

Automatic reconstruction of fault networks from seismicity catalogs including location uncertainty

Y. Wang,¹ G. Ouillon,² J. Woessner,¹ D. Sornette,³ and S. Husen¹

Received 5 March 2013; revised 9 October 2013; accepted 30 October 2013; published 22 November 2013.

[1] We introduce the anisotropic clustering of location uncertainty distributions (ACLD) method to reconstruct active fault networks on the basis of both earthquake locations and their estimated individual uncertainties. After a massive search through the large solution space of possible reconstructed fault networks, we apply six different validation procedures in order to select the corresponding best fault network. Two of the validation steps (cross validation and Bayesian information criterion (BIC) process the fit residuals, while the four others look for solutions that provide the best agreement with independently observed focal mechanisms. Tests on synthetic catalogs allow us to qualify the performance of the fitting method and of the various validation procedures. The ACLD method is able to provide solutions that are close to the expected ones, especially for the BIC and focal mechanism-based techniques. The clustering method complemented by the validation step based on focal mechanisms provides good solutions even in the presence of a significant spatial background seismicity rate. Our new fault reconstruction method is then applied to the Landers area in Southern California and compared with previous clustering methods. The results stress the importance of taking into account undersampled subfault structures as well as of the spatially inhomogeneous location uncertainties.

Citation: 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. Solid Earth*, 118, 5956–5975, doi:10.1002/2013JB010164.

1. Introduction

[2] Earthquake forecasts should ultimately be founded on the premise that seismicity and faulting are intimately interwoven: Earthquakes occur on faults, and faults grow and organize in complex networks through accumulation of earthquakes. The obvious character and the power of this well-established fact are obfuscated by serious difficulties in exploiting it for a better science of earthquakes and their prediction. Indeed, an intrinsic limitation of present efforts to forecast earthquakes lies in the fact that only a limited part of the full fault network has been revealed, notwithstanding the best efforts combining geological, geodetic, and geophysical methods (see *Mace and Keranen* [2012], for instance) together with past seismicity to illuminate fault structures [*Hiemer et al.*, 2013; *Plesch et al.*, 2007]. Nevertheless, these studies suggest that fault networks display multiscaling hierarchical properties [*Cowie et al.*,

1995], which are intimately associated with the modes of tensorial deformations accommodating large-scale tectonic driving forces [*Sornette*, 1991; *Sornette and Virieux*, 1992]. Neglecting the information from fault networks constitutes a major gap in the understanding of the spatial-temporal organization of earthquakes (see however early attempts by *Cowie et al.* [1995], *Cowie et al.* [1993], and *Sornette et al.* [1994]), thus limiting the quality and efficiency of most current earthquake forecasting methods. Including more realistic geometries and tensorial strain information associated with the underlying reconstructed fault networks will in the long term improve present attempts to develop better space-time models of earthquake triggering, which still lack information on fault localization by assuming diffuse seismicity unrelated to faults or assume very simplified structures [*Gerstenberger et al.*, 2005; *Ogata and Zhuang*, 2006; *Woessner et al.*, 2010]. A reliable association of earthquakes and faults is an important constraint to determine the spatial decay of earthquakes in aftershock sequences, which provides insights into the triggering mechanisms of earthquakes [Stein, 1999] and improves estimates of where aftershock hypocenters are located in comparison to the main shock properties [*Hauksson*, 2010; *Powers and Jordan*, 2010; *Woessner et al.*, 2006].

[3] Earthquake forecasting must issue statements about the likely spatial location of upcoming events. In an ideal case, we would like to forecast the set of faults or fault segments about to break in the near future. This would help predicting the expected ground motions due to radiated seismic waves, as well as anticipating problems due to surface faulting prone

Additional supporting information may be found in the online version of this article.

¹Swiss Seismological Service, ETH Zurich, Zurich, Switzerland.

²Lithophyse, Nice, France.

³Department of Management, Technology and Economics, ETH Zurich, Zurich, Switzerland.

Corresponding author: Y. Wang, Swiss Seismological Service, ETH Zurich, NO H 39.3, Sonneggstrasse 5, Zurich CH-8092, Switzerland. (yaming.wang@sed.ethz.ch)

©2013. American Geophysical Union. All Rights Reserved.
2169-9313/13/10.1002/2013JB010164

to cause damage on infrastructures. This goal is addressed with current fault-based approaches that use catalogs of mapped faults such as the Community Fault Model (CFM) in Southern California [*Basili et al.*, 2013; *Plesch et al.*, 2007], which, however, lack the small-scale structures that may contribute significantly to short and intermediate-term hazards. Moreover, as illustrated by the M_w 6.7 Northridge, 1994 earthquake, a significant number of large earthquakes continue to occur on faults that were not yet mapped and were only revealed by the earthquake itself. In the case of Northern California, most of the seismicity remains unexplained by the set of mapped faults as shown for example in *Wesson et al.* [2003], where most events are labeled under “BKGD”, for “background”, whereas they seem to occur on well-defined fault structures. Moreover, such extensive fault catalogs do not necessarily exist in other parts of the world exposed to intense seismic hazard.

[4] Using the magnitude of recorded events to determine empirically their contribution to the amount of slip over each fault patch, an improved knowledge of the underlying fault network may allow one to infer average slip rates on each fault at geological timescales and convert them into long-term average seismicity rates [*Gabrielov et al.*, 1996], possibly considering the information given by paleoseismological studies (see, for example, *Frankel et al.* [2002, the National Seismic Hazard Mapping Program]; *Petersen et al.* [2007, the Uniform California earthquake rupture forecast model]; *Field et al.* [2008]). This approach is used to provide long-term time-dependent or time-independent forecasts.

[5] The usual, and necessary, trick used in existing earthquake forecasting methods thus consists in smoothing the spatial structure of the earthquake catalog, in order to approximate the geological complexity of the local fault network. Only recently, forecast models were proposed that attempt to combine both seismicity and fault data sets in a common approach, yet blurring the knowledge of the fault structure by smoothing techniques [*Hiemer et al.*, 2013; *Rhoades and Stirling*, 2012]. Smoothing is performed using only the 2-D set of epicenters (and not the 3-D set of hypocenters), and this process always involves a set of arbitrary choices or parameters. The simplest smoothing consists in superimposing a regular grid onto the target area, thus coarse-graining the fault network at a homogeneous (and arbitrary) spatial resolution. A softer method consists in smoothing the set of declustered events with Gaussian kernels, whose bandwidths are adapted to optimize the quality of smoothing according to some metric [*Zechar and Jordan*, 2010], by using adaptive kernels or those respecting the distance to their closest neighbors [*Hiemer et al.*, 2013]. In general, events are simply replaced by kernels that are added up over the whole space and normalized so that the integral of the spatial density of events is equal to the number of events in the catalog. In many implementations, a smoothing is considered as optimal when it maximizes the score of the forecasts on an independent data set. It follows that the smoothing parameters do not stem from independent geological or physical knowledge. They thus look more like hidden parameters of the forecasting technique as a whole. Moreover, the use of square cells or isotropic kernels is totally opposite to what could be expected to best approximate a set of plane segments, whose orientations vary in space (see, for example, *Gaillot et al.* [2002] and *Courjault-Radé et al.* [2009] for the spatial analysis of sets of epicenters using anisotropic

wavelets, inspired by a methodology initially developed by *Ouillon et al.* [1996, 1995] for maps of fault or joint traces). In some cases, the bandwidth of the kernels may also depend on the size of the local events or on their spatial density: the larger the latter, the finer the resolution.

[6] The well-documented multiscale organization of earthquakes and faults precludes any objective choice of the most appropriate spatial resolution to study their dynamics. The only characteristic scales in such systems are the size of the system itself (at large scales), and the scale below which scale invariance breaks down without producing bonus information; typically, this is the smallest distance between pairs of events, or the size of the smallest fault, or the width of geological and rheologically different layers [*Ouillon et al.*, 1996]. From a statistical physics point of view, one may argue that taking account of the numerous “microscopic” spatial details of the seismicity process may only deteriorate our ability to model their dynamics and provide efficient forecasts, which is then a good reason to perform a smoothing. Another obvious reason is that events are always spatially located up to some finite uncertainties. However, *Werner et al.* [2011] brought into the debate new interesting elements by noticing that accounting for small magnitude earthquakes (down to $M=2$) in the input data set increased the likelihood of the forecasts. As increasing the number of small-scale earthquakes allows one to take account of smaller-scale details of the fault network, it follows that the smoothed seismicity rate of *Werner et al.* [2011] closely reflects the best possible approximation of the fault network they could hope to get. This result echoes the conclusion of *Zechar and Jordan* [2008] and *Woessner et al.* [2011] who suggest that future seismicity-based techniques should also use this set of faults as a data input.

[7] No independent and accurate geophysical technique exists that provides a detailed and complete 3-D map of active fault networks. As a consequence, we rely in this approach on seismicity itself as the best proxy to image the current fault network. Continuous and recent progress in earthquake location techniques now allows the manipulation of rather precise spatial data. For example, as absolute locations used to feature uncertainties of the order of a few kilometers in Southern California are now reestimated using relative location algorithms, the (relative) uncertainties are now shrinking down to only a few tens of meters [*Hauksson et al.*, 2012; *Waldauser and Schaff*, 2008]. Nonlinear location algorithms [*Husen et al.*, 2007, 2003; *Lomax et al.*, 2009] even allow the direct sampling of the full probability density function (hereafter pdf) of the location of each event. It follows that seismologists now have the opportunity to access to the detailed topology of the active part of the fault network, provided they have the tools to estimate the position, size, and orientation of fault segments from the precise location of events listed in earthquake catalogs, i.e., to extract the full value from these golden data.

[8] *Ouillon et al.* [2008] recently proposed a new method of pattern recognition that reconstructs the active part of a fault network from the spatial location of earthquake hypocenters. It is inspired from the seminal k -means method [*MacQueen*, 1967], which partitions a given data set into a set of (a priori isotropic) clusters by minimizing the global variance of the partition. *Ouillon et al.* [2008] generalized this method to the anisotropic case with a new algorithm,

which, in a nutshell, fits the spatial structure of the set of events with a set of finite-size plane segments. The number of segments used is increased until the residuals of the fit become comparable to the average hypocenter location uncertainty. One can then estimate the position, size, and orientation of each plane segment. *Ouillon et al.* [2008] applied this algorithm to synthetic data sets as well as to the aftershock cloud of the Landers, 1992 event, in Southern California, for which they showed that 16 planes were necessary to provide a fit compatible with the average location errors. Moreover, extrapolating the set of plane segments to the free surface, the predicted fault traces showed a good agreement with observed fault traces of the Southern California Community Fault Model (CFM) and also allowed to map faults of significant size that are not reported in the CFM.

[9] The main shortcoming of the *Ouillon et al.* [2008] clustering method is its rough account of location uncertainties, assumed to be constant for the whole catalog. In this paper, we improve on this method by taking account of the detailed and individual location uncertainties of each event, which control both the fit through the use of the expected squared distance (ESD) between an event and a plane and the resolution at which the latter is performed. As the fitting method is still strongly nonlinear, different runs generally converge toward different local minima of the residuals. We thus introduce new methodologies to validate the obtained solutions, as systematic and automatic comparison with existing fault maps, if existent, is a very difficult exercise in particular because it lacks a precise metric. We thus present six validation schemes: two of them based on the residuals of the fit and four others based on the compatibility of the fault networks with known focal mechanisms. The new method is then tested on simple and more complex synthetic fault networks, as well as on a new catalog of the Landers area.

2. The Optimal Anisotropic Data Clustering (OADC) Method

[10] The new clustering method proposed here is based on a pattern recognition technique called k -means, shortly described in *Ouillon et al.* [2008] and in more details in *Bishop* [2006], *Duda et al.* [2001], and *MacQueen* [1967]. This technique makes no assumption about the shape of the individual clusters. In that sense, it can be viewed as an “isotropic” processing of data. When dealing with earthquakes, it is desirable to cluster data within structures that can be identified as faults. In that case, the minimum a priori information that may help to constrain the pattern recognition process is that the clusters we look for should be highly anisotropic, i.e., that their thicknesses should be very small compared to their other dimensions.

[11] The OADC method of *Ouillon et al.* [2008] provides an attempt to reconstruct fault networks using solely the information contained within seismicity catalogs. Compared with other strategies, e.g., the Community Fault Model (CFM) of the Southern California Earthquake Center, it defines a general method that can identify active fault segments without taking into account direct observations such as maps of fault surface traces and/or subsurface borehole data nor indirect observations like seismic reflection profiles to map deeper structures. *Ouillon et al.* [2008] also provide a discussion of other seismicity clustering techniques.

[12] The OADC method is directly inspired from the original definition of the k -means method, yet generalizes it to strongly anisotropic clusters, whose thicknesses are assumed to be very small. Each fault segment is thus approximated by a finite rectangular plane, characterized by its dimension (length and width), orientation (strike and dip), and position of its center. Earthquakes are handled as pure data points, while a uniform and isotropic location uncertainty ϵ is assumed to hold for all events.

[13] The general algorithm of the method is the following:

[14] 1. Initialize N_0 planes with randomly chosen center positions, orientations, and dimensions.

[15] 2. For each earthquake \vec{O} in the catalog, compute the distance from it to each plane \vec{C} , determine the closest plane, and associate the former to the latter. Earthquake locations are treated as points, and Euclidean distances to the finite planes are computed. This first partition provides us a set of N_0 clusters of events.

[16] 3. For each cluster, perform a spatial principal component analysis and use the eigenvalues and eigenvectors to define their new dimensions, orientations, and center positions. The thickness of each cluster is given by the square root of the smallest eigenvalue. The two other eigenvalues provide the length and width of the cluster (see *Ouillon et al.* [2008] for details).

[17] 4. Assuming a uniform catalog spatial location uncertainty ϵ , the computation stops if the thickness of each cluster is smaller than ϵ , as the dispersion of events across each plane can be fully explained by location errors. If there is at least one cluster for which the thickness is larger than ϵ , then proceed to step 5.

