Automatic reconstruction of fault networks from seismicity catalogs: Three-dimensional optimal anisotropic dynamic clustering

G. Ouillon,[1] C. Ducorbier,[2] and D. Sornette[3,4]

Received 7 March 2007; revised 5 September 2007; accepted 17 October 2007; published 17 January 2008.

[1] We propose a new pattern recognition method that is able to reconstruct the three-dimensional structure of the active part of a fault network using the spatial location of earthquakes. The method is a generalization of the so-called dynamic clustering (or k means) method, that partitions a set of data points into clusters, using a global minimization criterion of the variance of the hypocenters locations about their center of mass. The new method improves on the original k means method by taking into account the full spatial covariance tensor of each cluster in order to partition the data set into fault-like, anisotropic clusters. Given a catalog of seismic events, the output is the optimal set of plane segments that fits the spatial structure of the data. Each plane segment is fully characterized by its location, size, and orientation. The main tunable parameter is the accuracy of the earthquake locations, which fixes the resolution, i.e., the residual variance of the fit. The resolution determines the number of fault segments needed to describe the earthquake catalog: the better the resolution, the finer the structure of the reconstructed fault segments. The algorithm successfully reconstructs the fault segments of synthetic earthquake catalogs. Applied to the real catalog constituted of a subset of the aftershock sequence of the 28 June 1992 Landers earthquake in southern California, the reconstructed plane segments fully agree with faults already known on geological maps or with blind faults that appear quite obvious in longer-term catalogs. Future improvements of the method are discussed, as well as its potential use in the multiscale study of the inner structure of fault zones.

Citation: 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.

1. Introduction

[2] In the last 20 years, a large amount of knowledge has been assembled on the phenomenology and physics of faulting and seismicity, but a clear physical and mechanical understanding of the links and interactions between and among them is still missing. Such knowledge may prove decisive to improve our understanding of seismic risk, especially the time-dependent risk associated with future large and potentially destructive seismic events.

[3] The basis of our paper is that seismic hazard assessment faces two fundamental bottlenecks: (1) catalogs are inherently incomplete with respect to the typical recurrence time of large earthquakes which are many times larger than the duration of the available catalogs and (2) earthquakes occur on faults, but most of them are still unknown, limiting our understanding and our assessment of seismic risks. The principal issue therefore lies in the association between earthquakes and faults, a necessary step to address the likely location of potential large events and the mechanical behavior of a fault network at timescales up to a few decades.

[4] Indeed, most of the information available on the structure of fault patterns comes from surface mapping at various scales. One can however question the reliability of such data sets, as earthquake catalogs generally include large quantities of events which do not seem to be linked to any such faults, but rather seem to be associated with blind faults. Comparing two-dimensional (2-D) map views of the long-term cumulative brittle deformation to the 3-D structure of the short-term incremental deformation is bound to provide limited insight at best. Failing to link appropriately the long-term cumulative brittle deformation to the short-term incremental deformation may lead to biased, incorrect models of the multiscale tectonic processes as a whole.

[5] The 3-D geometry of an active fault zone is often constrained by mapping the surface trace; dip angle at depth and depth extent are either constrained by results of controlled source seismology (if available) and the distribution

[1] Lithophyse, Nice, France.
[2] Géosciences Azur, UMR 6526 CNRS-UNSA-IRD, Mecatec Unit, Valbonne, France.
[3] Department of Management, Technology and Economics, ETH Zurich, Zürich, Switzerland.
[4] Department of Earth and Space Sciences and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California, USA.

0148-0227/08/2007JB005032$09.00

of hypocenter locations (and their associated focal mechanisms), or they are simply extrapolated using geometric constraints, if seismologic constraints are not available. For example, one of the most sophisticated fault models available, the Community Fault Model (CFM) of the Southern California Earthquake Center (SCEC), combines available information on surface traces, seismicity, seismic reflection profiles, borehole data, and other subsurface imaging techniques to provide three-dimensional representations of major strike-slip, blind-thrust, and oblique-reverse faults of southern California [*Plesch and Shaw*, 2002]. Each fault is represented by a triangulated surface in a precise geographic reference frame. However, the representation of a fault by a simple surface cannot reflect the fine-detailed structure seen in extinct fault zones and in drilling experiments of active faults [e.g., *Scholz*, 2002; *Faulkner et al.*, 2003]. These results suggest that fault zones actually consist of narrow earthquake-generating cores, possibly accompanied by small subsidiary faults. Detailed structural and seismicity analyses also reveal that many events still cannot be attributed to any known brittle structure [see, e.g., *Guzofski et al.*, 2007]. Our present contribution is to propose a general method to identify and locate active faults in seismically active regions by using a rigorous approach based on seismicity. Our goal is to gain a better understanding of the link between fault structures and earthquakes.

