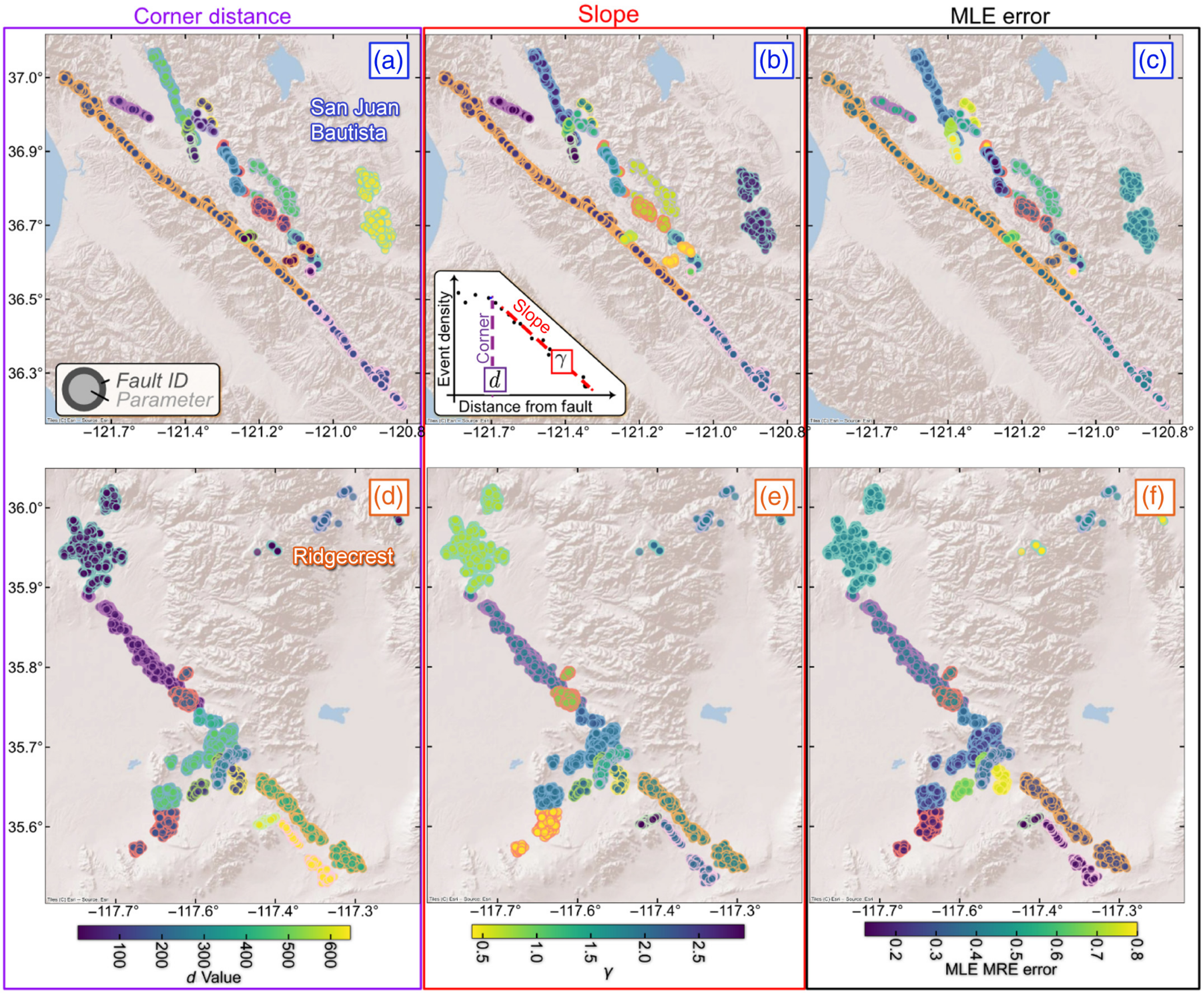


Figure 7. Map view display of the output of power law fit parameters *d* (a,d), γ (b,e), and the mean relative error (MRE) of the maximum-likelihood estimated (MLE) (c,f) fit to equation (4), for the two example datasets. The markers indicate the location of the earthquake cluster centroids used to generate the fault surfaces, the fill color indicates the parameter value, and the marker edge color indicates the merged clusters fault ID used to generate the modeled fault. The color map for the markers ranges from yellow to purple, where the yellower colors indicate potentially problematic fault surfaces. (a–c) Results for the San Juan Bautista region of California. (d–f) The parameter values for the 2019 Ridgecrest, California, earthquake sequence. The color version of this figure is available only in the electronic edition.