[18] 5. Split randomly the plane associated to the thickest cluster into m subplanes, increase N_0 accordingly by $m - 1$, and go back to step 2.

[19] This procedure, which is nothing but a nonlinear fitting technique, ensures that events will be partitioned into clusters with negligible thickness (up to location uncertainties), i.e., plane-like structures, which are the assumed a priori model for faults.

[20] Similarly to the classical k -means method, the OADC method may converge to a local minimum of the global clusters fit residual. One can solve this problem by running the clustering procedure several times, with different initial conditions, in order to explore the solution space and select the fault network model that achieves a genuine global minimum. However, as the method itself ensures that all fit residuals are smaller than location uncertainties, all solutions are therefore statistically equivalent. Picking one of them as the best one thus requires an independent validation process. Due to computational limitations, *Ouillon et al.* [2008] provided only ten runs on the Landers aftershocks data set; yet, they noticed that the method converged more often to one of the solutions than to any other (thus suggesting a validation based on the most frequently selected solution). For each solution, extrapolating all the planes they obtained to the free surface, thus generating the corresponding predicted surface fault traces, they noticed that the most frequent solution was also the one that fits best the observed natural fault traces in this area. While offering a validation procedure on an independent data set, this approach would prove cumbersome when dealing with much larger areas or with zones where no such fault traces maps or incomplete ones are available. Another drawback is

the subjectivity of the comparison, which is not based on any quantitative metric. The systematic validation of the obtained solutions is thus still an open problem.

[21] Another obvious limitation of the OADC method is the assumption made about location uncertainties, which are considered to be uniform and isotropic. This hypothesis is unrealistic since focal depth is often less well constrained than the epicentral location. Moreover, location uncertainty is strongly influenced by the velocity model error, the quality of waveform pickings, the station network geometry, etc. and is thus very heterogeneous in space and time [e.g., Husen and Hardebeck, 2010]. It thus follows that the clustering process should be more detailed in some areas and sparser in some others. The clustering method should take this heterogeneity into account.

3. Anisotropic Clustering of Location Uncertainty Distributions (ACLD)

[22] The original k -means method assumes that the uncertainty of the spatial location of data points is negligible. In the case of real physical systems, the story is different. For earthquakes, location uncertainty is an inherent property due to wave arrival time inaccuracy, velocity model errors, station network geometry, or outdated data sources like historical seismicity catalogs. When taking uncertainty into account, data can no longer be described as a point process, but as a more or less complex probability density function (hereafter pdf). Chau et al. [2006] claim that uncertainties can significantly affect the results provided by clustering techniques such as k -means. They thus introduce the *uk-means* algorithm (where “u” stands for “uncertain”, see supporting information), which incorporates uncertainty information and provides, when considering synthetic samples, more satisfying results than the standard algorithm.

[23] We now show how to extend the *uk-means* method of Chau et al. [2006] to the case where the cluster model \vec{C} is a plane, in the spirit of Ouillon et al. [2008], and the object to cluster \vec{O} is the pdf of an earthquake location. We term the new method the “anisotropic clustering of location uncertainty distributions” (ACLD).

[24] Chau et al. [2006] suggest using the expected squared distance (hereafter ESD), which, in our case, is defined as

$$\begin{aligned} d^2(\vec{O}, \vec{C}) &= \int_{\vec{x} \in \vec{O}} \|\vec{x} - \vec{C}\|^2 f(\vec{x}) d\vec{x} \\ &= \int_{\vec{x} \in \vec{O}} \inf_{\vec{c} \in \vec{C}} \|\vec{x} - \vec{c}\|^2 f(\vec{x}) d\vec{x} \end{aligned} \quad (1)$$

where $f(x)$ is the pdf of the earthquake location. While this distance is easily estimated in the case of an infinite plane \vec{C} , we also propose computationally efficient approximations in the case of a finite-size plane.

3.1. Expected Square Distance (ESD) Between a Probability Density Function and an Infinite Plane

[25] We consider an infinite plane within a Euclidean three-dimensional space. The coordinate system is chosen such that its origin is located on the plane, whose orientation is given by

two of the basis vectors, the third one being normal to it. Then, equation 1 can be rewritten as

$$d^2(\vec{O}, \vec{C}) = \int_{\vec{x} \in \vec{O}} \|x_3\|^2 f(\vec{x}) d\vec{x} \quad (2)$$

where x_3 is the third component of point $\vec{x} \in \vec{O}$. Noticing that

$$x_3^2 = (x_3 - k_3 + k_3)^2 = (x_3 - k_3)^2 + k_3^2 + 2(x_3 - k_3) \cdot k_3 \quad (3)$$

with k_3 being the third component of the centroid of \vec{O} and given that the contribution of the last right-hand term of equation 3 to the integral is zero, equation 2 becomes

$$d^2(\vec{O}, \vec{C}) = k_3^2 + \int_{\vec{x} \in \vec{O}} (x_3 - k_3)^2 f(\vec{x}) d\vec{x} \quad (4)$$

[26] The first term in the right-hand side is simply the squared distance between the centroid of \vec{O} and the infinite plane, while the second term is simply the variance of \vec{O} in the direction normal to the plane (which can be deduced from the pdf of \vec{O} and its covariance matrix). This is nothing but the variance decomposition theorem.

3.2. Expected Square Distance (ESD) Between a Probability Density Function and an Infinite Line or a Point

[27] Following a similar procedure when \vec{C} is a line, we can choose a coordinate system so that \vec{C} lies on the first axis. Then we get

$$\begin{aligned} d^2(\vec{O}, \vec{C}) &= k_2^2 + k_3^2 + \int_{\vec{x} \in \vec{O}} [(x_2 - k_2)^2 + (x_3 - k_3)^2] f(\vec{x}) d\vec{x} \\ &= \sum_{i=2}^3 k_i^2 + \sum_{i=2}^3 \int_{\vec{x} \in \vec{O}} (x_i - k_i)^2 f(\vec{x}) d\vec{x} \end{aligned} \quad (5)$$

[28] When \vec{C} is a point, we can choose a coordinate system so that \vec{C} lies at the origin. Then we get

$$d^2(\vec{O}, \vec{C}) = \sum_{i=1}^3 k_i^2 + \sum_{i=1}^3 \int_{\vec{x} \in \vec{O}} (x_i - k_i)^2 f(\vec{x}) d\vec{x} \quad (6)$$

[29] The interpretation of equations 5 and 6 is the same as for equation 4 except that we now compute the distance between the centroid and a line and use the relevant dimension for the variance decomposition. This last set of equations will prove very useful when approximating the distance between a pdf and a finite plane.

3.3. Expected Square Distance (ESD) Between a Probability Density Function (pdf) and a Finite Plane

[30] The anisotropic clustering of location uncertainty distributions (ACLD) method we propose still assumes that active fault segments can be modeled as rectangular finite planes. If it proves rather easy to compute the Euclidean

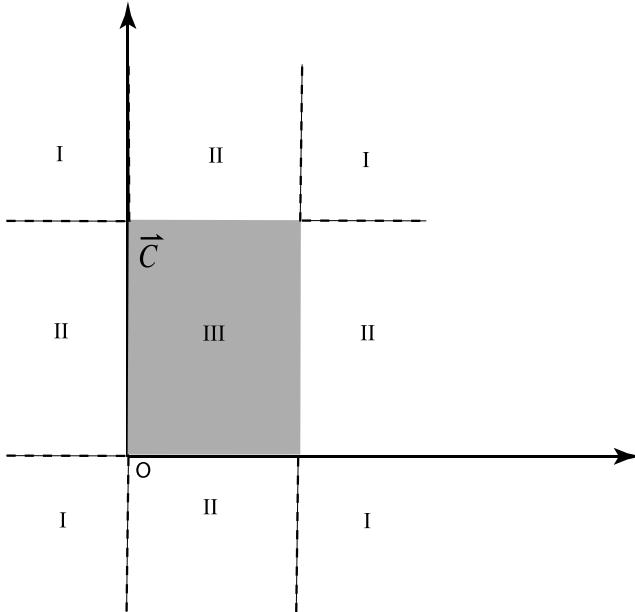


Figure 1. Partition of the 3-D space in order to compute the expected square distance of a pdf and a finite plane (shown in grey). The roman indices I, II, and III correspond to various approximations of the ESD (see main text, section 3.3).

distance between a point and a finite plane, the problem is a bit more difficult when observations are given through their pdfs. Indeed, we shall see that, using the variance decomposition theorem, we can only provide theoretical approximations to the expected squared distance between a pdf and a finite plane.

[31] Figure 1 illustrates the problem. The grey rectangle area represents a finite plane \vec{C} . Events may be located anywhere in the full 3-D space that surrounds it. We now consider any object \vec{O} in the 3-D space and its projection \vec{Q} along the direction normal to \vec{C} onto the infinite plane containing \vec{C} . The object \vec{Q} will be located within one or more of the nine sectors defined in Figure 1, each sector being indexed in roman numbers from I to III as shown in the figure. The object \vec{Q} can overlap several sectors depending on the shape of the support of its pdf.

[32] If \vec{Q} is completely included within sector III, then the ESD between \vec{O} and \vec{C} can be computed using equation 4, as the infinite plane assumption is valid. If \vec{Q} is completely included within a sector labeled II, the ESD is computed using equation 5 (after an appropriate change of coordinates) as the infinite line assumption is valid. If \vec{Q} is completely included within a sector labeled I, the ESD should be computed between the pdf and the closest corner of the finite plane, using equation 6. Indeed, a similar approach has been used in *Ouillon et al. [2008]* to compute the Euclidean distance between a given hypocenter and a given finite plane.

[33] In our case, the general problem is much more complex as we implicitly have to consider the distance between the finite plane and every point where the pdf of \vec{O} is defined. This implies that the projection \vec{Q} is characterized by a pdf that may overlap several distinct sectors, so that none of the above simple formulae 4–6 can be used anymore. In that

case, only a direct Monte Carlo approach provides an accurate estimate of the ESD. As it would prove computationally too heavy when handling large catalogs and sets of faults, we propose a simplification: We first consider only the centroid of \vec{O} and its own projection. If the latter is contained within sector III, we use formula 4 as an approximation to the ESD. If it is contained within a sector labeled II, we use formula 5. If it is contained within a sector labeled I, we use formula 6. This approximation is obviously wrong when the size of the finite plane is much smaller than the spatial extent of the domain where the pdf of \vec{Q} is defined. However, in practice we found that for most of cases, location uncertainties are much smaller than the size of potential fitting fault plane we can resolve.

3.4. Anisotropic Clustering of Location Uncertainty Distributions Algorithm

[34] Assume that an earthquake catalog provides the location of each event with a pdf. We can characterize the location with its centroid (hereafter, the hypocenter) and its covariance matrix. The new clustering algorithm we propose is the following:

[35] 1. Split randomly the earthquake catalog into 2 distinct subsets: the training set (which is the one to be fitted) and the validation set (which is the one used to qualify or discriminate different clustering models).

[36] 2. Initialize a number of N_0 faults with random positions, orientations, and dimensions.

[37] 3. For each earthquake in the training subset, associate the earthquake to the closest plane according to the ESD. We thus get a partition of events into a set of N_0 clusters.

[38] 4. For each cluster i , compute the covariance matrix of the locations of its associated hypocenters, and find its eigenvalues and eigenvectors. By doing so, the dimensions and orientations of each cluster can be computed. The smallest eigenvalue $\lambda_{i,3}$ provides the thickness of the corresponding cluster.

[39] 5. For each cluster, compute the average individual variance ε_i of the hypocenters' location pdf in the direction normal to the cluster.

[40] 6. For each cluster, compare its thickness $\lambda_{i,3}$ with the average location uncertainty ε_i of its associated events. If $\varepsilon_i \geq \lambda_{i,3}$ for all clusters, the computation stops, as location errors alone can explain the finite thickness of each cluster. We then proceed to step 8. If there is at least one cluster for which, then we proceed to step 7 as we need more planes to explain the data.

[41] 7. We split randomly the thickest cluster into m other planes, and go back to step 3 (increasing N_0 accordingly by $m - 1$).

[42] 8. We compute the residual of the fit of the validation data set conditioned on the fault network model of the training data set (from step 6).

[43] 9. We repeat steps 1–8 many times (typically several thousands) and rank all models according to their validation fit residuals obtained in step 8.

[44] For this study, in step 7 we use $m = 2$. The proposed algorithm accounts for individual event location uncertainties, both in the computation of the ESD between an event and the planes and in the criterion used to continue or stop the fitting process. The stopping criterion thus does not assume a spatially uniform location uncertainty, but is adapted to the case

of space-dependent location quality. This property is particularly welcome in the case of earthquakes for which location uncertainties heavily depend on the spatial structure of the station networks.

[45] One should also be aware that the full three-dimensional confidence interval is different from the confidence interval in 1-D. In order to compute the variance of the pdf in the direction normal to the plane, we have to project the 68% three-dimensional confidence ellipsoid onto that normal direction. Yet, after the projection, the confidence level increases to higher levels so that the correct quantiles have to be estimated [Press *et al.*, 2007, p. 811, Figure 15.6.3].

[46] By subdividing the data set, we implement a cross-validation technique to the predictive skill of the clustering approach. Our procedure separates randomly the full data set into two independent subsets, generates the fault model that fits the training data set, and evaluates it by estimating how well it predicts the independent validation set. The process is repeated several times, each trial corresponding to different training and validation sets, and we select the one with the best validation result. How to generate the training and validation data sets is a question in itself. On the one hand, if there are not enough earthquakes in the training set, it will lead to a spurious fit with a very bad validation score; on the other hand, if there are not enough earthquakes in the validation set, residuals may fluctuate and depend strongly on the particular choice of the validation set. Using 95% of the data as the training set and 5% as the validation set are standard values used in pattern recognition algorithms [Bishop, 2006]. Yet, from synthetic tests where the original fault networks are known, we checked that it generally provides robust results.