[6] For this, the present paper presents a new pattern recognition method which determines an optimal set of plane segments, hereafter labeled as "faults," that best fit a given set of earthquake hypocenters. In a nutshell, our simple planar fault recognition technique uses a principal component analysis of earthquake hypocenters dynamically divided into distinct clusters, whereby the number of potential faults is iteratively increased until the least important principal component eigenvalue of every cluster drops below the earthquake location error: this then defines distinctly planar features labeled faults.

[7] In the first part of this paper, we provide a short overview of pattern recognition methods used to detect linear or planar features in images, such as the Hough and wavelet transforms. We then present in more detail the k means method, which provides a partition of a set of data points into clusters, before detailing our new method, a generalization of k means to the case of anisotropic structures. We illustrate this new technique through benchmark tests on a synthetic data set, before providing results obtained on a subset of the Landers aftershock sequence. The final discussion of our results sketches the next set of improvements for this method which will be developed in following papers and outlines its potential use in the multiscale study of fault zone structures.

2. Line and Plane Detection in Image Analysis

[8] The detection of linear and planar structures in seismotectonics has a long history, but still suffers from a lack of quantitative methods. From the early years of instrumental seismology, the main instrument to identify faults from earthquakes has indeed been the eye. It is by visual inspection that tectonic plates boundaries or the Benioff plane in subduction zones have been delineated. Visual inspection remains the main approach to identify blind faults at smaller scales. As location errors can sometimes blur the structures one is looking for, it is common to remove events with too large uncertainties in their location. Another method, proposed by *Jones and Stewart* [1997], consists in moving events within their ellipsoids of confidence to collapse the data onto a simpler structure. This method provides spectacular results on real data, but may also create spurious patterns as a result of the collapse of random sets onto artificial organized structures [*Nicholson et al.*, 2000].

[9] Very few efforts have been devoted to the development of automatic digital detection of linear or planar spatial features from earthquake catalogs. We shall now describe two major approaches in image analysis, and outline their advantages and drawbacks, before turning to dynamic clustering and our proposed dedicated extension.

2.1. Hough Transform

[10] The Hough transform [see *Duda and Hart*, 1972] is a technique that is widely used in digital image analysis, with applications, for example, in the detection and characterization of edges, or the reconstruction of trajectories in particle physics experiments. The Hough transform is conceptually very close to the 2-D radon transform. The original Hough transform identifies linear features within an image, but it can be extended to other shapes (like circles or ellipses). We shall anyway present it for straight line detection in two dimensions for the sake of simplicity. The basic idea is that, given a set of data points, there is an infinite number of lines that can pass through each point. The aim of the Hough transform is to determine the set of lines that get through several points.

[11] Each line can be parameterized using two parameters, r and θ, in the Hough space. The parameter r is the distance between this line and the origin, and θ is the orientation of the normal to that line (and is measured from the abscissa axis). Then, the infinite number of lines passing through a given point defines a curve $r(\theta)$ in the Hough plane. All lines passing through a point located at (x_0, y_0) obey the equation $r(\theta) = x_0 \cos(\theta) + y_0 \sin(\theta)$. If we now consider several points of our data set that are located on the same straight line in direct space, their associated $r(\theta)$ curves will all cross at the same (r, θ) point in the Hough space. The corresponding line is then fully identified. Note that several distinct sets of $r(\theta)$ curves can cross in the Hough space at different (r, θ) locations, so that as many linear features can be extracted at once. This idea can be extended to arbitrary 2-D structures (such as circles or ellipses) or to 3-D spaces and planar features, provided that one extends the dimension of the Hough space correspondingly. For example, this dimension is equal to 4 when one wants to detect straight lines in a 3-D real space, so that the method is not used in practice in that case (or rather in a much simplified form, provided there is only a very small number of such lines to detect [see *Bhattacharya et al.*, 2000]). The dimension of the Hough space drops to 3 when dealing with planar features, which is still quite difficult to handle computationally.

[12] The main problem of this technique is that it does not provide directly the spatial extension of the linear or planar features, only their positions in space, as each line (or plane) has an infinite size. A postprocessing step must then be

performed in order to compute the finite extension, by computing the largest distance between pairs of points belonging to the same structure, for example. Another drawback is that the Hough transform does not take account of the uncertainties in the location of the set of data points, a crucial parameter in seismology. All in all, it appears that the Hough transform is very efficient only when dealing with a small number of clean ordered patterns (see however some specific strategies that can be defined to deal with real fracture planes [*Sarti and Tubaro*, 2002]). Another major argument against the use of the Hough transform for fault segment reconstruction from seismic catalogs is that there is no way to include any other information one may have about seismic events, such as their seismic moment tensors, focal mechanisms or magnitudes.