Recommendations for implementation

It is important to note that several key parameters influence the outcome of the algorithm. For example, we found that setting the minimum cluster size to 30–40 events allows for the generation of longer faults. Using large minimum cluster sizes either omits small dense clusters or lumps them together and reduces the resolution. Conversely, using minimum cluster sizes that are too small often produces unstable results because cluster orientation calculations become uncertain with few data points. In some rare cases, clustering may produce two clusters, with one containing a large portion of events and another cluster with few events. This issue can be resolved by reclustering the large cluster and breaking it down into smaller clusters.

The merging thresholds also impact the results. In these examples, we use a threshold value for **M** 0.75 from equation (3), which produces results that agree with previous models (Figs. 3 and 4). Increasing the merging threshold to higher values reduces the likelihood of merging. Another important parameter is the number of nearest neighbors used for comparing directions and orientations. As may be expected, increasing the number of neighbors may result in merging over large distances, which typically increases fault continuity.

We also define ranges of the two parameters C and ϵ that control the SVR fit surface. Optimal values for these parameters are selected using cross-validation with a randomized search over

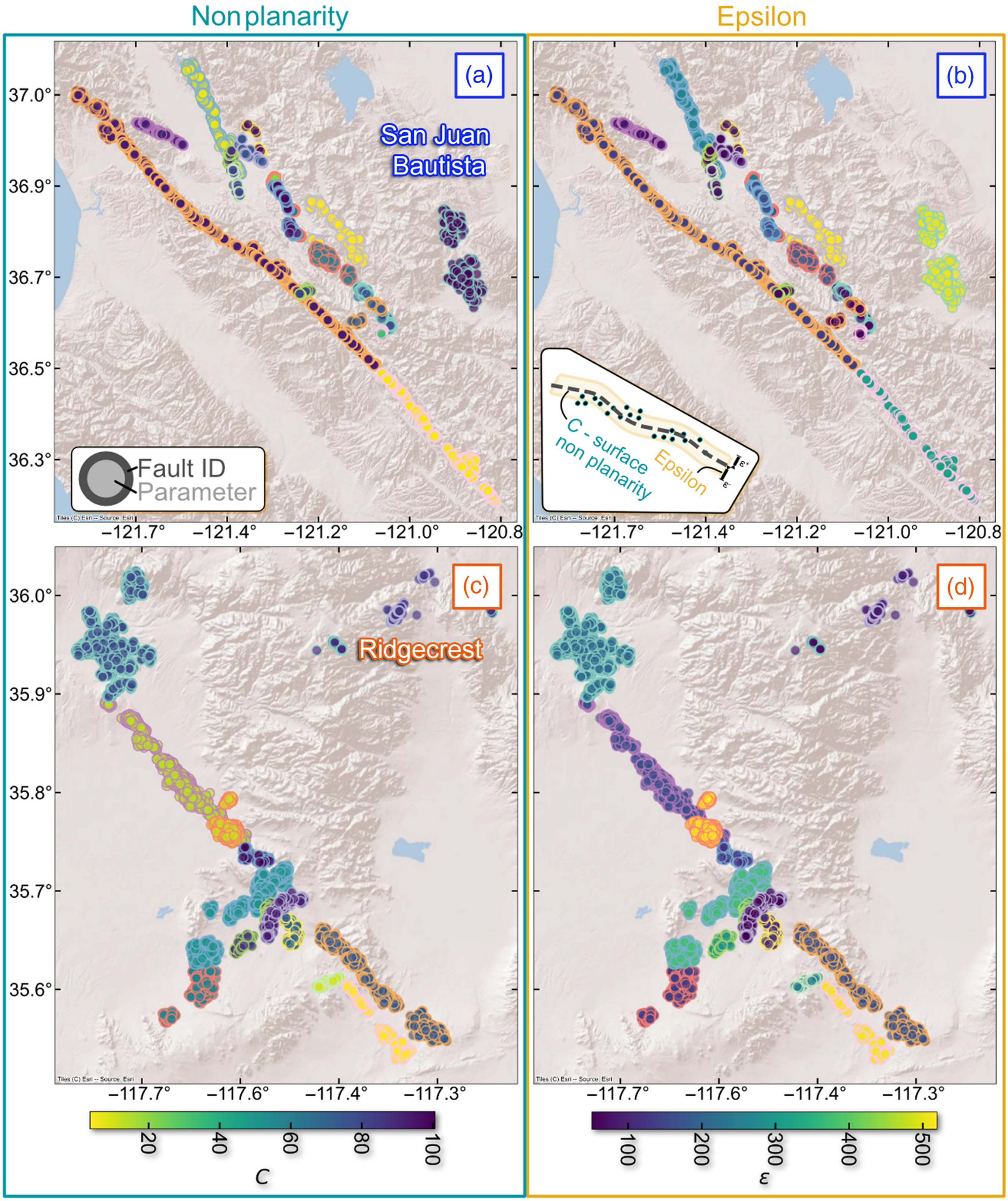


Figure 8. Map view display of the parameter outputs: C, that describes the surface nonplanarity (a,c) and ϵ (b,d) for the two example datasets. The markers indicate the location of the earthquake cluster centroids used to generate the fault surfaces, the fill color indicates the parameter value, and the marker edge color indicates the merged cluster's fault ID that was used to generate the modeled fault using the method presented in the text. (a,b) Results for the San Juan Bautista region of California. (c,d) The parameter values for the 2019 Ridgecrest, California, earthquake sequence. The color version of this figure is available only in the electronic edition.

a uniform distribution for both parameters. The tolerance zone ϵ around the fitted surface, where errors are not penalized, accommodates location errors and microseismicity that occurs within the fault damage zone. Because the length scale of absolute hypocentral location uncertainties (Fig. S1) and damage zone dimensions are generally on the order of hundreds of meters, we set ϵ to range from 1 to 500 m; this may be adjusted for other use cases. For example, larger permissible ϵ may be required in regions with sparse or asymmetric seismic station coverage that yield different uncertainty ellipsoids. Conversely, smaller values might be appropriate for very precise near source studies. The parameter C controls the model's complexity and tolerance for errors when larger values correspond to a greater degree of nonplanarity (see Fig. S8). In our tests, we find a range of 1–100 for C balances model complexity and smoothness, with 1 yielding near-planar outputs and 100 allowing for reasonable surface detail. Larger C values can be applied as needed depending on expected fault geometry and data intricacies.