[47] The main assumption of this algorithm is that the hypocenter corresponds to the expectation hypocenter location [Lomax *et al.*, 2000]. In the framework of probabilistic earthquake location the hypocenter location is usually associated with the maximum likelihood point [Tarantola and Valette, 1982]. The assumption that the hypocenter is not very different from the maximum likelihood point would be valid if and only if the pdf of the location of the event is compact, i.e., small in size, which has no a priori reason to be true. We shall discuss later the conditions for which this assumption might be approximately valid in the case of natural earthquake catalogs.

3.5. Validation Strategies

[48] The new clustering method automatically explores a very large solution space. In order to find the “best” solution, we follow a purely statistical strategy, i.e., cross validation. However, other validation strategies might be more appropriate. In the following, we will introduce three other criteria: one residual-based statistical strategy called Bayesian information criterion (BIC) [see Schwarz, 1978] and two metrics based on observed focal mechanisms.

3.5.1. Bayesian Information Criterion

[49] BIC is a commonly used statistical criterion for model selection that takes both the likelihood function and model complexity into consideration. During clustering, it is possible to increase the likelihood by adding more faults, at the cost of increasing the complexity of the model. By adding a penalty term for the number of faults, the BIC merges the likelihood and complexity of the solution together. Assuming that the

distribution of earthquakes across the fitting planes is a normal distribution, the BIC can be expressed as

$$\text{BIC} = n \cdot \ln(\hat{\sigma}) + k \cdot \ln(n) \quad (7)$$

where n is the number of events used for the fit, $\hat{\sigma}$ is the unbiased variance estimation of the earthquake distribution across the fitting planes, and k is the number of faults in the tested model. Thus, by minimizing the BIC, we may find the best network from the solution space that provides both a large likelihood and a simple model structure. The difference with the cross-validation scheme is that the latter is performed using the validation data set, whereas the BIC uses the training data set. It is also important to notice that, during the clustering process, we randomly partition the whole data set into training and validation sets. It means that, for each clustering run, the training set changes so that the computed BIC is not strictly derived from the same training set. However, considering that we deal with large data sets among which 95% of each single one is used as training sets, the BIC remains a robust estimator.

3.5.2. Focal Mechanism μ Metrics

[50] The focal mechanism of an earthquake describes the potential orientations of the rupture plane and slip vector. If events are clustered together on a given fault plane, we may expect them to be characterized by similar focal mechanisms, the latter being also consistent with the orientation of the fitting plane. This provides a mechanical approach to validation. At the end of each fit, we thus adopt the following procedure:

[51] 1. For each cluster, select the available focal mechanisms of events.

[52] 2. For each focal mechanism, compute the normal vector to each of the two nodal planes.

[53] 3. For each nodal normal vector, compute its dot product with the vector normal to the cluster (defined as pointing upward). If one of the dot products is negative, replace the nodal normal vector by its opposite and change the sign of the dot product.

[54] 4. From both nodal normal vectors, choose the one that maximizes the dot product.

[55] 5. Once steps 2–4 have been fulfilled for each event of the cluster, stack all the selected nodal normal vectors, and compute the angle μ between the resultant and the normal vector to the cluster.

[56] Step 5 is performed after weighting each selected nodal normal vector according to the magnitude M of the corresponding event. The weight is taken as 10^{aM} . If $a=0$, then all events have the same weight and the measured angular discrepancy is mainly controlled by the smallest events. If $a=1/2$, then each event is weighted proportionally to its empirically assumed slip amount, while it is weighted by its energy or moment if we set $a=3/2$ (and in that case the angular discrepancy is controlled by the largest event in the cluster). As the plane segments we infer from our network reconstruction are, among other parameters, characterized by a size (i.e., an area) and a direction, it thus makes sense to compare them with the average direction of rupture events weighted by their individual rupture area. This is why we choose to set $a=1$. Moreover, if the local Gutenberg-Richter b value is close to 1, each magnitude range contributes equally to the estimated angular discrepancy, yet, it is not a necessary assumption of the methodology.

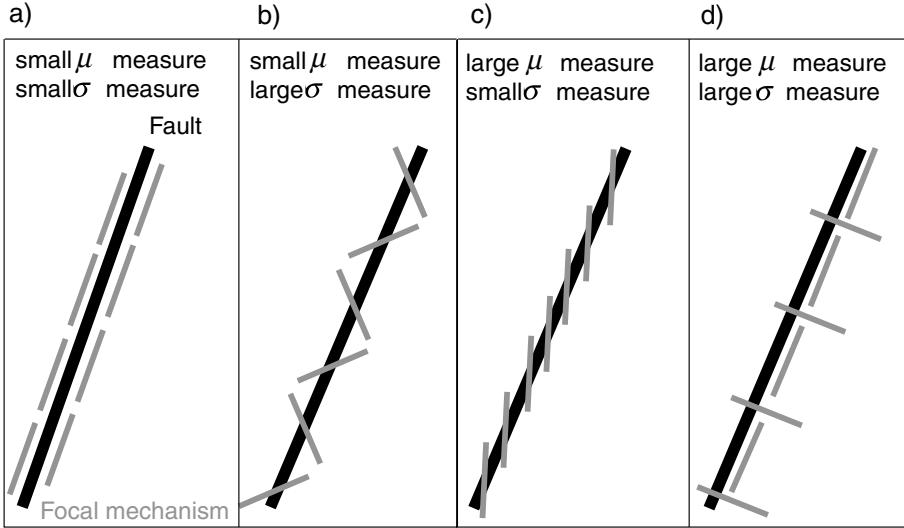


Figure 2. Examples of microstructure and macrostructure relationships in fault zones to justify the use of different criteria based on focal mechanisms (see main text, sections 3.5.2 and 3.5.3). Thick black line denotes general orientation of the fault zone (macrostructure); thin gray lines indicate orientation of shorter individual fault planes within the fault zone (microstructure).

[57] We first define a weighted average normal vector to the selected nodal plane of events on fault plane F_i as

$$\vec{V}_{E_i} = \frac{1}{\left\| \sum_{k=1}^{m(i)} \vec{v}_{E_{i,k}} \cdot 10^{\alpha \cdot M_{i,k}} \right\|} \sum_{k=1}^{m(i)} \vec{v}_{E_{i,k}} \cdot 10^{\alpha \cdot M_{i,k}} \quad (8)$$

where $m(i)$ = the number of events in fault plane F_i and $\vec{v}_{E_{i,-}}$ = the normal vector to the selected nodal plane of a given event on fault plane F_i . We then define a global angular discrepancy of the full set of planes as the μ_{fault} measure. It is formally expressed as

$$\mu_{fault} = \frac{\sum_{i=1}^n \cos^{-1} |\vec{V}_{F_i} \cdot \vec{V}_{E_i}|}{n} \quad (9)$$

where n = the number of fault planes and \vec{V}_{F_i} = the normal vector to fault plane F_i . The weighting strategy of equation 9 implies that we simply compute an average angular misfit over all faults (hence the associated subscript on the left-hand side). Similarly, we can also perform the average over all events. We obtain

$$\mu_{event} = \frac{\sum_{i=1}^n m(i) \cos^{-1} |\vec{V}_{F_i} \cdot \vec{V}_{E_i}|}{\sum_{i=1}^n m(i)} \quad (10)$$

[58] Minimizing both estimators will select networks where the orientation of inverted fault planes is the closest to the average orientation of the focal mechanisms. In summary, the μ metric measures the magnitude weighted average direction of the normal vectors of the “observed” focal mechanisms to the normal vector of the fault plane derived within the clustering approach.

3.5.3. Focal Mechanism σ Metrics

[59] Events grouped together by our fitting procedure may also feature roughly similar focal mechanisms, whose orientation may be different from the one of the fitting plane (see sketch in Figure 2 and explanations below). Following the same procedure as above from steps 1 to 4, we change step 5 as:

[60] 6. Once steps 2–4 have been fulfilled for each event, stack all the selected nodal normal vectors, and compute the average angle between each individual selected nodal vector and the resultant stacked vector.

[61] The associated measures are defined σ_{fault} and σ_{event} , depending on the way they are averaged. They are similar to standard deviation in statistics; yet, we compute them using the L1 norm (and not the L2 norm). The reason is that, in the case when the distribution of angles is not Gaussian but fatter tailed, using the L1 norm provides results less sensitive to large outliers. Using the same notations as above, the mathematical expressions are

$$\sigma_{fault} = \frac{\sum_{i=1}^n \frac{1}{m(i)} \sum_{j=1}^{m(i)} \cos^{-1} |\vec{v}_{E_{i,j}} \cdot \vec{V}_{E_i}|}{n} \quad (11)$$

$$\sigma_{event} = \frac{\sum_{i=1}^n \sum_{j=1}^{m(i)} \cos^{-1} |\vec{v}_{E_{i,j}} \cdot \vec{V}_{E_i}|}{\sum_{i=1}^n m(i)} \quad (12)$$

σ measures the angular difference from each single normal vector to the fault plane from the clustering approach and then averages, which results in a quite different metric.

[62] Figure 2 shows examples of applying μ and σ measures. On each plot, the black line indicates the trend of the fault zone, while the gray lines indicate the potential orientations of shorter individual ruptures within the fault zone, all

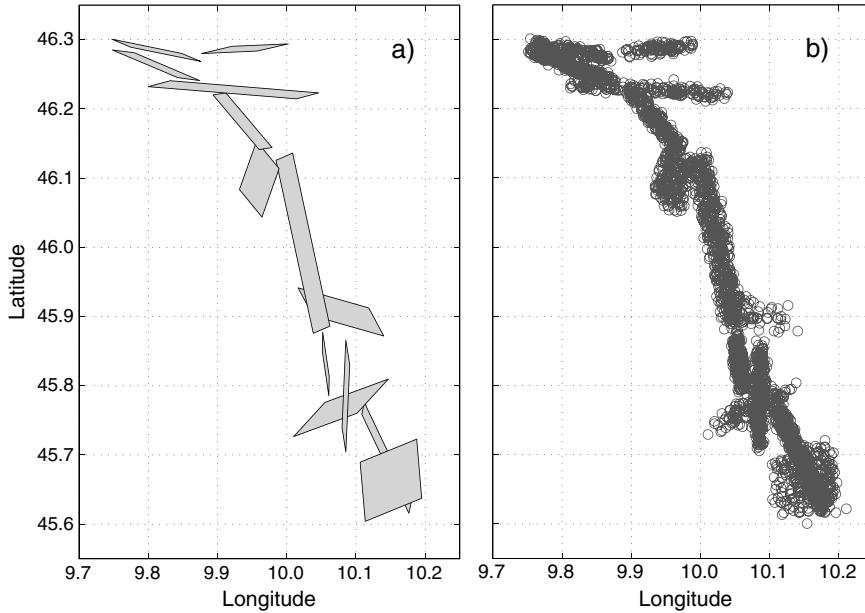


Figure 3. Synthetic data derived from the analysis of *Ouillon et al.* [2008] on the Landers fault network. (a) Fault network consisting of 13 faults. (b) Epicenter map of the synthetic relocated 3103 events.

events being clustered in the same macroscopic fault zone (see section 6 for a further discussion of the influence of the fault zone complexity on the results of clustering). When the rupture planes are quasi colinear with the fault trend, then both μ and σ values are small (Figure 2a). Figure 2b shows a series of planes for which orientations oscillate around the trend of the fault. In that case, the μ value is still small while the σ value is larger. Figure 2c shows the case of an en echelon distribution of rupture segments, which will provide a finite and possibly large μ value and a very small σ value. The last example (Figure 2d) shows a series of alternating conjugate rupture planes, which will be associated with large values of both μ and σ . These two measures derived from focal mechanisms can quantify the degree of agreement of the reconstructed fault network with local focal mechanisms. They provide tools in model selection with consideration of tectonic knowledge compared to pure statistical approaches such as cross validation or BIC.

4. Tests of the ACLUD Method on Synthetic Catalogs Featuring Location Uncertainties

[63] The previous section has introduced a new clustering scheme to automatically reconstruct fault structures from seismicity catalogs including location uncertainty information. We apply the approach to synthetic catalogs to understand its sensitivity to different structural complexities.

4.1. Generation of Data Sets

[64] Locating earthquakes results in a posterior probability density function of an event location [*Moser et al.*, 1992; *Tarantola and Valette*, 1982; *Wittlinger et al.*, 1993]. The pdf may possess any arbitrary shape and may be visualized using scatter density plots, which are obtained by drawing samples from the posterior pdf with their number being proportional to the probability [*Husen et al.*, 2003; *Lomax et al.*, 2000]. From these samples, the 68% confidence ellipsoid can

be computed by a singular value decomposition of the corresponding covariance matrix and consists in a rough approximation of the spatial uncertainty of the location estimate. The expectation hypocenter is at the center of the confidence ellipsoid, and the maximum likelihood hypocenter will always be located within the densest part of the pdf, so that both locations do not necessarily coincide.

[65] In this section, we generate synthetic earthquake catalogs using the NonLinLoc software package [*Lomax et al.*, 2000], version 5.2, <http://alomax.free.fr/nlloc/>. Compared to traditional, linearized approaches, NonLinLoc is superior in that it computes the posterior pdf using nonlinear, global searching techniques. The general method we use to generate a synthetic earthquake catalog is the following. We first impose the geometry of the original fault network, which consists in a collection of rectangular planes with variable locations, sizes, and orientations. We then assume that all earthquakes occur exactly on those planes and generate P waves. We then randomly distribute a given number of earthquakes on those planes. For each event, we randomly choose a set of 11 stations which constitute a set of observations. For a given velocity model, theoretical travel times between the true hypocenters and a set of given stations are computed. Random perturbations are added to the arrival times mimicking the uncertainty in picking waveform onset, which allows us to proceed to the inverse problem of computing the location of the events as well as their uncertainties by using NonLinLoc.