2.2. Optimal Anisotropic Wavelet Coefficient Method

[13] The characterization of anisotropy is a basic task in structural geology, tectonics and seismology. The orientations of structures stem from mechanical boundary conditions, which are important to retrieve. They control the future evolution of the system, which we are interested to predict. Tools for quantifying anisotropy generally materialize in rose diagrams (for 2-D fracture maps for instance) or in stereographical projections (when data are sampled in 3-D space). The main drawback of these representations is that they are scale-dependent. For example, as shown by *Ouillon* [1995] and *Ouillon et al.* [1995], in the case of en echelon fractures, small-scale and large-scale features will possess fundamentally different anisotropy properties. In that case, mapping at small scales will not give any clue on the large-scale properties of the system. The so-called Optimal Anisotropic Wavelet Coefficient (OAWC) method, presented in the above publications, was designed to specifically address this problem.

[14] For 2-D signals, such as fault maps and fracture maps, a wavelet is a band-pass filter which is characterized, in real space, by a width (called resolution), a length and an azimuth. In a nutshell, the OAWC method consists in fixing the resolution and convolving the wavelet with the map of fault traces, varying its length and azimuth. For any given position, when the result of the convolution (which is the wavelet coefficient) reaches a maximum, one stores the associated azimuth of the wavelet and switches to the next spatial location. Building a histogram of such optimal azimuths provides a rose diagram at the considered resolution. Performing this process at different resolutions allows one to describe the evolution of anisotropy with scale from a single initial data set. Looking at various maps of fractures and faults patterns in Saudi Arabia, *Ouillon* [1995] and *Ouillon et al.* [1995] were able to show that anisotropy properties change drastically at resolutions that can be related to the thicknesses of the mechanical layers involved in the brittle deformation process (from sandstone beds up to the continental crust).

[15] The OAWC method has been extended to characterize the anisotropy of mineralization in thin plates, as well as to detect faults or lineaments from earthquake epicenter maps [*Darrozes et al.*, 1998; *Gaillot et al.*, 1997, 1999, 2002; *Grégoire et al.*, 1998]. In this second class of applications, the issue concerning the finite accuracy of event locations is addressed by considering a wavelet resolution larger than the spatial uncertainties. However, this method does not provide direct information about the size of the structures and indeed does not manipulate structures as such. Each point in space is characterized by anisotropy properties, but the method does not naturally construct clusters from neighboring points with similar characteristics. Moreover, its extension to 3-D signals and their associated 3-D patterns would prove unreasonably time consuming (as space has to be discretized at a scale smaller than the chosen resolution) and would suffer from edge effects near the top and bottom of the seismogenic zone. Note that some specific wavelets can be defined to account for such edge effects [*Cohen et al.*, 1993], but the known algorithms in this case deal with orthonormal bases of wavelets, restricting the choice of the analyzing scales to a discrete set obeying a geometric series of common ratio 2 (whereas the sizes of tectonic structures have no reason to be restricted to such a set). Moreover, the generalization of this technique (coined wavelets on the interval) to 3-D anisotropic wavelets still remains to be done. The current 2-D version of the OAWC method could be applied in areas where only vertical to subvertical faulting occurs, but we would miss the segmentation of such faults at depth, if any.

[16] Notwithstanding its power and flexibility, the wavelet transform is thus no better than the Hough transform for our present problem of reconstructing a network of fault segments best associated with a given seismic catalog.

3. Dynamic Clustering

3.1. Dynamic Clustering in Two Dimensions

[17] Dynamic clustering (generally known as k means in the pattern recognition literature) is a very general image processing technique that allows one to partition a set of data points using a simple variance criterion. The method is described in many papers or textbooks (see, e.g., *MacQueen* [1967], *Bishop* [2006], or *Duda et al.* [2001]) so it will suffice for our purposes to perform a practical demonstration in two dimensions, but extension of the method to 3-D is straightforward.

[18] We first define a set of N_d data points (x_i,y_i), with $i = 1,\ldots,N_d$. Figure 1a is an example with $N_d = 9$. Our goal is to partition this data set into a given number N_c of clusters, with $N_c < N_d$. In the example illustrated in Figure 1, we chose $N_c = 2$. The first step consists in adding $N_c = 2$ more points to the graph at random locations represented by square symbols (as shown on Figure 1b). Another method consists in choosing those points to coincide with randomly chosen data points. We now link each point from the original set (represented by solid circles) to the closest of the randomly added points (squares), and obtain a first partition of the data set into two clusters (Figure 1c). The first cluster features two data points, while the second cluster features seven data points in our example. Each square is now moved to the barycenter of the associated cluster (Figure 1d). As the squares change position, so do the distances between squares and solid circles. We repeat the partitioning procedure by linking once again each solid circle to the closest square, which updates the partition (see Figure 1e). Clusters now feature four and five data points, respectively. Squares are then moved once again to the barycenters of the associated clusters (see Figure 1f). We