The method presented in this study offers a semiautomated and uniform approach to 3D fault modeling. This uniform approach offers scalability and consistency across different regions. Our results may provide complementary, alternative, or additional information to expertly interpreted fault model studies. We suggest that the method presented here may be a good starting point for 3D fault mapping and that a hybrid approach that includes some additional quality controls and user interpretation is likely best. In addition, the method can be modified based on the desired use cases and information available, such as using surface traces information to shift fault surfaces in space or incorporating focal mechanism information into the clustering step.

Conclusion

The SURF algorithm presented in this study provides a novel, low-user-input approach for constructing 3D fault geometries from seismicity patterns. By integrating density-based clustering with spatially constrained cluster merging, this workflow effectively models continuous fault surfaces, crossing faults, and subparallel fault structures. A key advantage of this approach is that it fits fault surfaces to the data without imposing a priori assumptions about their shape. To objectively assess model reliability, we evaluate the spatial distribution of seismicity around the modeled surfaces, ensuring alignment with established fault zone characteristics.

When applied to the SAF–CF junction and the 2019 Ridgecrest earthquake sequence, the SURF algorithm produces fault geometries that closely match previously published models. This agreement highlights its effectiveness in delineating complex fault structures using only earthquake catalogs. The simplicity and scalability of this workflow make it a promising tool for further fault modeling applications, offering an efficient, data-driven alternative for understanding subsurface fault structures and improving seismic hazard assessments.

Although the method performs well in the case studies presented, its applicability to regions with sparser seismicity or more diffuse fault structures warrants further investigation. Future improvements could incorporate additional geophysical constraints, such as geodetic data or geological mapping, to enhance model accuracy. In addition, refining clustering and merging criteria may improve fault delineation in complex settings, particularly where fault intersections and diffuse seismicity clouds are present. Beyond fault modeling, this workflow could aid in real-time seismic monitoring efforts by providing rapid estimates of fault geometry following significant earthquake sequences.