[66] To generate the set of associated synthetic focal mechanisms, we first assume that the rake of the slip vector on each plane is zero. For each event, the strike and dip are assumed to be identical to the ones of the input plane to which it belongs. We then add an independent Gaussian random perturbation respectively to the strike, dip, and rake of the event. Those perturbed angles are then used to compute the strike and dip of the auxiliary plane, thus providing a complete focal mechanism.

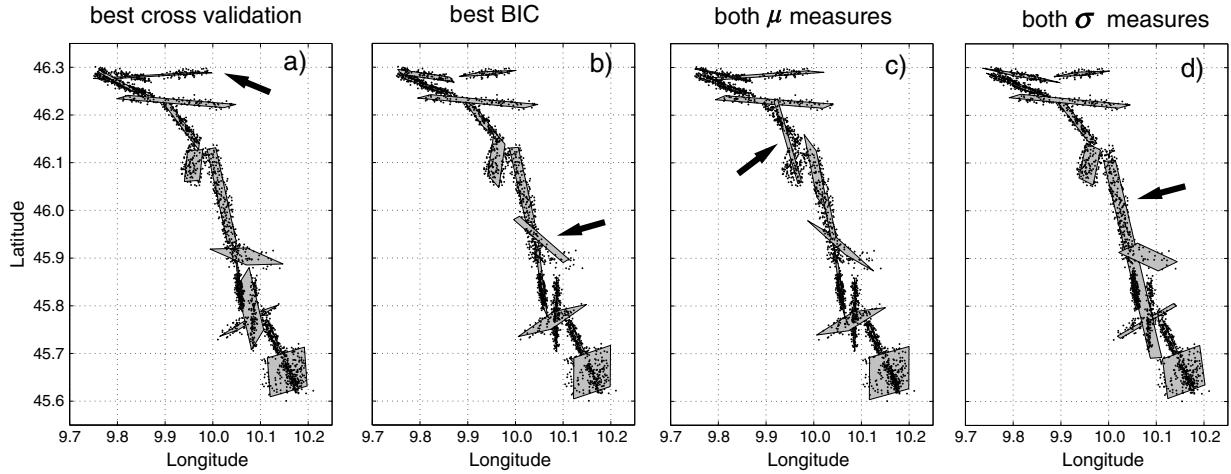


Figure 4. Result of our clustering method applied to the synthetic data consisting of 13 original fault planes and 3103 events presented in Figure 3. Planes pointed by arrows are spurious faults discussed in section 4.2.

[67] Note that we did not take account of the possible errors on the velocity model, which would provide systematic errors on both locations and focal mechanisms.

[68] The catalog of relocated hypocenter locations including their scatter density clouds is then fitted with a set of finite planes, using the ACLUD algorithm as defined in the previous section. The best solution which depends on the validation technique is then compared to the original input fault network.

[69] As a first test, we generated a very simple synthetic data set consisting in three vertical faults featuring 4000 events in all (thus similar to the one studied in *Ouillon et al.* [2008]) and characterized by their full pdf. The new clustering technique we propose successfully reconstructed the fault network whatever the validation criterion we used (see supporting information). We shall now test it on a more realistic and complex case.

4.2. Synthetic Catalog With Complex Geometry Inspired From Ouillon Et Al.

[70] This synthetic data set outlines a more complex and realistic case. Figure 3a shows the structure of the reconstructed fault network in the area of the 1992 M_w 7.3 Landers earthquake by *Ouillon et al.* [2008]. It features 13 planes with a dip larger than 45° (the three other planes, dipping less than 45° , have been removed as they certainly are spurious planes—see *Ouillon et al.* [2008]). The original catalog used in *Ouillon et al.* [2008] includes 3103 events, which we now assume to occur randomly and uniformly on those planes. We define a virtual station network, similar to the simpler one used in the example shown in the supporting information, in order to compute theoretical wave travel times to 11 randomly chosen stations, and add Gaussian errors with a standard deviation of 0.1 s to simulate picking errors. Figure 3b shows the spatial distribution of the relocated 3103 events. To generate the set of synthetic focal mechanisms, we add an independent Gaussian random perturbation respectively to the strike, dip, and rake of each event with a standard deviation of 10° . Those perturbed angles are then used to compute the strike and dip of the auxiliary plane, thus providing a complete focal mechanism. For an 80° dipping fault, we performed a Monte Carlo simulation in order to compute the

angular difference between the normal vectors of the correct and perturbed mechanisms. We found that the mean value of the angular difference is 11.5° , which is comparable with the quality Class A and B focal mechanisms computed by the HASH (HArdebeck and SHearer) approach [*Hardebeck and Shearer*, 2002]. As focal mechanisms are characterized by a 3-D orientation, their statistics is very different from linear variables. Our approach is thus a simplified procedure to simulate uncertainties. *Kagan* [2005] introduces a more rigorous way to randomly rotate double-couple focal mechanisms in 3-D, which one has to use if simulating broader distributions than ours.

[71] Note that, in section 3.5, we defined two statistical measures derived from focal mechanisms, which can be used to evaluate each reconstructed fault network. We can also assess the individual contribution of each cluster with respect to those global measures. We then similarly define for each cluster two individual measures of focal mechanism consistency, μ_F and σ_F :

$$\mu_F = \cos^{-1} \left| \vec{V}_F \cdot \vec{V}_E \right| \quad (13)$$

$$\sigma_F = \frac{1}{m} \sum_{i=1}^m \cos^{-1} \left| \vec{v}_{E_i} \cdot \vec{V}_E \right| \quad (14)$$

where m = number of events within fault F , \vec{V}_F = normal vector to the given fault plane F , \vec{V}_E = weighted average normal vector to the selected nodal plane of events on fault F , and \vec{v}_{E_i} = normal vector to the selected nodal plane of a given event on fault F .

[72] A large μ_F value indicates that the average focal mechanism rupture plane deviates significantly from the fitted fault plane. A large σ_F value indicates a significant dispersion of the orientations of focal mechanisms within the cluster.

[73] We performed 6000 runs with different initial conditions of the random number generator which controls the fault splitting step, and obtained as many solutions. We now discuss the results obtained using the six validation techniques discussed in section 3.5.

[74] 1. Cross validation: Figure 4a shows the selected reconstructed network, featuring 14 planes. One can notice

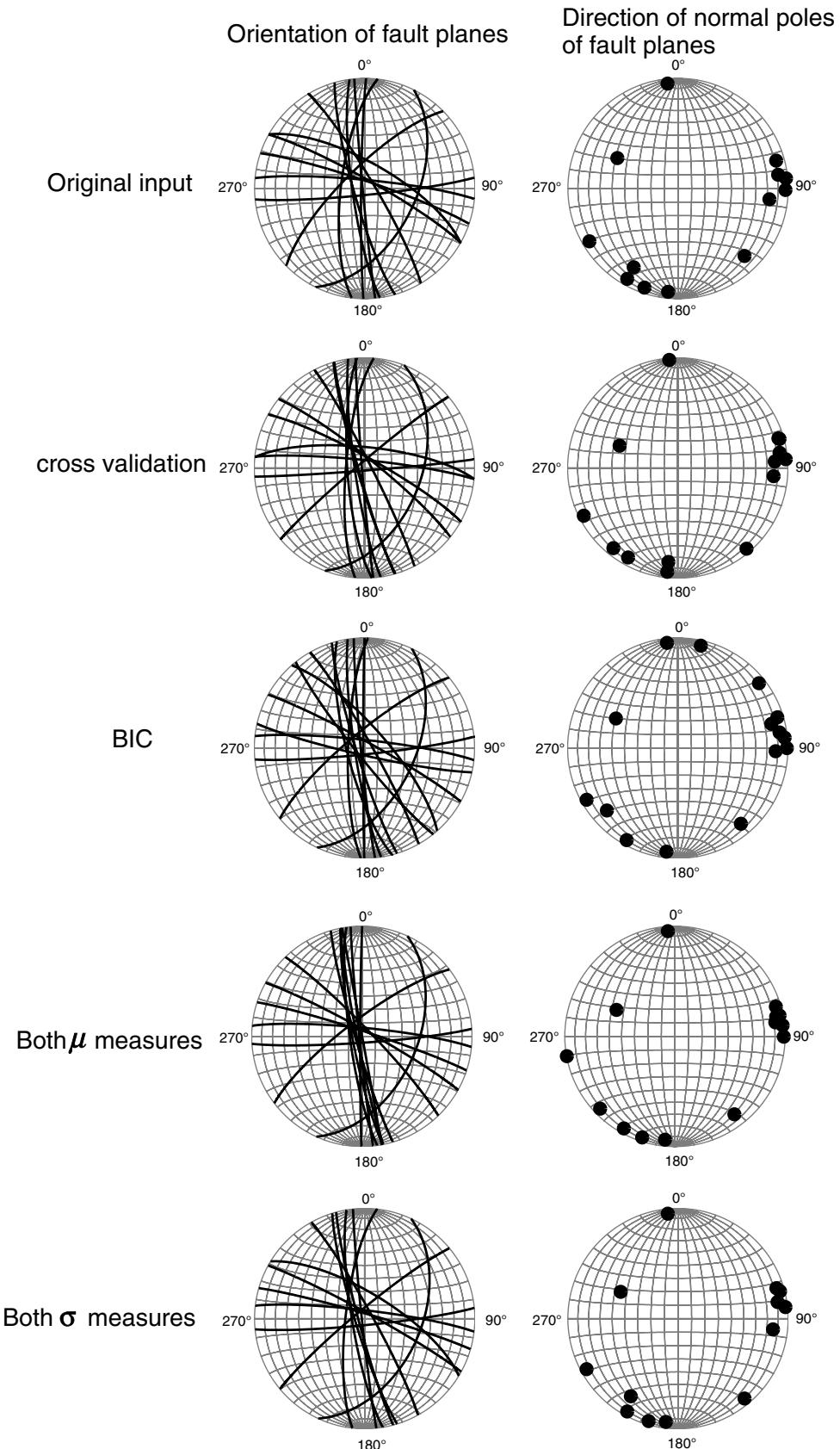


Figure 5. Stereo plots of the original input network and solutions chosen by the six validation criteria. (left column) Curves indicate the orientations of fault traces. (right column) Dots show directions of the normal poles of fault planes.

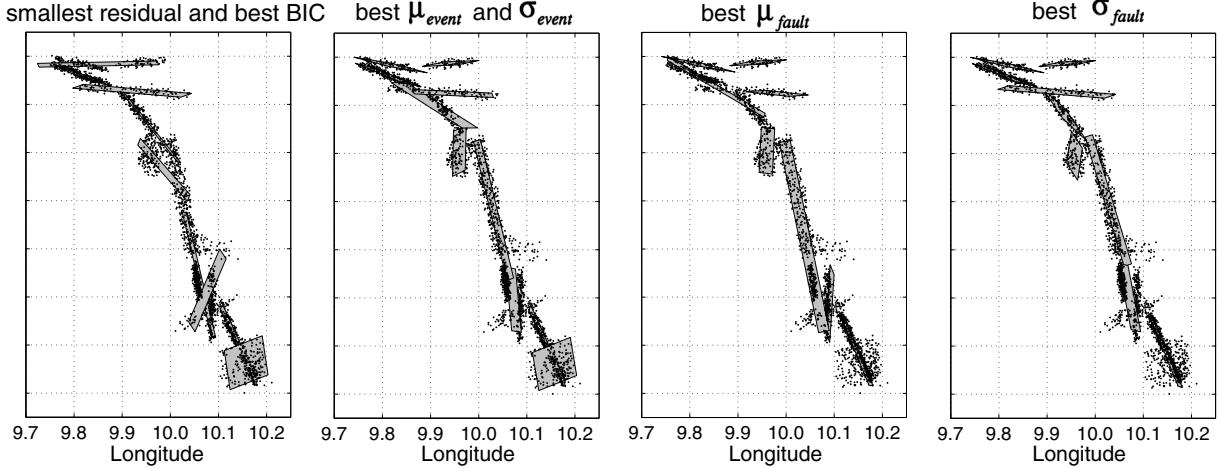


Figure 6. Result of the OADC clustering method of *Ouillon et al.* [2008] on a synthetic data set consisting of 13 original faults and 3103 events.

that two faults in the northern end are merged into a single plane. This is due to the fact that locations quality in this region is deteriorated due to a poor station coverage at the northern end. Such a poor coverage also occurs for the southern end, where the two crossing faults are reconstructed as a set of three faults. This kind of local overfitting is often observed in such situations and is due to the splitting step of the clustering process.

[75] 2. BIC: Figure 4b shows the selected reconstructed network, featuring 15 planes, which is different from the one selected by cross validation. Whereas the structure is now correctly inverted in the northern part, one can observe a small fault in the middle region pointed by the arrow, whose orientation is clearly rotated clockwise compared to the original synthetic network (Figure 3a). The reason is that the BIC gives more weight to the fault planes featuring more events. The density of events on this fault is the smallest among all 15 faults (for which this parameter ranges from $0.6/\text{km}^2$ to $5.0/\text{km}^2$). The reconstruction of such low event density faults can be unstable as their weight in the global criterion is very small. We also noticed that its individual σ_F value is the largest (for which this parameter ranges from 12° to 29°), indicating that the focal mechanisms of events clustered on this fault are very scattered.

[76] 3. μ metrics: both μ_{event} and μ_{fault} metrics select the same solution, shown in Figure 4c, featuring 14 planes. One can observe that two faults in the northern middle region have been merged into a single one (indicated by a small arrow). The distribution of μ_F and σ_F values of all 14 planes range from 1° to 43° and 12° to 23° , respectively. The individual μ_F value and σ_F value of these merged faults are both the largest over all 14 faults. This indicates that the focal mechanisms of the events clustered on this fault are neither consistent with each other nor with the orientation of the fitting plane. This thus makes the fault suspicious. More runs would be necessary to sufficiently sample the solution space and get a fully correct solution.