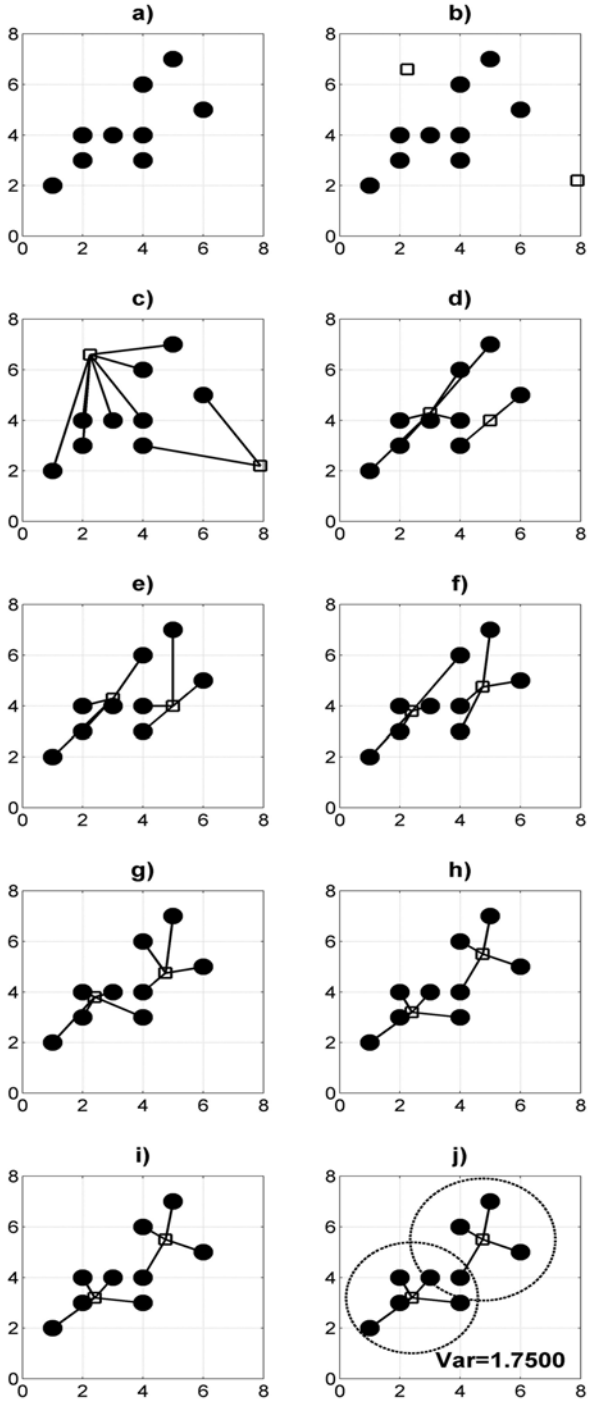

Figure 1. Presentation of the k means method. (a) Set of data points to be partitioned into two clusters. (b) Same as Figure 1a with the addition of the two starting seeds for the clustering procedure (squares). (c and d) First iteration. (e and f) Second iteration. (g and h) Third iteration. (i and j) Fourth and last iteration (with total final variance). The circles in dashed line in Figure 1j give an estimation of the size of clusters.

associate once more each solid circle to the closest square (Figure 1g), which updates the partition once again. The squares being moved to the barycenters of the corresponding clusters, we obtain configuration (Figure 1h). The next iteration is shown on Figures 1i and 1j, and one can check that it does not change the partition anymore (that is, the location of the squares), so that the method converges to a fixed structure, which is the final partition of the data set (the signification of the dashed line circles will be given below).

[19] For any given partition, we can compute its global variance by

$$J = \frac{1}{N_d} \sum_{i_c=1}^{N_c} \sum_{j=1}^{N_{ic}} D_j^2 \tag{1}$$

where i_c runs over all clusters and D_j is the distance between point j belonging to the cluster i_c and the barycenter of that cluster (which features N_{ic} data points). It can be shown [*Bottou and Bengio*, 1995] that the final partition obtained after convergence minimizes the global variance. However, there can exist several local minima for the global variance in the partition space, so that one has to run the k means clustering procedure several times, using different initial locations of the squares in order to select the partition which corresponds to the genuine global minimum.

[20] Figure 1j shows the final partition delineated with two dashed line circles that we obtained by starting from state (Figure 1b). Each circle is centered on the barycenter of a cluster, and its diameter is taken as $\sqrt{12}$ times the square root of the variance of that cluster (the factor $\sqrt{12}$ will be explained in section 3.2). The role of those circles is simply to illustrate the size of each cluster. The total variance of the partition (Figure 1j) is 1.7500 (note that coordinates are dimensionless in our example).