Data and Resources

The SURF software Alongi and Skoumal (2025) is publically available. Python (https://www.python.org), the NumPy package (https://numpy.org), the Pandas package (https://pandas.pydata.org), the SciPy package (https://scipy.org), the PyVista package (https://docs.pyvista.org), the pyproj package (https://pyproj4.github.io/pyproj), the scikit-learn package (https://scikit-learn.org), and the Networkx package (https://networkx.org) were used. These websites were last accessed in March 2025. The northern California double-difference earthquake catalogs for the Bay Area were obtained from Lamont–Doherty Earth Observatory, Columbia University California Data Center, https://nocaldd.ldeo.columbia.edu/catalog (last accessed December 2024). Northern California Earthquake Data Center, NCSS catalog for this study were accessed through the Northern California Earthquake Data Center (NCEDC), doi: 10.7932/NCEDC (last accessed December 2024). Ridgecrest catalog was obtained from the U.S. Geological Survey (USGS) data release that may be accessed at doi: 10.5066/P9JN6H0N. Mapped fault traces used in figures were obtained from the Quaternary Fault and Fold Database (https://www.usgs.gov/programs/earthquake-hazards/faults). See the supplemental material for additional figures and animations that support main article.

Declaration of Competing Interests

The authors acknowledge that there are no conflicts of interest recorded.

Acknowledgments

Stations used in this study are operated by the U.S. Geological Survey; the University of California, Berkeley; and the California Division of Mines and Geology. Ole Kaven, Ruijia Wang, and an anonymous reviewer provided helpful reviews and comments that helped improve the article. The authors give thanks to Tim Clements for testing and helping to improve the seismicity utilized for reconstructing faults (SURF) algorithm.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. government.

References

Aagaard, B., and E. Hirakawa (2021). San Francisco bay region 3D seismic velocity model v21. 1, *U.S. Geol. Surv. Data Release* [Data Set], doi: 10.5066/P9TRDCHE.

Alaei, B., and A. Torabi (2017). Seismic imaging of fault damaged zone and its scaling relation with displacement, *Interpretation* **5,** no. 4, SP83–SP93, doi: 10.1190/INT-2016-0230.1.

Alongi, T., E. E. Brodsky, J. Kluesner, and D. Brothers (2022). Using active source seismology to image the Palos Verdes fault damage zone as a function of distance, depth, and geology, *Earth Planet. Sci. Lett.* **600,** 117,871, doi: 10.1016/j.epsl.2022.117871.

Alongi, T., E. E. Brodsky, J. Kluesner, and D. Brothers (2024). Characteristics of the fault damage zone from high-resolution seismic imaging along the Palos Verdes fault, California, *AGU Adv.* **5,** no. 4, e2023AV001155, doi: 10.1029/2023AV001155.

Alongi, T., and R. Skoumal (2025). SURF: A python package for reconstructing faults from seismicity patterns (Version 1.0.0), *U.S. Geol. Surv. Soft. Rel.* doi: 10.5066/P1MM9HMW.

Antoine, S. L., Y. Klinger, A. Delorme, K. Wang, R. Bürgmann, and R. D. Gold (2021). Diffuse deformation and surface faulting distribution from submetric image correlation along the 2019 Ridgecrest, California, ruptures, *Bull. Seismol. Soc. Am.* **111,** no. 5, 2275–2302.

Atterholt, J., Z. Zhan, and Y. Yang (2022). Fault zone imaging with distributed acoustic sensing: Body-to-surface wave scattering, *J. Geophys. Res.* **127,** no. 11, e2022JB025052, doi: 10.1029/2022JB025052.

Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching, *Commun. ACM* **18,** no. 9, 509–517.

Berg, S. S., and T. Skar (2005). Controls on damage zone asymmetry of a normal fault zone: Outcrop analyses of a segment of the Moab fault, SE Utah, *J. Struct. Geol.* **27,** no. 10, 1803–1822, doi: 10.1016/j.jsg.2005.04.012.

Bishop, C. M., and N. M. Nasrabadi (2006). *Pattern Recognition and Machine Learning*, Vol. 4, Springer, New York, New York.

Boettcher, M. S., and T. Jordan (2004). Earthquake scaling relations for mid-ocean ridge transform faults, *J. Geophys. Res.* **109,** no. B12, doi: 10.1029/2004JB003110.

Caine, J. S., J. P. Evans, and C. B. Forster (1996). Fault zone architecture and permeability structure, *Geology* **24,** no. 11, 1025, doi: 10.1130/0091-7613(1996)024<1025:FZAAPS>2.3.CO;2.

Campello, R. J., D. Moulavi, and J. Sander (2013). Density-based clustering based on hierarchical density estimates, *Pacific-Asia Conf. on Knowledge Discovery and Data Mining*, 160–172.