[77] 4. σ metrics: Figure 4d shows the reconstructed network chosen by both σ_{event} and σ_{fault} metrics, featuring 13 planes. Three faults in the central region are merged into a single large fault (see the arrow). This comes from the fact that the orientations of those three faults are very similar.

The individual μ_F and σ_F values of this merged fault are close to the average of the values obtained on the other planes. We thus have no way to diagnose this cluster as abnormal. This may stem from the fact that the faults that generate those events are located close to each other and feature orientation differences less than the uncertainties on the focal mechanisms orientations.

[78] Figure 5 shows the stereo plots of the original input faults and of the four solutions favored by the six different criteria. Plots in the left column indicate the orientations of fault traces. Dots in the right column indicate the directions of the normal vectors to the fault planes. Qualitatively, there

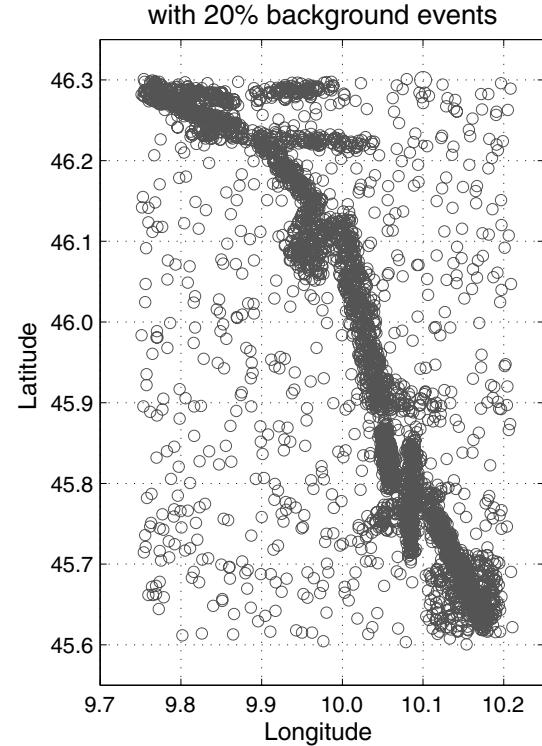


Figure 7. Epicentral map of the synthetic data set with 20% background events, giving a total of 3724 events.

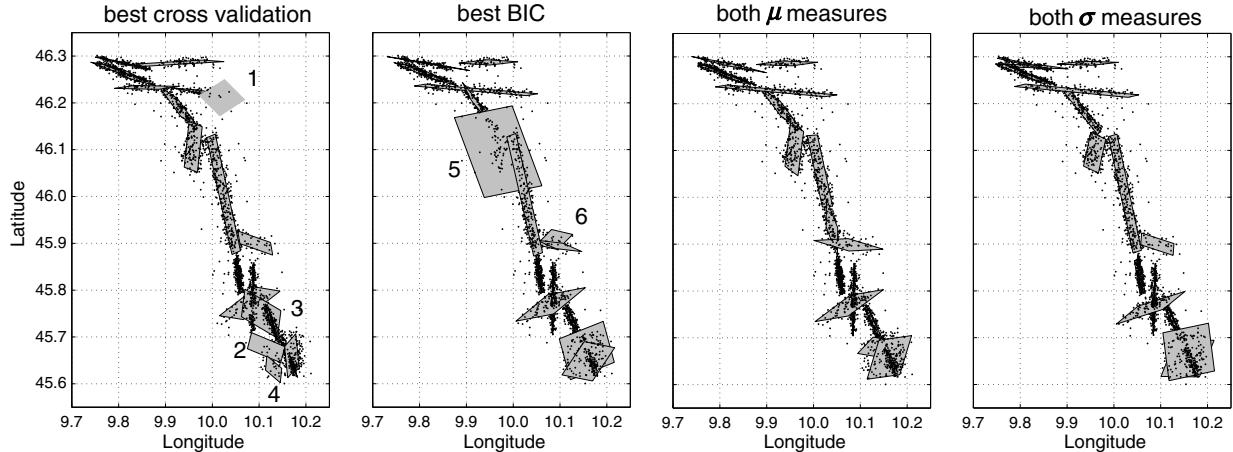


Figure 8. Result of our clustering method applied to the synthetic data set consisting 13 original faults with background seismicity. Solutions chosen by cross validation and BIC feature horizontal planes pointed by numbers are discussed in the text.

is a nice agreement between all the reconstructed networks and the true network (Figure 5, first row).

[79] This little example shows that inverting a complex but realistic structure, given realistic location uncertainty estimates, is not an easy task. However, the inverted networks, if not identical to the original one, are very similar to the original synthetic ones using the selection criteria. All validation criteria feature reasonable solutions: None of them is particularly better or worse than any of the others and the selections based on pure statistical techniques give similar fault networks as those based on tectonic constraints.

4.3. Comparison of the ACLUD Method With the OADC Method

[80] The OADC method uses a single, uniform, and isotropic location uncertainty for the whole catalog as the clustering stopping criterion. For the synthetic Landers catalog, we computed an average location uncertainty of 1.10 km. Using this value as the stopping threshold, we performed 6000 runs using the code of *Ouillon et al.* [2008]. The OADC method does not feature the cross-validation procedure, so that all events are used as the training data set. However, we can still rank all 6000 solutions based on their final clustering global residuals. Figure 6 shows the four solutions chosen by the six following criteria: best global clustering residual, BIC, and the four focal mechanism criteria previously defined. All those four solutions selected from different criteria clearly miss the small-scale structure of the network. Obviously, clustering has been forced to end too early due to using an inappropriate average location uncertainty estimate, especially in the central region. As location uncertainties in the central region are smaller than close to the northern and southern edges (due to a better station coverage), the stopping criterion that resembles the location uncertainty should be smaller in the central region than in the edge regions. Clustering thus stopped too early in the central region and made the structure coarser. Comparing with our new method, we thus clearly see the advantage of using the true location uncertainty of each event. Comparing the four solutions, we notice that the three of them chosen by

focal mechanism criteria are superior to the ones chosen by both the global clustering residuals and the BIC criteria (see Figure 6). These three solutions cover most of the input fault planes, yet do not include planes sampled by a small numbers of events. However, despite its simplicity, the main advantage of OADC is its fast convergence.

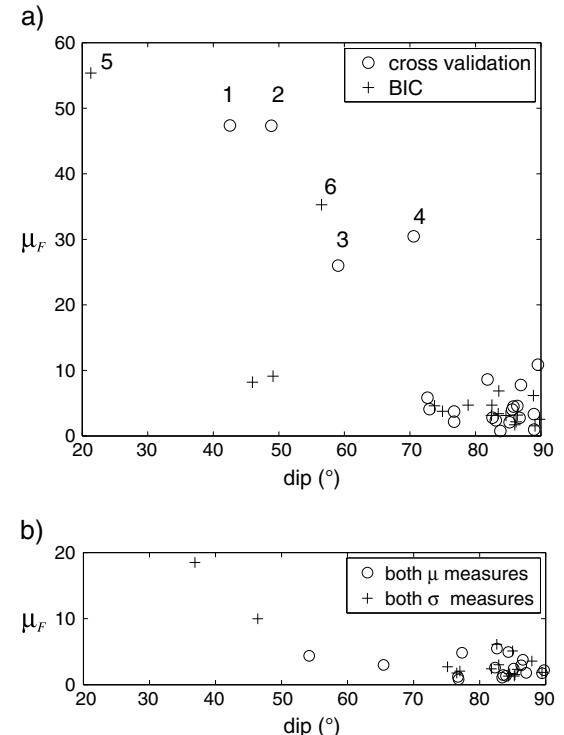


Figure 9. A function of dip (μ_F value) of each reconstructed fault for solutions chosen by different validation criteria. The synthetic data set consists of 13 original faults with background seismicity. Solutions chosen by cross validation and BIC feature horizontal planes with large values pointed by numbers and are discussed in the text.

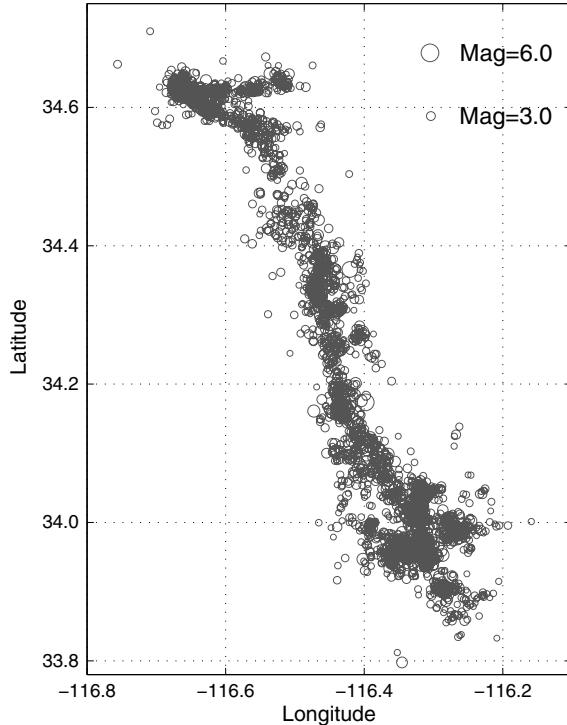


Figure 10. Epicentral seismicity map of the Landers area, 1984–2004. There were 3360 events chosen with magnitude 2, with more than 11 observations, located within an area well covered by the station network (primary azimuthal gap smaller than 180° , and ratio of the epicentral distance to the closest station over focal depth smaller than 1.5).

4.4. Synthetic Data With Background Events

[81] The previous section showed that our technique is able to reasonably reconstruct the structure of the synthetic fault network. We now test a new assumption where the catalog of events consists in the same set as before, but now we add background events. In nature, such events also occur on faults but the latter are, for our approach, undersampled by seismicity; thus, a clustering technique cannot reconstruct the structure. Specifically, we add another 20% background events to the synthetic data set uniformly distributed in the 3-D space (see Figure 7). The latitude, longitude, and depth ranges are identical to the ones of the fault-related events, providing a total number of 3724 events. For the sake of simplicity, their focal mechanism is chosen randomly among the set of the original 3103 events.

[82] Our new clustering technique follows the same approach as the OADC method to detect and remove background events. The detection is based on a local density criterion, as well as on the impossibility to associate an event with a given cluster without increasing too much its thickness. However, background events are not removed from the data set if they are located close to a fault, as they are then undistinguishable from other events.

[83] Results obtained using the different selection procedures are shown in Figure 8, after 6000 runs. Both purely statistical criteria (cross validation and BIC) select models with clearly spurious faults. For example, for cross validation, we observe a large nearly horizontal plane in the northern area,

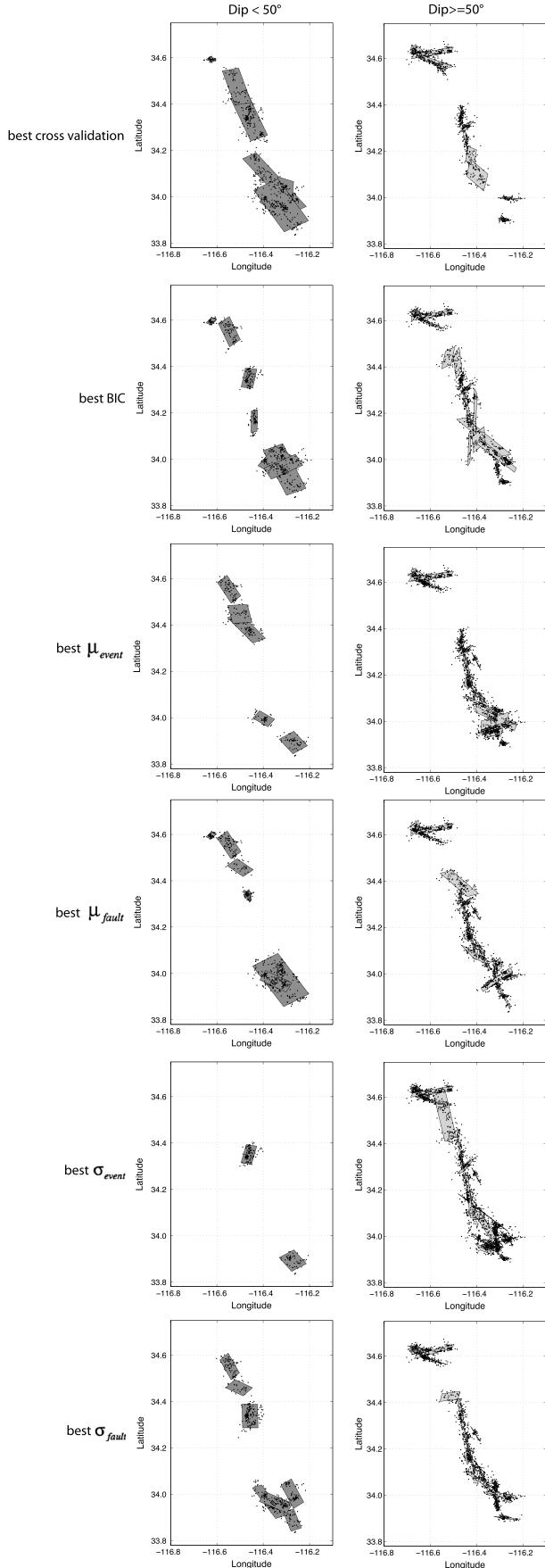
while in the southern region, original planes are divided into many small planes. Similarly, for the solution selected by the BIC, a large nearly horizontal plane is generated at latitudes 46.0° – 46.2° . Those low-dipping planes are indicated by numbers on Figures 8 and 9. The best results emerge when using models selected by criteria based on focal mechanisms. Looking at the properties of each cluster (see Figure 9), we notice that the reconstructed horizontal faults marked 1 and 2 have very large μ_F values. This suggests that these shallow-dipping planes disagree with their associated focal mechanisms. Results chosen by cross validation and BIC clearly show the effect of these nearby background events, which distort the inverted network and require to introduce spurious shallow-dipping faults to decrease the variance of the fit. In contrast, due to the fact that background events come mostly with arbitrary mechanisms, the validation criteria based on focal mechanisms detect more efficiently the associated inconsistencies and favor more realistic solutions.