[21] We performed this k means clustering process using hundreds of different initial positions for the squares, and were able to find two other local variance minima (with corresponding partitions illustrated on Figures 2a and 2b). The final variance of the configuration shown in Figure 2a is 1.6296, while the one we computed for Figure 2b is 2.0926 (the variance of the whole initial data set being 4.9722). The partition which minimizes the global variance is thus the one shown in Figure 2a. The three found minima

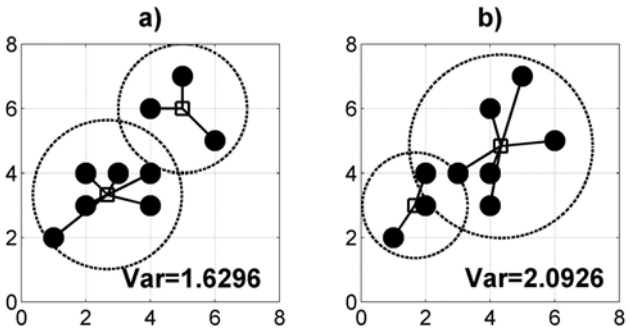

Figure 2. Final partitions for the same data set as in Figure 1a obtained for two different starting configurations of the seed points (squares), with values of the total variances.

are quite similar to each other, and they all appear to be quite reasonable partitions. The closeness of their variances as well as their visual structures suggest that each of these three partitions could be suitable solutions selected by a human operator, reflecting the so-called ''subjectivity.'' Note also that the final partitions display data points that lay almost halfway between two squares, so that adding some noise on the positions of the data points (to simulate earthquake location uncertainties) would certainly make a given final partition switch to one of the two others, and vice versa. It is sometimes suggested to postprocess the results with simulated annealing or genetic algorithms to avoid such local minima [*Kanungo et al.*, 2004]. As the choice of the initial cluster positions completely determines the final partition, *Bradley and Fayyad* [1998] propose a refining method to better choose the starting configuration, consisting in running first a k means procedure over different subsamples of the data set. It is important to note that the k means method and its improvements generally assume that clusters are distributed isotropically around their centers.

3.2. Problem of Anisotropy in Three Dimensions: A New Minimization Criterion

[22] The k means method, as described in section 3.1, is an iterative procedure which minimizes the sum of variances over a given number of clusters. Using a criterion in terms of the minimization of the total variance is certainly not adequate when data points have to be partitioned into planar clusters, as we plan to do by associating earthquakes with faults. As an alternative, we propose a minimization criterion which takes account of the whole covariance tensor of each cluster. As we change the nature of the criterion and give up looking for isotropic clusters, we refer to our method using the more general term of dynamic clustering.

[23] The covariance tensor of a given cluster of points embedded in a 3-D space is equivalent to the covariance matrix of their coordinates (x, y, z). The covariance matrix C reads

$$C = \begin{pmatrix} \sigma_x^2 & \mathrm{cov}(x,y) & \mathrm{cov}(x,z) \\ \mathrm{cov}(x,y) & \sigma_y^2 & \mathrm{cov}(y,z) \\ \mathrm{cov}(x,z) & \mathrm{cov}(y,z) & \sigma_z^2 \end{pmatrix} \quad (2)$$

where the diagonal terms are the variances of the variables x, y and z, while the off-diagonal terms are the covariances of the pairs of variables. Diagonalizing the covariance matrix provides a set of three eigenvalues and their associated eigenvectors. The largest eigenvalue λ_1^2 provides information on the largest dimension of the cluster (thereafter considered as its length), and its associated eigenvector $\mathbf{u}_1$ provides the direction along which the length is striking. The second largest eigenvalue λ_2^2 (and its eigenvector $\mathbf{u}_2$) gives the same kind of information for the width of the cluster, while the smallest eigenvalue λ_3^2 (and its eigenvector $\mathbf{u}_3$) gives an information on the thickness of the cluster. If we now consider that data points cluster around a fault, eigenvalues and eigenvectors provide the dimensions and orientations of the fault plane, the third eigenvector being normal to the fault plane.

[24] The relationship between an eigenvalue and the dimension of the fault along the associated direction depends on the specific distribution of points on the plane. In the following, we shall assume for simplicity that data points are distributed uniformly over a plane. It is easily shown that if x is a 1-D random variable distributed uniformly over the interval $[0, L]$, then we have:

$$\mathrm{Var}(x) = \frac{L^2}{12} \quad (3)$$

so that $L = \sqrt{12 \times \mathrm{Var}(x)}$. We generalize this argument to a $2D$ space so that, if events are uniformly distributed over a fault of length L and width W, then L is equal to $\lambda_1 \sqrt{12}$, and W is equal to $\lambda_2 \sqrt{12}$. The square root of the third eigenvalue is the standard deviation of the location of events perpendicularly to the fault plane. If the fault plane really represents the fault associated with the earthquakes, then λ_3 should be of the order of the location uncertainty. Our idea is to partition the data points by minimizing the sum of all λ_3^2 values obtained for each clustering, so that the partition will converge to a set of clusters that tend to be as thin as possible in one direction, while being of arbitrary dimension in other directions. This procedure defines a set of fault-like objects in the most natural way possible. For each cluster, the third eigenvector $\mathbf{u}_3$ is then sufficient to determine the strike and dip of the fault. Note that we implicitly assume that faults are planar objects. The case of curved faults is beyond the scope of this paper, so that such faults will be considered as a juxtaposition of smaller, planar patches.