Candela, T., and E. E. Brodsky (2016). The minimum scale of grooving on faults, *Geology* **44,** no. 8, 603–606.

Candela, T., F. Renard, Y. Klinger, K. Mair, J. Schmittbuhl, and E. E. Brodsky (2012). Roughness of fault surfaces over nine decades of length scales, *J. Geophys. Res.* 117, doi: 10.1029/2011JB009041.

Carena, S., J. Suppe, and H. Kao (2002). Active detachment of Taiwan illuminated by small earthquakes and its control of first-order topography, *Geology* **30,** no. 10, 935–938.

Cattania, C., and P. Segall (2021). Precursory slow slip and foreshocks on rough faults, *J. Geophys. Res.* **126,** no. 4, e2020JB020430, doi: 10.1029/2020JB020430.

Chester, J. S., F. M. Chester, and A. K. Kronenberg (2005). Fracture surface energy of the Punchbowl fault, San Andreas system, *Nature* **437,** no. 7055, 133–136.

Choi, J.-H., P. Edwards, K. Ko, and Y.-S. Kim (2016). Definition and classification of fault damage zones: A review and a new methodological approach, *Earth Sci. Rev.* **152,** 70–87, doi: 10.1016/j.earscirev.2015.11.006.

Chu, S. X., V. C. Tsai, D. T. Trugman, and G. Hirth (2021). Fault interactions enhance high-frequency earthquake radiation, *Geophys. Res. Lett.* **48,** no. 20, e2021GL095271, doi: 10.1029/2021GL095271.

Cocco, M., S. Aretusini, C. Cornelio, S. B. Nielsen, E. Spagnuolo, E. Tinti, and G. Di Toro (2023). Fracture energy and breakdown work during earthquakes, *Annu. Rev. Earth Planet. Sci.* **51,** no. 1, 217–252.

Cochran, E. S., Y.-G. Li, P. M. Shearer, S. Barbot, Y. Fialko, and J. E. Vidale (2009). Seismic and geodetic evidence for extensive, long-lived fault damage zones, *Geology* **37,** no. 4, 315–318, doi: 10.1130/G25306A.1.

Cochran, E. S., R. J. Skoumal, D. McPhillips, Z. E. Ross, and K. M. Keranen (2020). Activation of optimally and unfavourably oriented faults in a uniform local stress field during the 2011 Prague, Oklahoma, sequence, *Geophys. J. Int.* **222,** no. 1, 153–168.

DuRoss, C. B., R. D. Gold, T. E. Dawson, K. M. Scharer, K. J. Kendrick, S. O. Akciz, S. J. Angster, J. Bachhuber, S. Bacon, S. E. Bennett, *et al.* (2020). Surface displacement distributions for the July 2019 Ridgecrest, California, earthquake ruptures, *Bull. Seismol. Soc. Am.* **110,** no. 4, 1400–1418.

Faulkner, D., C. Jackson, R. Lunn, R. Schlische, Z. Shipton, C. Wibberley, and M. Withjack (2010). A review of recent developments concerning the structure, mechanics and fluid flow properties of fault zones, *J. Struct. Geol.* **32,** no. 11, 1557–1575, doi: 10.1016/j.jsg.2010.06.009.

Fialko, Y. (2004). Evidence of fluid-filled upper crust from observations of postseismic deformation due to the 1992 M_w 7.3 Landers earthquake, *J. Geophys. Res.* **109,** doi: 10.1029/2004JB002985.

Fialko, Y. (2021). Estimation of absolute stress in the hypocentral region of the 2019 Ridgecrest, California, earthquakes, *J. Geophys. Res.* **126,** no. 7, e2021JB022000, doi: 10.1029/2021JB022000.

Fialko, Y., and Z. Jin (2021). Simple shear origin of the cross-faults ruptured in the 2019 Ridgecrest earthquake sequence, *Nature Geosci.* **14,** no. 7, 513–518.

Field, E. H., R. J. Arrowsmith, G. P. Biasi, P. Bird, T. E. Dawson, K. R. Felzer, D. D. Jackson, K. M. Johnson, T. H. Jordan, C. Madden, *et al.* (2014). Uniform California earthquake rupture forecast, version 3 (UCERF 3)—The time-independent model, *Bull. Seismol. Soc. Am.* **104,** no. 3, 1122–1180.

Fischler, M. A., and R. C. Bolles (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM* **24,** no. 6, 381–395.