4.5. Summary of Synthetic Tests

[84] The synthetic tests show that our new ACLUD method successfully reconstructs fault networks, both in the case of simple or more realistic and complex structures. The tests show that, due to location uncertainties, faults that are close in space and orientation may merge into a single structure. Comparing with the previous OADC method proposed by *Ouillon et al.* [2008], the new method improves the results by considering location uncertainties of each individual event, thus allowing us to invert the structure more finely within areas benefiting from a better station coverage. The new method also improved the validation step, as we automated the computation of six criteria, two of them being purely statistical indices of the fit (cross validation and BIC), the four others being based on the comparison between the inverted network and the observed focal mechanisms. While all those criteria provide reasonable selected models in the absence of background events, criteria based on focal mechanisms outperform the others when such background events are present. We even obtain better solutions when including background events, which may be due to a different exploration of the solution space. For real data sets, this implies performing an extensive simulation effort to reconstruct a fault network, similar to larger-scale Monte Carlo simulations. The multiple selection criteria and their characteristics also suggest that the technique does not allow us to pinpoint single best solutions but rather emphasize that possible solution groups exist, which are likely a result of undersampling of the structures with earthquakes.

5. Application to the Landers Aftershock Series

[85] We now apply our new clustering technique to a real data set in the area of the 1992 M_w 7.3 Landers earthquake, already studied by *Ouillon et al.* [2008]; this allows for a comparison of results. The catalog we used has not been published by the Southern California Earthquake Data Center and we obtained the permission to use this data set by E. Hauksson (California Institute of Technology, personal communication, 2010). The catalog has been located using the NonLinLoc method described in the supporting information. It contains 20 years of data from 1984 to 2004, with depth ranging from 1.37 km to 26.99 km. This catalog neither features the complete description of the original pdf of event



locations nor the corresponding covariance matrices that we need to input into our clustering scheme. Uncertainty is simply characterized by the lengths and orientations of the axes of the 68% confidence ellipsoid. Note that the corresponding derivation of the covariance matrix can be rigorously achieved only when the location pdf is Gaussian, a condition which generally holds only in areas well covered by a dense network of stations [Husen and Hardebeck, 2010; Lomax *et al.*, 2009]. We assume that this is the case in the Landers area, due to the presence of numerous stations belonging to the permanent Southern California network, as well as due to the set of temporary stations installed during the Joshua-Tree-Landers earthquake sequence. This is also the reason why we selected a subset of events that are most likely to be located with Gaussian uncertainties, i.e., those whose locations are particularly well constrained according to the criteria we defined in a companion work (Wang *et al.*, in preparation). We finally retained only events located using more than 11 stations, with local magnitude $M \geq 2$, and located within an area well covered by the station network (primary azimuthal gap smaller than 180° , ratio of the epicentral distance to the closest station over focal depth smaller than 1.5; see Bondár *et al.* [2004]), yielding a final subset of 3360 events (see Figure 10), comparable in size with our most complex synthetic example.

[86] The focal mechanism catalog we used is computed by the HASH method, using the locations derived by waveform cross correlation and a 3-D velocity model [Hauksson *et al.*, 2012; Yang *et al.*, 2012]. We only used the quality Class A and B focal mechanisms that show to about 60% focal mechanism errors of up to 20° [Hardebeck and Shearer, 2002]. Note that we used locations from the unpublished catalog derived with NonLinLoc to for clustering; we associated the focal mechanisms using the event IDs to apply the validation metrics. The focal mechanism error provided is an average of the uncertainties in strike, dip, and rake, mainly governed by the uncertainty of the rake angle. Given that the solutions are provided by using first-motion polarities and S wave amplitude ratios, we assume that the actual uncertainties of the strike and dip, which are important for our measures μ and σ , may be smaller.

[87] As the clustering technique can be considered itself partly as a stochastic process, we performed 30,000 different runs in order to reasonably sample the complex landscape of the solution space. Figure 11 shows the fault networks corresponding to the best solutions selected from the six validation procedures. Plots present the horizontal projection of the fitting plane segments as well as the epicenters of their associated events. For the sake of clarity, the clusters obtained for each fit are split into 2 subsets depending on their dip: clusters with dip larger than 50° (left plot) and clusters with dip smaller than 50° (right plot). As the Landers area is

Figure 11. Results of our clustering method applied to the Landers area using the six validation criteria. Results are presented separately for small-dipping faults ($dip < 50^\circ$, left) and large-dipping faults ($dip \geq 50^\circ$, right). We assume the solutions using cross validation and BIC as unrealistic due to the many low-dipping faults in comparison to the tectonically motivated validation measures.

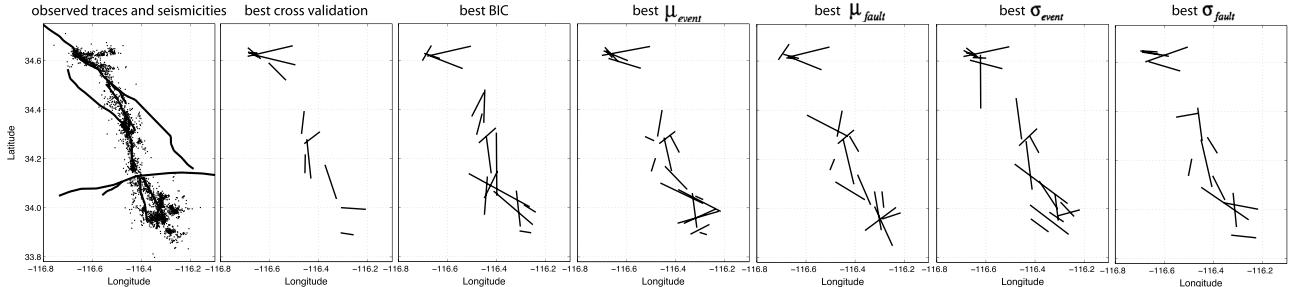


Figure 12. Observed surface traces and seismicity of the Landers area (left plot) and predicted sets of fault traces for each selected reconstructed network.

dominated by strike slip faulting on nearly vertical faults, we think, in the spirit of *Ouillon et al.* [2008], that the large-dip clusters may represent genuine underlying faults, while the low-dip clusters mainly represent spurious structures artificially introduced in order to decrease the local residual of the fit in areas of diffuse seismicity.

[88] Each of the validation techniques yields a different solution. Clearly, there are a large number of events that are clustered on low-dip faults ($dip < 50^\circ$) in the model selected by cross validation. Looking at the properties of each cluster, we notice that there is a clear decrease of μ_F value with increasing dip, suggesting that low-dip planes disagree with their associated focal mechanisms. Thus, the solution selected by cross validation seems not to be realistic. The other validation processes yield solutions that offer a nice agreement in the northern part of the network (which can then be considered as reasonably well inverted); yet, significant differences occur at other locations. If we leave aside the BIC solution for reasons explained in the section dealing with synthetic examples, we are left with four solutions that all agree well with focal mechanisms and among which no definitive and objective choice can be made.

[89] The fact that these validation techniques yield different selected solutions may come from the interplay of two main factors: the multiscale structure of individual faults and the spatial extent of earthquake location uncertainties. Many studies show that faults feature a complex inner structure consisting of a complex subnetwork of subfaults and secondary brittle structures [*Tchalenko*, 1970; *Tchalenko and Ambraseys*, 1970]. If the time span of the catalog is much shorter than the typical timescale necessary to activate rupture on every substructure, then most of the subfaults will feature very few events, precluding their detailed reconstruction. Furthermore, if location uncertainties are larger than the typical spacing of subfaults, the solution to the fit of the full network is not unique either and different validation techniques will favor different solutions.

[90] Following the same approach as *Ouillon et al.* [2008], we also computed the predicted surface traces of the reconstructed faults for each selected model. The idea is to prolong fault planes to the surface and compare them with the observed traces compiled by the CFM (see Figure 12). None of the six predicted trace maps fully agrees with the observed surface fault traces. It may stem from the fact that the catalog we used is only 20 years long, whereas surface fault traces derive from millions of years of tectonic deformation. The active part of this network is thus necessarily a subset of the full network, so that the correspondence between

both sets of fault traces is necessarily imperfect. Surprisingly, *Ouillon et al.* [2008] obtained a solution with a more realistic predicted map of fault traces in the same area.

6. Discussion and Conclusions

6.1. Summary of the Results

[91] In this paper, we introduced a new technique (the ACLUD method) to reconstruct active fault networks which improves on the method of *Ouillon et al.* [2008] as it uses both earthquake locations and their estimated individual uncertainties. After a massive yet nonexhaustive search through the very large solution space, the full set of potential solutions is submitted to six different validation procedures in order to select the corresponding best solutions. Two of the validation steps (cross validation and BIC) process the fit residuals, while the four others look for solutions that provide the best agreement with independently observed focal mechanisms. Tests on synthetic catalogs allowed us to qualify the performance of the fitting method and of the various validation procedures. The method is able to provide exact reconstructions in the case of very simple structures yet is not able to find the input network when structures display more complexity and realistic location uncertainties. However, the solutions provided by each validation step are close to the expected one, especially for the BIC and focal mechanism-based techniques. Adding a uniform spatial background seismicity rate, both validation techniques based on fit residuals fail, while the ones based on focal mechanisms consistency show a much better agreement with the expected solution. The use of a uniform background density is compatible with a total lack of prior assumption about its spatial structure. Moreover, background events are generally spatially isolated as their magnitude is too small to trigger a sufficient number of close aftershocks (which would help in defining a local structure). Those low magnitude events are thus, naturally, prone to larger location uncertainties, which randomize their structure even more. Using more complex distributions, like fractal or multifractal ones, which should also be anisotropic, would require defining more arbitrary parameters. *Ouillon and Sornette* [2011] show the spatial distribution of the selected background events are not clustered at all, thus do not define any fractal structure. They rather define a more or less uniform process which spatial rate slowly changes with location. Thus, if we consider that background events are events which do not cluster on well-defined planes, then we see that they do not seem to obey any fractal distribution.

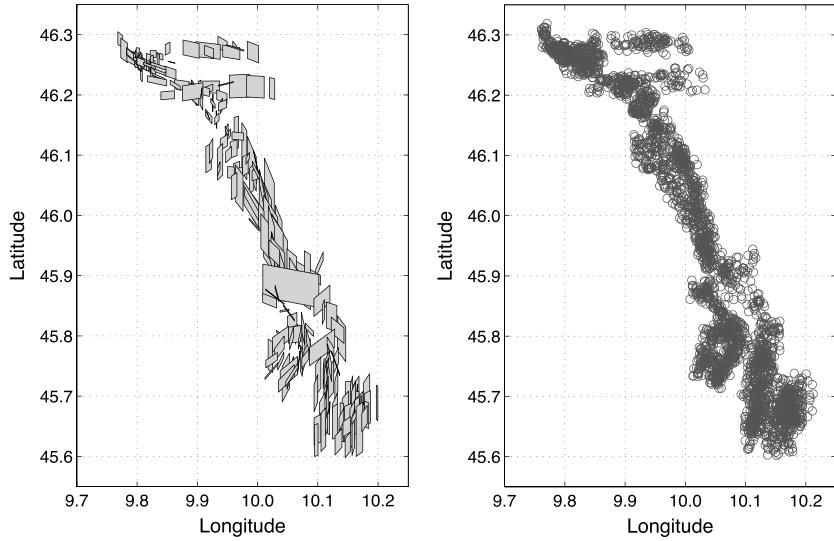


Figure 13. Synthetic multiscale (left) fault network and (right) seismicity, consisting respectively of 220 subfaults and 3153 events.

[92] We compared the results obtained by our new ACLUD technique with the ones obtained on the same data set using the OADC code developed by *Ouillon et al.* [2008]. Despite a slight difference in the nature of one of the validation procedures, we showed that the new method improves significantly on the OADC method, because accounting for individual location uncertainties of events allowed a more detailed fit of faults in areas where such uncertainties were small. It also showed that the results provided by the OADC method also improved when using validation steps based on focal mechanisms consistency. This last observation thus suggests the systematic use of such validation tools, whatever the underlying clustering technique. This also suggests that including focal mechanisms into the clustering scheme itself will provide a more consistent and efficient exploration of the solution space.

[93] The technique has also been applied to a real data set, namely the Landers area. This study confirms that cross validation provides a poor-quality solution, as the network features a significant number of planes with a very low dip, at odds with the prior structural knowledge we have about the nature of faulting in that area. The obtained fault networks also show a poor agreement with focal mechanisms. Comparing the predicted map of fault traces for each of the six selected solutions to the actually observed map did not allow us to draw any conclusion. The reason *Ouillon et al.* [2008] obtained a solution with a more realistic predicted map of fault traces in the same area remains unclear as they did not use the same catalog. The latter may have been of lower quality than ours, which in turn allowed them to fit correctly the gross features of the network. In our case, a better assessment of locations and uncertainties may better reveal the genuine small-scale complexity of the network, which may in turn impact on the quality of the fit, for various reasons that we explain below.