4. A New Method: The 3-D Optimal Anisotropic Dynamic Clustering

4.1. Definition of the 3-D Optimal Anisotropic Dynamic Clustering

[25] The general problem we have to solve is to partition a set of earthquake hypocenters into separate clusters labeled as faults. Section 3 has presented the general method to determine the size of a cluster as well as a new minimization criterion for anisotropic dynamic clustering, for a given preassigned number of clusters. We shall now give a brief overview of a new algorithm, which determines automatically the optimal number of clusters N_c to be used in the partition. Allowing N_c to be optimized dynamically is particularly important in order for the fault network to be entirely deduced from the data with no a priori bias. Few methods have been proposed to determine N_c in the case of k means. The X means method of *Pelleg and Moore* [2000] splits local clusters using a Bayesian Information Criterion in order to maximize the model's posterior probability. *Hamerly and Elkan* [2003] introduce the G means method which consists in testing the Gaussian structure of each cluster. If the test fails, the cluster is split in two subclusters. Both methods iteratively increase N_c until some statistical criterion is reached. As those criteria are currently defined only for isotropic cluster structures, and would not be easily extended to a 3-D anisotropic problem, we will use a simpler approached, inspired from the global k means of *Likas et al.* [2003].

[26] In this method, the first step consists in performing a k means over a data set featuring N_d points using $N_c = 1$

(which simply consists in computing the position of the barycenter of the whole data set). *Likas et al.* [2003] then increase N_c to 2 and start N_d other k means procedures with different starting configurations. Each of the N_d different starting configurations consists of one initial barycenter placed at the barycenter of the whole data set, and a second one placed at the location of one of the data points. The final configuration that provides the smallest variance is considered as the optimal one for $N_c = 2$. One then takes $N_c = 3$ and runs N_d more k means. The starting configurations now consists of the optimal solution obtained with $N_c = 2$, and one of the data points, and so on. A rigorous proof that we reach a global minimum for each value of N_c is still missing, but Likas et al. argue that it is the case based on extensive empirical tests. Note however that, while this method allows one to optimally choose the starting positions of the N_c clusters, it does not provide N_c in itself. We thus have to propose a criterion to end the process, and adapt it to anisotropic clusters.

[27] To simplify at this stage of development (this will be refined in the future), we consider that the location uncertainty of earthquakes is uniform in the whole catalog and equal to Δ. The idea is thus that all clusters should be characterized by $\lambda_3 < \Delta$. If not, part of the variance may be explained by something else than location errors, i.e., another fault. The algorithm can be described as follows.

[28] 1. We first consider N_0 faults with random positions, orientations and sizes.

[29] 2. For each earthquake in the catalog, we search for the closest fault and associate the former to the latter. This provides a partition of seismic events in clusters.

[30] 3. For each cluster, we compute the position of its barycenter as well as its covariance matrix. We perform a principal component analysis which provides its eigenvectors and eigenvalues. From those, we compute the strike, the dip, the length and the width, i.e., the characteristics of the fault which explain best the given cluster. The center of the fault is located at the barycenter of the cluster.

[31] 4. If we have $\lambda_3 < \Delta$ for all clusters, the computation stops, as the dispersion of events around faults can be fully explained by location errors. If not, we go back to step 2 and loop until we converge to a fixed geometry, in the spirit of the k means method. Then, if there is at least one cluster for which $\lambda_3 > \Delta$, we go to step 5.

[32] 5. We split the thickest cluster (the one which possesses the largest λ_3 value) by removing the associated fault and replacing it by at least two new faults with random positions within that cluster, and whose lengths are half that of the removed fault. The rationale for this choice is that the cluster with the largest λ_3 is the one worst fitted by a single fault, leading to the hypothesis that a better fit will be obtained by using more faults. This step increases by one unit the total number of fault N_0 used in the partition.

[33] 6. We go back to step 2.

[34] The computation then stops when there are enough faults in the system to split the set of earthquake hypocenters in clusters that all obey the condition $\lambda_3 < \Delta$. In addition, we remove from further analysis those clusters which contain too few events (less than four) or the earthquakes which are spatially isolated. We refer to this new procedure as the Optimum Anisotropic Dynamic Clustering (OADC). Note that it presently uses only hypocentral location parameters. Inclusion of other data such as rupture mechanisms will be discussed in section 5.