Fletcher, J. M., M. E. Oskin, and O. J. Teran (2016). The role of a keystone fault in triggering the complex El Mayor-Cucapah earthquake rupture, *Nature Geosci.* **9,** no. 4, 303–307.

Goebel, T., T. Candela, C. Sammis, T. Becker, G. Dresen, and D. Schorlemmer (2014). Seismic event distributions and off-fault damage during frictional sliding of saw-cut surfaces with pre-defined roughness, *Geophys. J. Int.* **196,** no. 1, 612–625.

Hatem, A. E., M. L. Cooke, and E. H. Madden (2015). Evolving efficiency of restraining bends within wet kaolin analog experiments, *J. Geophys. Res.* **120,** no. 3, 1975–1992.

Hudnut, K. W., B. A. Brooks, K. Scharer, J. L. Hernandez, T. E. Dawson, M. E. Oskin, J. Ramon Arrowsmith, C. A. Goulet, K. Blake, M. L. Boggs, *et al.* (2020). Airborne lidar and electro-optical imagery along surface ruptures of the 2019 Ridgecrest earthquake sequence, southern California, *Seismol. Res. Lett.* **91,** no. 4, 2096–2107, doi: 10.1785/0220190338.

Jourdon, A., D. A. May, J. N. Hayek, and A.-A. Gabriel (2025). 3D reconstruction of complex fault systems from volumetric geodynamic shear zones using medial axis transform, *Geochem. Geophys. Geosys.* **26,** no. 6, e2025GC012169, doi: 10.1029/2025GC012169.

Kaven, J. O., and D. D. Pollard (2013). Geometry of crustal faults: Identification from seismicity and implications for slip and stress transfer models, *J. Geophys. Res.* **118,** no. 9, 5058–5070.

Li, Y.-G., J. E. Vidale, and E. S. Cochran (2004). Low-velocity damaged structure of the San Andreas fault at Parkfield from fault zone trapped waves: Low-velocity damaged structure of the San Andreas fault, *Geophys. Res. Lett.* **31,** no. 12, doi: 10.1029/2003GL019044.

Lienkaemper, J. J., F. S. McFarland, R. W. Simpson, and S. J. Caskey (2014). Using surface creep rate to infer fraction locked for sections of the San Andreas fault system in northern California from alignment array and GPS data, *Bull. Seismol. Soc. Am.* **104,** no. 6, 3094–3114.

Liu, C., T. Lay, E. E. Brodsky, K. Dascher-Cousineau, and X. Xiong (2019). Coseismic rupture process of the large 2019 Ridgecrest earthquakes from joint inversion of geodetic and seismological observations, *Geophys. Res. Lett.* **46,** no. 21, 11,820–11,829.

Lomax, A. (2020). Absolute location of 2019 Ridgecrest seismicity reveals a shallow M_w 7.1 hypocenter, migrating and pulsing M_w 7.1 foreshocks, and duplex M_w 6.4 ruptures, *Bull. Seismol. Soc. Am.* **110,** no. 4, 1845–1858.

Lozos, J. C., D. D. Oglesby, B. Duan, and S. G. Wesnousky (2011). The effects of double fault bends on rupture propagation: A geometrical parameter study, *Bull. Seismol. Soc. Am.* **101,** no. 1, 385–398.

McGuire, J., and Y. Ben-Zion (2005). High-resolution imaging of the bear valley section of the San Andreas fault at seismogenic depths with fault-zone head waves and relocated seismicity, *Geophys. J. Int.* **163,** no. 1, 152–164.

Nevitt, J. M., B. A. Brooks, J. L. Hardebeck, and B. T. Aagaard (2023). 2019 M 7.1 Ridgecrest earthquake slip distribution controlled by fault geometry inherited from independence dike swarm, *Nat. Commun.* **14,** no. 1, 1546.

Ouillon, G., C. Ducorbier, and D. Sornette (2008). Automatic reconstruction of fault networks from seismicity catalogs: Three-dimensional optimal anisotropic dynamic clustering, *J. Geophys. Res.* **113,** doi: 10.1029/2007JB005032.

Park, Y., G. C. Beroza, and W. L. Ellsworth (2022). Basement fault activation before larger earthquakes in Oklahoma and Kansas, *Seism. Rec.* **2,** no. 3, 197–206.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.* (2011). Scikit-learn: Machine learning in python, *J. Machine Learn. Res.* **12,** 2825–2830.