6.2. Undersampled Multiscale Faults

[94] Many field observations suggest that faults feature a complex inner structure [Klinger *et al.*, 2005; Tchalenko and Ambraseys, 1970], consisting of a complex network of subfaults and secondary brittle structures (like Riedel shears

or flower structures, for instance). Some of the substructures may themselves feature a complex inner zone, which thus replicates itself in a more or less self-similar manner. This process necessarily holds down to a lower cutoff scale, which might be of the order of a few rock grain sizes, so that the full fault should ideally be modeled as a closely packed array of a very large number of potentially seismically active subfeatures. This view has been one of the arguments raised by *Ouillon and Sornette* [2011] to justify the use of a Gaussian mixture approach to cluster earthquakes. If we now assume that we can compile a catalog of all events occurring on such a fault, whatever their size and over a very long period of time, with vanishing location uncertainties, then our method would invert correctly the full underlying structure. If the time span of the catalog is much shorter than the typical timescale necessary to activate rupture on every substructure, then the subfaults will be undersampled by the seismicity process, as most of them would feature very few events, if any. In that case, any method will fail to retrieve the correct structure of the fault zone, and our method would only provide a coarse-grained solution, which may not be necessarily unique. If we now add location uncertainties that are larger than the typical spacing of subfaults, and sometimes comparable to the spacing of the macro faults, the coarse-graining problem will be transferred to even larger scales, so that the solution to the fit of the full network will not be unique either: Different validation techniques will provide different preferred solutions.

[95] In order to illustrate this reasoning, we extended the complexity of the synthetic Landers network of section 4 down to smaller scales, using an algorithm inspired from the theory of iterated function systems [Barnsley, 1988; Hutchinson, 1981], a popular technique used to build synthetic fractal sets. In a nutshell, this technique consists in replicating a given fault into another set of randomly rotated and scaled down copies of itself. The set of copies is then used to replace the original fault. The copies are themselves replaced by a similar set of rotated and scaled down copies as well, and so on, down to a given fine-scale resolution. For the sake of simplicity, this segmentation is imposed along the strike of

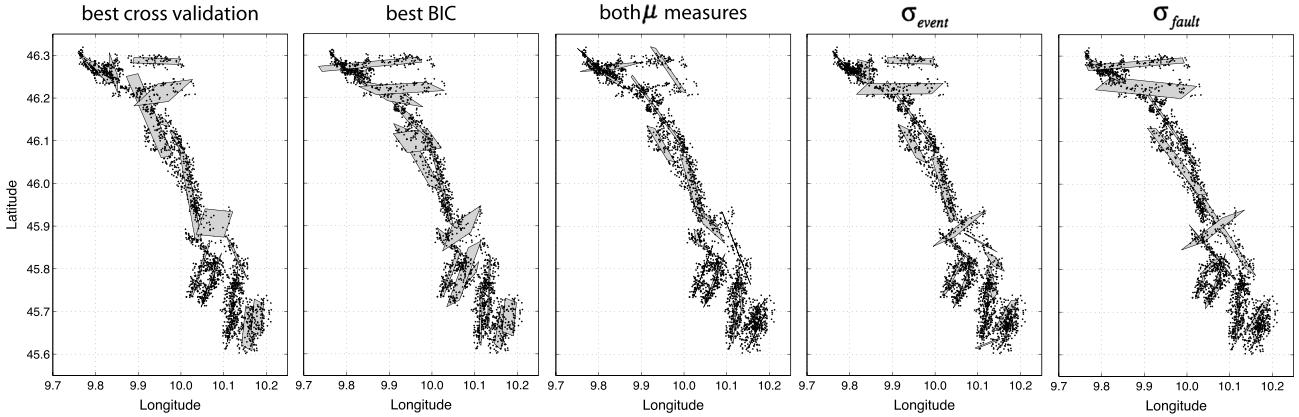


Figure 14. Result of our clustering method applied to the 220 synthetic multiscale fault network consisting of 3103 events. Only 17 to 19 planes were generated.

the fault, each subplane extending to the same depth as the original fault. An example is shown in Figure 13 and features 220 subfaults (instead of the 13 original planes). The small-scale structure appears to be very complex, yet the large-scale structure is similar to the one presented in Figure 3. We then distribute the same set of 3103 events over this new set of subfaults. The network has been generated so that there is, on average, between 10 and 20 events on each segment, but some subfaults may feature only one or two of them. (Details are given in the supporting information.)

[96] Using the same method as in section 4, we generate a new catalog of events providing both their expected locations and their uncertainties. Focal mechanisms are first chosen as fully compatible with the orientation of the subfault to which the event is attributed, before we add a 10° uncertainty on strike, dip, and rake. This catalog is then processed by our nonlinear fitting method, using 6000 runs. This smaller number of runs is a consequence of the much larger duration of individual inversions due to the larger complexity of the data set, which necessitates a longer time to explore the space of models.

[97] Figure 14 shows the solutions selected by the six validation methods. None of them is able to reconstruct the full set of 220 planes, as expected. All proposed networks feature only 17 to 19 faults, as undersampled subfaults are indeed merged into simpler structures in order to cluster a sufficient number of events (at least four, as we imposed). None of the solutions are identical, reflecting the nonuniqueness of the solution provided by the different criteria.

6.3. Overfitting, Underfitting, and Validation Techniques

[98] The two validation tools based on residuals, i.e., cross validation and BIC, were used in order to avoid problems of overfitting. However, we showed in the previous section that we primarily face a problem of underfitting. This observation necessarily questions the use of such validation strategies for clustering techniques. We also showed that both cross validation and BIC were unable to select the correct solution when a set of background events is superimposed over the more correlated set of earthquakes. This thus leads us to conclude that the use of such criteria is certainly much less adapted to the selection of the correct solution than the use of focal mechanisms, which bring their share of information about the dynamics of the network. Up to now, we only use part

of the information contained within focal mechanisms, as we only checked the consistency of the orientation of one of the nodal planes and of the fitting planes. We thus deliberately forgot the rake. In the future, this observation should be included as well in order to better constrain solutions, thus providing a coherent set of slip vectors within the same fault.

6.4. Future Developments

[99] Our unsupervised clustering technique uses only the spatial information contained within seismicity catalogs. We showed that the model validation criteria derived from focal mechanisms are in better agreement with the true model when dealing with synthetics. A natural idea is then to include more prior seismic information into the clustering procedure itself, like waveform correlation coefficients, focal mechanism similarities, and so on. However, the design of a cost function able to take account of all those different data necessitates defining a proper weighting strategy. We rather suggest using this extra knowledge to make decisions at decisive steps of the clustering process.

[100] Despite the fact that earthquake catalogs depict events as point processes, those events indeed define a collection of stress tensors (and their time histories during the rupture process), distributed over a set of finite planar, subplanar, or fractal structures. Earthquakes define stress and strain singularities, which obviously interact through stress transmission: Earthquakes are triggered by the accumulation of stress at plates' boundaries as well as by stress fluctuations induced by previous events. Earthquakes are also increments of deformation that reveal the development and growth of faults. In return, earthquakes are constrained to occur on such faults. The geometry of the set of events is thus governed both by the applied boundary conditions and the mechanical interactions between events. The overall orientation of faults is mainly governed by the principal directions of the applied boundary stress tensor, while the inner structure and complexity of faults are mainly dominated by interactions between events.

[101] These interactions may propagate over very large distances and timescales, through cascades of domino effects. Indeed, faults are complex geological structures that are often considered as self-affine surfaces or self-similar aggregates of smaller-scale planar features. This means that such objects are significantly correlated over a substantial range of spatial scales. The basic idea we have in mind is that such

a correlation must also translate into the dynamical signature of faulting, i.e., the dynamics of the associated earthquakes. Here, we do not use the term “dynamic” as associated to the temporal distribution of individual events (that is also given in earthquake catalogs) but to the rupture process of individual events. The idea is that if two events occur within a short spatial distance and belong to the same fault, then there is a “large” probability that their rupture processes will be similar (which is the basic meaning of correlation). This similarity should, on average, decrease with the distance between events. As all the information we have about the dynamics of faulting is contained within the recorded waveforms, it is thus reasonable to assume that events belonging to the same fault segment will radiate, on average, similar waveforms. Indeed, this similarity is observed and exploited for source model inversion and strong ground motion modeling in using small events as empirical Green’s functions [e.g., Woessner *et al.*, 2002].

[102] The most critical and arbitrary step of the clustering algorithm is the one where the locally worst cluster is split into two subclusters in order to improve the fit. The chosen cluster is the one with the largest thickness (so that it relies on arguments based on local fit residuals), and the split process is purely random. We suggest that, for a given number of clusters, we may first assess the μ_F and σ_F values for all individual faults. The one(s) with the largest values may then be the chosen ones to be split when increasing the number of planes, so that the splitting is now based on more mechanical grounds. We may also separate those clusters from the rest of the catalog, fit them separately, and put them back into the whole data set. This would allow fitting separately less complex structures within smaller solution spaces, converging more quickly to a reliable solution. The randomness of the splitting may also be questioned, as we know that the standard k -means algorithm is very sensitive to initial conditions (i.e., the locations of the initial seeds) and that some of them are more optimal than others. In our case, the location, size, and orientation of the new planes generated by splitting certainly have a large impact on the reliability of the final solution. Recently, both the k -means++ [Arthur and Vassilvitskii, 2007] and the k -means \parallel [Bahmani *et al.*, 2012] have been proposed in order to provide better initial conditions to k -means. In k -means++, the first seed is chosen randomly among the data points. All the other seeds are then chosen sequentially from the remaining data points with a probability proportional to their distance squared to the closest previous seed. The k -means \parallel is an improvement of k -means++ to deal with large data sets. This technique thus allows one to generate a more or less uniform set of seeds.

[103] The most important obstacle to such clustering techniques is certainly the size of the catalogs to be processed. Up to now, we only considered sets of a few thousands of events, but the full California catalog for instance features up to half a million data points. Processing such large data sets is clearly out of reach of our current algorithm. We may improve it by parallelizing some steps (such as the computations of distances) and also by choosing more efficiently the initial conditions (as outlined above with the k -means++ approach). This limitation to process very large catalogs also holds for other clustering techniques, such as the Gaussian mixture expectation-maximization (EM) approach of Oullon and Sornette [2011]. In the latter paper, a catalog is approximated

as a superposition of Gaussian kernels, whose optimal number is determined through a cross-validation strategy. A set of 4000 events occurring in the Mount Lewis area necessitated about 100 Gaussian kernels for fitting. This large number of objects to fit the data is explained by the fact that the fitting procedure is very sensitive to density fluctuations along a given fault—whereas this is not in the case when fitting with planes. Such a fault, fitted by one single plane following our approach, may require several kernels in the EM approach, which increases the necessary computational resources. We would thus rather use our k -means-based approach to first fit the main faults, then switch to an EM approach to infer more precisely the structure of the fault zones, in the spirit of Oullon and Sornette [2011], who were able to provide a typical segmentation scale—an information of prime importance to model high-frequency ground shaking.

[104] The proposed clustering algorithm accounts for the full location pdf of each event and the full covariance matrix of the location errors, implemented as the stopping criterion. Four solution selection criteria based on focal mechanisms are used to find the optimal solutions. It implies that (1) location catalogs should contain detailed location error information and (2) focal mechanisms are known in great number, and with sufficient accuracy. The prerequisites currently limit the scope of this method to only a few suitable regions. Indeed, our method calls for increased efforts to generate catalogs that provide a more detailed description of the uncertainties as quantitative pattern recognition methods can now handle them—this is nothing else than is required by users of earthquake catalogs, ranging from earthquake forecasting to seismic hazard assessment. We used, for example, NonLinLoc [Lomax *et al.*, 2000] to perform relocations as this method is able to provide the full description of possible locations, and it is an accepted method. This can be implemented for any network and may become a standard method in the future. Similarly, there are more and more methods to better describe the uncertainties of focal mechanisms. Yang *et al.*’s [2012] is only one example using the HASH method from Hardebeck and Shearer [2002]. Another example for better constraining focal mechanism uncertainties is using a Bayesian approach by Arnold and Townend [2007] and Lund and Townend [2007]. We are convinced that this is not a limit of the method but rather defines quality requirements for future studies.

[105] Both OADC and ACLUD assume that faults can be modeled as perfectly planar objects with a vanishingly small thickness, so that each event can be attributed with certainty (i.e., unit probability) to the closest plane. It implies that location uncertainties alone explain the fit discrepancies. Instead of using planes, Oullon and Sornette [2011] used an expectation-maximization scheme featuring Gaussian kernels to fit the data. Generalizing to fractal (or self-affine) fault models would necessitate to extend the method to stable laws (such as Cauchy or Lévy laws), which display power law tails. Yet, no simple implementation of such kernels exists, even in 1-D.

[106] The proposed clustering approach retains the potential to improve the spatial forecasting skills of current forecast models, especially those that attempt short-term near real-time forecasts and are prone to be used for operational earthquake forecasting. Forecast models such as the Short-Term Earthquake Probability model [Gerstenberger *et al.*,

2005; Woessner et al., 2010] or the class of epidemic-type earthquake forecast models [Helmstetter et al., 2006; Ogata and Zhuang, 2006] have been shown to be mostly limited in their spatial predictive skill [Woessner et al., 2011]. Thus, we expect that including the proposed method will improve the forecast skills at least during strong aftershock sequences and may help to improve current efforts to provide meaningful operation earthquake forecasting [Jordan et al., 2011].

[107] **Acknowledgments.** We would like to thank Egill Hauksson for providing us the Landers catalog and Stefan Wiemer for his valuable comments. This work was funded by the Swiss Science Foundation under grant 200021-117829 and the European Commission through funds provided within the FP7 project GEISER, grant agreement 241321-2.