4.2. Test on a Synthetic Data Set

[35] This algorithm has been tested on a variety of synthetic sets of events. The synthetic catalogs of earthquakes have been constructed as follows. We first choose the number of faults as well as their characteristics. We then put at random some data points on those planes, and add some noise to simulate the location uncertainty characterizing the location of natural events. This set of data points is then used as an input for the OADC algorithm presented in section 4.1. Our goal is to check if the algorithm is capable to find the characteristics of the original fault network. As we want to ensure that no a priori knowledge of the data set can contaminate the determination of the clusters, we start the algorithm with an initial number of faults equal to $N_0 = 1$.

[36] We shall begin by detailing the example shown in Figure 3a. This data set has been generated using two parallel planes, each of them being normal to a third one. All planes are vertical. Two planes strike N90° (and are located 11 km apart) while the third one strikes N0°. All planes are 20 km long and 10 km wide. We locate randomly (using the uniform random generator ran3 routine of *Press et al.* [1992]) two hundred earthquakes on each of those three planes. For each coordinate of each earthquake, we then use the same routine to generate a uniform random deviate within the interval [-0.01 km; +0.01 km], resulting in a symmetric distribution of events across each plane. We set $\Delta = 0.01$ km and start the 3-D OADC algorithm on this data set. As we start with $N_0 = 1$, the algorithm first fits the data set with a single plane determined from the covariance matrix of the whole catalog, giving the solution shown in Figure 3b. Note that the best fitting plane is horizontal, and thus does not reflect at all the original geometry. Its length is 19.424 km and its width is 16.293 km. For that plane, we find $\lambda_3 = 5.031$ km $> \Delta$, so that a second plane is introduced in order to decrease λ_3. The algorithm then converges to the fit proposed in Figure 3c. Plane A strikes N90° and dips S86°. Its length is 16.945 km and its width is 10.089 km. Its λ_3 value is 2.7055 km. Plane B strikes N90° and dips N87°. Its length and width are 15.682 km and 10.060 km, respectively. Its λ_3 value is 2.855 km. The cluster associated to plane B being the thickest (and its thickness being larger than Δ), we replace its best fitting plane by two new segments, so that there are now three planes available to fit the data. After a few iterations, the algorithm converges to the structure shown in Figure 3d. One can check that the spatial location of each plane is correct, while the azimuths and dips are determined within an error of less than 0.02 degrees. The dimensions of the planes are determined within 2%. The value of λ_3 for each plane is slightly smaller than Δ. The algorithm has thus found by itself that it needed three planes to fit correctly the data set, the remaining variance being explained by the location errors. We processed several times the same data set, using different values for the initial seed of the random number generator. This seed controls the position of the new plane segments we add in the thickest cluster at each step. The algorithm sometimes converges to a final geometry featuring four planes, as shown on Figure 4. Note that two