Perrin, C., F. Waldhauser, and C. H. Scholz (2021). The shear deformation zone and the smoothing of faults with displacement, *J. Geophys. Res.* doi: 10.1029/2020JB020447.

Petersen, M. D., A. M. Shumway, P. M. Powers, E. H. Field, M. P. Moschetti, K. S. Jaiswal, K. R. Milner, S. Rezaeian, A. D. Frankel, A. L. Llenos, *et al.* (2024). The 2023 US 50-State National Seismic Hazard Model: Overview and implications, *Earthq. Spectra* **40,** no. 1, 5–88.

Pierce, I., A. Williams, R. D. Koehler, and C. Chupik (2020). High-resolution structure-from-motion models and orthophotos of the southern sections of the 2019 M_w 7.1 and 6.4 Ridgecrest earthquakes surface ruptures, *Seismol. Res. Lett.* **91,** no. 4, 2124–2126.

Plesch, A., J. H. Shaw, C. Benson, W. A. Bryant, S. Carena, M. Cooke, J. Dolan, G. Fuis, E. Gath, L. Grant, *et al.* (2007). Community fault model (CFM) for southern California, *Bull. Seismol. Soc. Am.* **97,** no. 6, 1793–1802, doi: 10.1785/0120050211.

Plesch, A., J. H. Shaw, Z. E. Ross, and E. Hauksson (2020). Detailed 3D fault representations for the 2019 Ridgecrest, California, earthquake sequence, *Bull. Seismol. Soc. Am.* **110,** no. 4, 1818–1831.

Poliakov, A. N., R. Dmowska, and J. R. Rice (2002). Dynamic shear rupture interactions with fault bends and off-axis secondary faulting, *J. Geophys. Res.* **107,** no. B11, ESE-6.

Pollitz, F. F., J. R. Murray, J. L. Svarc, C. Wicks, E. Roeloffs, S. E. Minson, K. Scharer, K. Kendrick, K. W. Hudnut, J. Nevitt, *et al.* (2020). Kinematics of fault slip associated with the 4-6 July 2019 Ridgecrest, California, earthquake sequence, *Bull. Seismol. Soc. Am.* **110,** no. 4, 1688–1700.

Powers, P. M., and T. H. Jordan (2010). Distribution of seismicity across strike-slip faults in California, *J. Geophys. Res.* **115,** doi: 10.1029/2008JB006234.

Rodriguez Padilla, A. M., and M. E. Oskin (2023). Displacement hazard from distributed ruptures in strike-slip earthquakes, *Bull. Seismol. Soc. Am.* **113,** no. 6, 2730–2745.

Rodriguez Padilla, A. M., M. E. Oskin, C. W. Milliner, and A. Plesch (2022). Accrual of widespread rock damage from the 2019 Ridgecrest earthquakes, *Nature Geosci.* **15,** no. 3, 222–226.

Rodriguez Padilla, A. M., M. A. Quintana, R. M. Prado, B. J. Aguilar, T. A. Shea, M. E. Oskin, and L. Garcia (2022). Near-field high-resolution maps of the Ridgecrest earthquakes from aerial imagery, *Bull. Seismol. Soc. Am.* **93,** no. 1, 494–499.

Ross, Z. E., B. Idini, Z. Jia, O. L. Stephenson, M. Zhong, X. Wang, Z. Zhan, M. Simons, E. J. Fielding, S.-H. Yun, *et al.* (2019). Hierarchical interlocked orthogonal faulting in the 2019 Ridgecrest earthquake sequence, *Science* **366,** no. 6463, 346–351.

Savage, H. M., and E. E. Brodsky (2011). Collateral damage: Evolution with displacement of fracture distribution and secondary fault strands in fault damage zones, *J. Geophys. Res.* **116,** doi: 10.1029/2010JB007665.

Sawaki, Y., T. Shiina, K. Sagae, Y. Sato, H. Horikawa, A. Miyakawa, K. Imanishi, and T. Uchide (2025). Fault geometries of the 2024 M_w 7.5 Noto peninsula earthquake from hypocenter-based hierarchical clustering of point-cloud normal vectors, *J. Geophys. Res.* **130,** no. 4, e2024JB030233, doi: 10.1029/2024JB030233.

Scott, C. P., J. R. Arrowsmith, E. Nissen, L. Lajoie, T. Maruyama, and T. Chiba (2018). The *M* 7 2016 Kumamoto, Japan, earthquake: 3-D deformation along the fault and within the damage zone constrained from differential lidar topography, *J. Geophys. Res.* **123,** no. 7, 6138–6155, doi: 10.1029/2018JB015581.