References

- Arnold, R., and J. Townend (2007), A Bayesian approach to estimating tectonic stress from seismological data, *Geophys. J. Int.*, 170(3), 1336–1356, doi:10.1111/j.1365-246x.2007.03485.x.
- Arthur, D., and S. Vassilvitskii (2007), k-means plus plus: The advantages of careful seeding, Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 1027–1035.
- Bahmani, B., B. Moseley, A. Vattani, R. Kumar, and S. Vassilvitskii (2012), Scalable k-means++, *Proc. VLDB Endow.*, 5(7), 622–633.
- Barnsley, M. (1988), *Fractals Everywhere*, Academic Press, New York.
- Bishop, C. M. (2006), *Pattern Recognition and Machine Learning*, Springer, New York.
- Bondár, I., S. C. Myers, E. R. Engdahl, and E. A. Bergman (2004), Epicentre accuracy based on seismic network criteria, *Geophys. J. Int.*, 156(3), 483–496, doi:10.1111/j.1365-246x.2004.02070.x.
- Chau, M., R. Cheng, B. Kao, and J. Ng (2006), Uncertain data mining: An example in clustering location data, *Lect. Notes Artif. Int.*, 3918, 199–204.
- Courjault-Radé, P., J. Darrozes, and P. Gaillot (2009), The M = 5.1 1980 Arudy earthquake sequence (western Pyrenees, France): A revisited multi-scale integrated seismologic, geomorphologic and tectonic investigation, *Int. J. Earth Sci. (Geol Rundsch.)*, 98(7), 1705–1719, doi:10.1007/s00531-008-0320-5.
- Cowie, P. A., C. Vanneste, and D. Sornette (1993), Statistical physics model for the spatiotemporal evolution of faults, *J. Geophys. Res.*, 98(B12), 21,809–21,821, doi:10.1029/93jb02223.
- Cowie, P. A., D. Sornette, and C. Vanneste (1995), Multifractal scaling properties of a growing fault population, *Geophys. J. Int.*, 122(2), 457–469, doi:10.1111/j.1365-246x.1995.tb07007.x.
- Duda, R. O., P. E. Hart, and D. G. Stork (2001), *Pattern Classification*, Wiley, New York.
- Field, E. H., K. R. Milner, and the 2007 Working Group on California Earthquake Probabilities (2008), Forecasting California's earthquakes—What can we expect in the next 30 years?, *U.S. Geological Survey*.
- Frankel, A. D., D. L. Carver, and R. A. Williams (2002), Nonlinear and linear site response and basin effects in Seattle for the M 6.8 Nisqually, Washington, earthquake, *Bull. Seismol. Soc. Am.*, 92(6), 2090–2109, doi:10.1785/0120010254.
- Gabrielov, A., V. KeilisBorok, and D. D. Jackson (1996), Geometric incompatibility in a fault system, *Proc. Natl. Acad. Sci. U. S. A.*, 93(9), 3838–3842, doi:10.1073/pnas.93.9.3838.
- Gaillot, P., J. Darrozes, P. Courjault-Rade, and D. Amorese (2002), Structural analysis of hypocentral distribution of an earthquake sequence using anisotropic wavelets: Method and application, *J. Geophys. Res.*, 107(B10), 2218, doi:10.1029/2001jb000212.
- Gerstenberger, M. C., S. Wiemer, L. M. Jones, and P. A. Reasenberg (2005), Real-time forecasts of tomorrow's earthquakes in California, *Nature*, 435(7040), 328–331, doi:10.1038/Nature03622.
- Hardebeck, J. L., and P. M. Shearer (2002), A new method for determining first-motion focal mechanisms, *Bull. Seismol. Soc. Am.*, 92(6), 2264–2276, doi:10.1785/0120010200.
- Hauksson, E. (2010), Spatial separation of large earthquakes, aftershocks, and background seismicity: Analysis of interseismic and coseismic seismicity patterns in Southern California, *Pure Appl. Geophys.*, 167(8–9), 979–997, doi:10.1007/S00024-010-0083-3.
- Hauksson, E., W. Z. Yang, and P. M. Shearer (2012), Waveform relocated earthquake catalog for Southern California (1981 to June 2011), *Bull. Seismol. Soc. Am.*, 102(5), 2239–2244, doi:10.1785/0120120010.
- Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2006), Comparison of short-term and time-independent earthquake forecast models for Southern California, *Bull. Seismol. Soc. Am.*, 96(1), 90–106, doi:10.1785/0120050067.
- Hiemer, S., D. D. Jackson, Q. Wang, Y. Y. Kagan, J. Woessner, J. D. Zechar, and S. Wiemer (2013), A stochastic forecast of California earthquakes based on fault slip and smoothed seismicity, *Bull. Seismol. Soc. Am.*, 103(2A), 799–810, doi:10.1785/0120120168.
- Husen, S., and J. L. Hardebeck (2010), Earthquake location accuracy, *Community Online Resource for Statistical Seismicity Analysis*, doi:10.5078/corsa-55815573.
- Husen, S., E. Kissling, N. Deichmann, S. Wiemer, D. Giardini, and M. Baer (2003), Probabilistic earthquake location in complex three-dimensional velocity models: Application to Switzerland, *J. Geophys. Res.*, 108(B2), 2077, doi:10.1029/2002jb001778.
- Husen, S., C. Bachmann, and D. Giardini (2007), Locally triggered seismicity in the central Swiss Alps following the large rainfall event of August 2005, *Geophys. J. Int.*, 171(3), 1126–1134, doi:10.1111/j.1365-246x.2007.03561.x.
- Hutchinson, J. E. (1981), Fractals and self similarity, *Indiana U Math J.*, 30(5), 713–747, doi:10.1512/iumj.1981.30.30055.
- Jordan, T. H., Y. T. Chen, P. Gasparini, R. Madariaga, I. Main, W. Marzocchi, G. Papadopoulos, G. Sobolev, K. Yamaoka, and J. Zschau (2011), Operational earthquake forecasting: State of knowledge and guidelines for utilization, *Ann. Geophys.-Italy*, 54(4), 315–391, doi:10.4401/ag-5350.
- Kagan, Y. Y. (2005), Double-couple earthquake focal mechanism: Random rotation and display, *Geophys. J. Int.*, 163(3), 1065–1072, doi:10.1111/j.1365-246x.2005.02781.x.
- Klinger, Y., X. W. Xu, P. Tapponnier, J. Van der Woerd, C. Lasserre, and G. King (2005), High-resolution satellite imagery mapping of the surface rupture and slip distribution of the M-W similar to 7.8, 14 November 2001 Kokoxili earthquake, Kunlun Fault, northern Tibet, China, *Bull. Seismol. Soc. Am.*, 95(5), 1970–1987, doi:10.1785/0120040233.
- Lomax, A., J. Virieux, P. Volant, and C. Berge (2000), Probabilistic earthquake location in 3D and layered models: Introduction of a Metropolis-Gibbs method and comparison with linear locations, in *Adv. Seismic Event Location*, edited by C. H. Thurber and N. Rabinowitz, pp. 101–134, Kluwer, Amsterdam, The Netherlands.
- Lomax, A., A. Michelini, and A. Curtis (2009), Earthquake location, direct, global-search methods, in *Encyclopedia of Complexity and System Science*, edited by R. A. Meyers, pp. 2249–2473, Springer, New York, doi:10.1007/978-0-387-30440-3.
- Lund, B., and J. Townend (2007), Calculating horizontal stress orientations with full or partial knowledge of the tectonic stress tensor, *Geophys. J. Int.*, 170(3), 1328–1335, doi:10.1111/j.1365-246x.2007.03468.x.
- Mace, C. G., and K. M. Keranen (2012), Oblique fault systems crossing the Seattle Basin: Geophysical evidence for additional shallow fault systems in the central Puget Lowland, *J. Geophys. Res.*, 117, B03105, doi:10.1029/2011jb008722.
- MacQueen, J. B. (1967), Some methods for classification and analysis of multivariate observations, paper presented at Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press.
- Moser, T. J., T. Vanek, and G. Nolet (1992), Hypocenter determination in strongly heterogeneous Earth models using the shortest-path method, *J. Geophys. Res.*, 97(B5), 6563–6572, doi:10.1029/91JB03176.
- Ogata, Y., and H. C. Zhuang (2006), Space-time ETAS models and an improved extension, *Tectonophysics*, 413(1–2), 13–23, doi:10.1016/j.tecto.2005.10.016.
- Ouillon, G., and D. Sornette (2011), Segmentation of fault networks determined from spatial clustering of earthquakes, *J. Geophys. Res.*, 116, B02306, doi:10.1029/2010jb007752.
- Ouillon, G., D. Sornette, and C. Castaing (1995), Organisation of joints and faults from 1-cm to 100-km scales revealed by optimized anisotropic wavelet coefficient method and multifractal analysis, *Nonlinear Proc. Geophys.*, 2(3–4), 158–177.
- Ouillon, G., C. Castaing, and D. Sornette (1996), Hierarchical geometry of faulting, *J. Geophys. Res.*, 101(B3), 5477–5487, doi:10.1029/95jb02242.
- 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, B01306, doi:10.1029/2007jb005032.
- Petersen, M. D., T. Q. Cao, K. W. Campbell, and A. D. Frankel (2007), Time-independent and time-dependent seismic hazard assessment for the State of California: Uniform California earthquake rupture forecast model 1.0, *Seismol. Res. Lett.*, 78(1), 99–109, doi:10.1785/Gssrl.78.1.99.
- Plesch, A., et al. (2007), Community fault model (CFM) for Southern California, *Bull. Seismol. Soc. Am.*, 97(6), 1793–1802, doi:10.1785/0120050211.
- Powers, P. M., and T. H. Jordan (2010), Distribution of seismicity across strike-slip faults in California, *J. Geophys. Res.*, 115, B05305, doi:10.1029/2008jb006234.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (2007), *Numerical Recipes: The Art of Scientific Computing*, 3rd ed., Cambridge University Press, New York.
- Rhoades, D. A., and M. W. Stirling (2012), An earthquake likelihood model based on proximity to mapped faults and catalogued earthquakes, *Bull. Seismol. Soc. Am.*, 102(4), 1593–1599, doi:10.1785/0120110326.

- Schwarz, G. (1978), Estimating dimension of a model, *Ann. Stat.*, 6(2), 461–464, doi:10.1214/Aos/1176344136.
- Sornette, D. (1991), Self-organized criticality in plate-tectonics, *Nato Adv. Sci. I C-Mat.*, 349, 57–106.
- Sornette, D., and J. Virieux (1992), Linking short-timescale deformation to long-timescale tectonics, *Nature*, 357(6377), 401–404, doi:10.1038/357401a0.
- Sornette, D., P. Miltenberger, and C. Vanneste (1994), Statistical physics of fault patterns self-organized by repeated earthquakes, *Pure Appl. Geophys.*, 142(3–4), 491–527, doi:10.1007/Bf00876052.
- Stein, R. S. (1999), The role of stress transfer in earthquake occurrence, *Nature*, 402(6762), 605–609, doi:10.1038/45144.
- Tarantola, A., and B. Valette (1982), Inverse problems = Quest for information, *J. Geophys.*, 50, 159–170.
- Tchalenko, J. S. (1970), Similarities between shear zones of different magnitudes, *Geol. Soc. Am. Bull.*, 81(6), 1625–1640, doi:10.1130/0016-7606(1970)81[1625:Sbszod]2.0.Co;2.
- Tchalenko, J. S., and N. N. Ambraseys (1970), Structural analysis of Dasht-e-Bayaz (Iran) earthquake fractures, *Geol. Soc. Am. Bull.*, 81(1), 41–60, doi:10.1130/0016-7606(1970)81[41:Saotdb]2.0.Co;2.
- Basili, R., et al. (2013), The European database of seismogenic faults (EDSF) compiled in the framework of the Project SHARE, <http://diss.rm.ingv.it/share-edsf/>, doi:10.6092/INGV.IT-SHARE-EDSF.
- Waldhauser, F., and D. P. Schaff (2008), Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods, *J. Geophys. Res.*, 113, B08311, doi:10.1029/2007jb005479.
- Werner, M. J., K. Ide, and D. Sornette (2011), Earthquake forecasting based on data assimilation: Sequential Monte Carlo methods for renewal point processes, *Nonlinear Proc. Geoph.*, 18(1), 49–70, doi:10.5194/Npg-18-49-2011.
- Wesson, R. L., W. H. Bakun, and D. M. Perkins (2003), Association of earthquakes and faults in the San Francisco Bay area using Bayesian inference, *Bull. Seismol. Soc. Am.*, 93(3), 1306–1332, doi:10.1785/0120020085.
- Wittlinger, G., G. Herquel, and T. Nakache (1993), Earthquake location in strongly heterogeneous media, *Geophys. J. Int.*, 115(3), 759–777, doi:10.1111/j.1365-246X.1993.tb01491.x.
- Woessner, J., M. Treml, and F. Wenzel (2002), Simulation of M-W = 6.0 earthquakes in the Upper Rhinegraben using empirical Green functions, *Geophys. J. Int.*, 151(2), 487–500, doi:10.1046/J.1365-246x.2002.01785.X.
- Woessner, J., D. Schorlemmer, S. Wiemer, and P. M. Mai (2006), Spatial correlation of aftershock locations and on-fault main shock properties, *J. Geophys. Res.*, 111, B08301, doi:10.1029/2005jb003961.
- Woessner, J., A. Christoffersen, J. D. Zechar, and D. Monelli (2010), Building self-consistent, short-term earthquake probability (STEP) models: Improved strategies and calibration procedures, *Ann. Geophys.-Italy*, 53(3), 141–154, doi:10.4401/Ag-4812.
- Woessner, J., S. Hainzl, W. Marzocchi, M. J. Werner, A. M. Lombardi, F. Catalli, B. Enescu, M. Cocco, M. C. Gerstenberger, and S. Wiemer (2011), A retrospective comparative forecast test on the 1992 Landers sequence, *J. Geophys. Res.*, 116, B05305, doi:10.1029/2010jb007846.
- Yang, W. Z., E. Hauksson, and P. M. Shearer (2012), Computing a large refined catalog of focal mechanisms for Southern California (1981–2010): Temporal stability of the style of faulting, *Bull. Seismol. Soc. Am.*, 102(3), 1179–1194, doi:10.1785/0120110311.
- Zechar, J. D., and T. H. Jordan (2008), Testing alarm-based earthquake predictions, *Geophys. J. Int.*, 172(2), 715–724, doi:10.1111/J.1365-246x.2007.03676.X.
- Zechar, J. D., and T. H. Jordan (2010), Simple smoothed seismicity earthquake forecasts for Italy, *Ann. Geophys.-Italy*, 53(3), 99–105, doi:10.4401/Ag-4845.