Segall, P., and D. Pollard (1980). Mechanics of discontinuous faults, *J. Geophys. Res.* **85,** no. B8, 4337–4350.

Shelly, D. R. (2020a). A high-resolution seismic catalog for the initial 2019 Ridgecrest earthquake sequence, doi: 10.5066/P9JN6H0N.

Shelly, D. R. (2020b). A high-resolution seismic catalog for the initial 2019 Ridgecrest earthquake sequence: Foreshocks, aftershocks, and faulting complexity, *Seismol. Res. Lett.* **91,** no. 4, 1971–1978.

Shipton, Z. K., and P. A. Cowie (2001). Damage zone and slip-surface evolution over mm to km scales in high-porosity Navajo sandstone, Utah, *J. Struct. Geol.* 20, doi: 10.1016/S0191-8141(01)00035-9.

Skoumal, R. J., J. O. Kaven, and J. I. Walter (2019). Characterizing seismogenic fault structures in Oklahoma using a relocated template-matched catalog, *Seismol. Res. Lett.* **90,** no. 4, 1535–1543.

Smola, A. J., and B. Schölkopf (2004). A tutorial on support vector regression, *Stat. Comput.* **14,** 199–222.

Thurber, C., S. Roecker, W. Ellsworth, Y. Chen, W. Lutter, and R. Sessions (1997). Two-dimensional seismic image of the San Andreas fault in the northern Gabilan range, central California: Evidence for fluids in the fault zone, *Geophys. Res. Lett.* **24,** no. 13, 1591–1594.

Thurber, C., S. Roecker, K. Roberts, M. Gold, L. Powell, and K. Rittger (2003). Earthquake locations and three-dimensional fault zone structure along the creeping section of the San Andreas fault near Parkfield, CA: Preparing for SAFOD, *Geophys. Res. Lett.* **30,** no. 3, doi: 10.1029/2002GL016004.

Truttmann, S., T. Diehl, and M. Herwegh (2023). Hypocenter-based 3D imaging of active faults: Method and applications in the southwestern Swiss Alps, *J. Geophys. Res.* **128,** no. 6, e2023JB026352, doi: 10.1029/2023JB026352.

Ulrich, T., A.-A. Gabriel, J.-P. Ampuero, and W. Xu (2019). Dynamic viability of the 2016 M_w 7.8 Kaikōura earthquake cascade on weak crustal faults, *Nat. Commun.* **10,** no. 1, 1–16.

Waldhauser, F., and D. P. Schaff (2021). A comprehensive search for repeating earthquakes in northern California: Implications for fault creep, slip rates, slip partitioning, and transient stress, *J. Geophys. Res.* **126,** no. 11, e2021JB022495, doi: 10.1029/2021JB022495.

Wang, X., and Z. Zhan (2020). Seismotectonics and fault geometries of the 2019 Ridgecrest sequence: Insight from aftershock moment tensor catalog using 3-D Green's functions, *J. Geophys. Res.* **125,** no. 5, e2020JB019577, doi: 10.1029/2020JB019577.

Wang, Y., G. Ouillon, J. Woessner, D. Sornette, and S. Husen (2013). Automatic reconstruction of fault networks from seismicity catalogs including location uncertainty, *J. Geophys. Res.* **118,** no. 11, 5956–5975.

Wesnousky, S. G. (2006). Predicting the endpoints of earthquake ruptures, *Nature* **444,** no. 7117, 358–360, doi: 10.1038/nature05275.

Willemse, E. J., and D. D. Pollard (2000). Normal fault growth: Evolution of tipline shapes and slip distribution, in *Aspects of Tectonic Faulting: In Honour of Georg Mandl*, F. K. Lehner and J. L. Urai (Editors), Springer, Berlin, Heidelberg, 193–226.

Young, E. K., M. E. Oskin, and A. M. Rodriguez Padilla (2024). Reproducibility of remote mapping of the 2019 Ridgecrest earthquake surface ruptures, *Seismol. Res. Lett.* **95,** no. 1, 288–298.

Zaliapin, I., and Y. Ben-Zion (2011). Asymmetric distribution of aftershocks on large faults in California, *Geophys. J. Int.* **185,** no. 3, 1288–1304.

Zhang, H., and C. Thurber (2006). Development and applications of double-difference seismic tomography, *Pure Appl. Geophys.* **163,** 373–403.

Manuscript received 22 April 2025