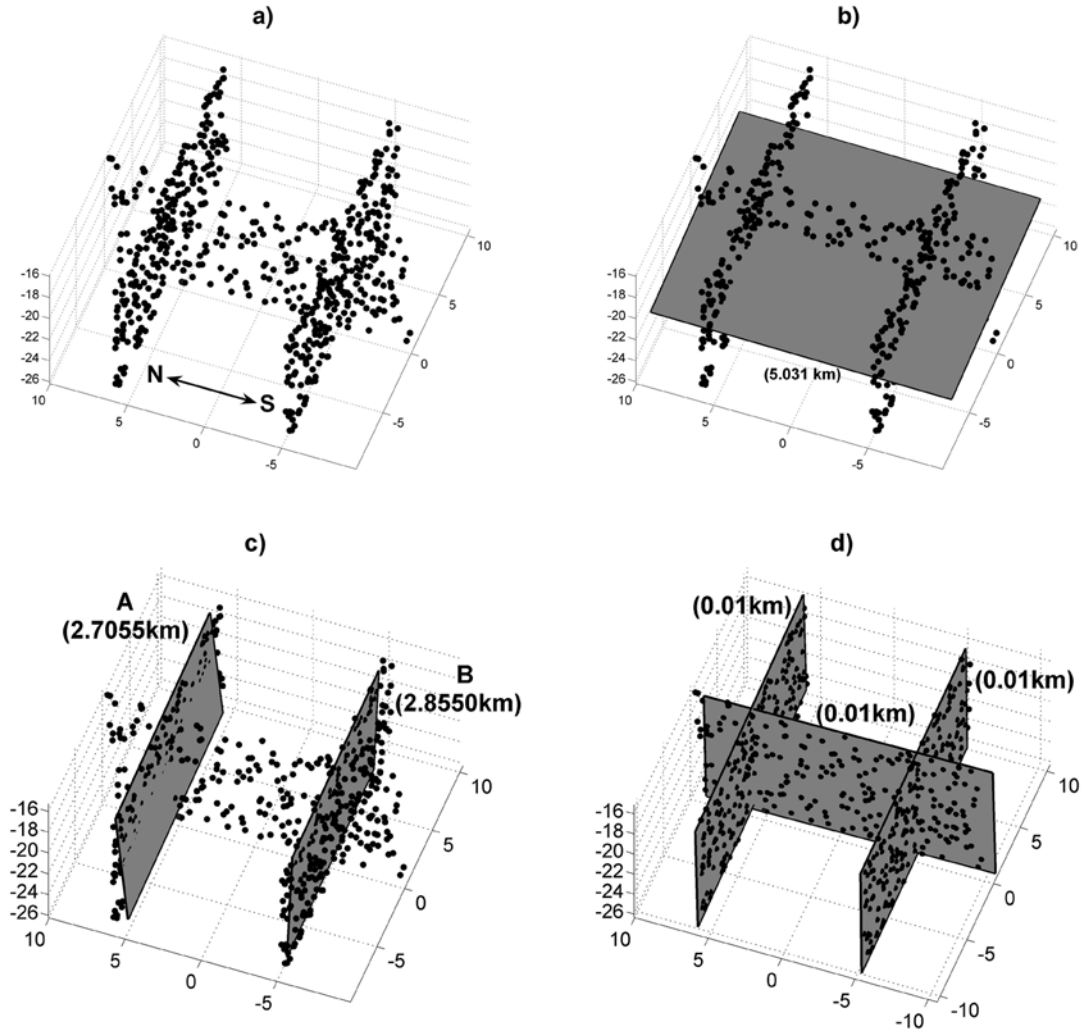

Figure 3. Step by step progression of the OADC procedure proposed in this article, applied to a synthetic data set. (a) The 3-D data set of earthquake hypocenters. (b) A fit with one plane. (c) Best fit with two planes A and B described in the text. (d) Best fit with three planes. In this last case, we recover the correct set of planes with which the data set had been generated. Numbers in parentheses indicate the value of λ_3 for each plane.

of the initial planes are found correctly while the third one has simply been split in two subplanes, but the positions and orientations of all planes still provide a very nice inversion of the data, that would not modify a tectonic interpretation. A postprocessing step, currently under investigation, is required in order to merge such subplanes. Indeed, each time we add new planes, we should perform several trials in order to find the best way to introduce them, in the spirit of the global k means presented above. However, in our case, we would have to test exhaustively all possible positions, sizes and orientations of the newly introduced planes, which is impossible. Future work will focus on this step of the method.

[37] Using synthetic tests developed to demonstrate the power of the method, we considered several other fault geometries, with varying strikes, dips and dimensions. We report that the spatial extent of the reconstructed fault structure was found with great accuracy in each case. Our synthetic tests have been performed with a seismicity uniformly distributed on the fault planes. In the future, we plan to perform more in-depth tests with synthetic catalogs on more complex multiscale fault networks in the presence of a possible nonuniform complex spatial organization of seismicity, so as to better mimic real catalogs. However, since, as we shall see in section 4.3, the extents of faults retrieved from natural data sets seem reasonable as well, we conjecture that our method is robust with respect to the presence of heterogeneity.

4.3. Application to the Landers Aftershock Sequence

4.3.1. Implementation

[38] As the method yields correct results on simple synthetic examples, we are encouraged to apply it to real catalogs, such as the aftershock sequence of the 1992 Landers earthquake in southern California. We used the locations provided by *Shearer et al.* [2005] and considered only the first two weeks of aftershock activity, i.e., a set of 3503 events. We used such a limited subset of the full

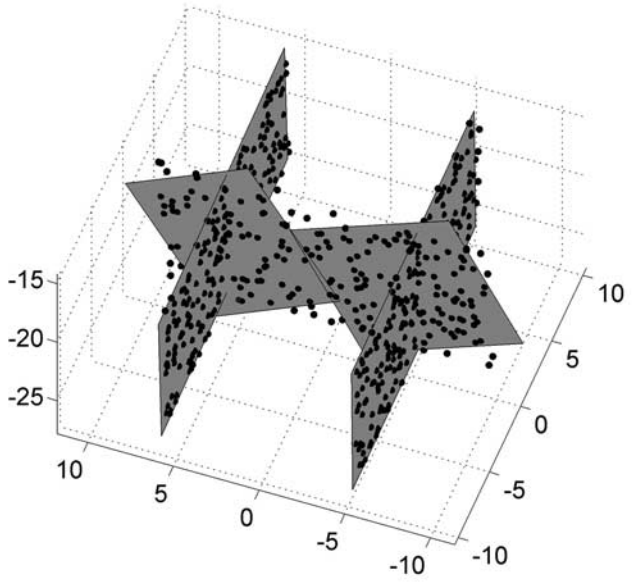

Figure 4. Final partition for the same data set as in Figure 3a obtained for a different starting random number generator seed. Note that the N-S plane is now split into two subplanes.

sequence due to limits set by the required computation time. In the future, we plan to parallelize the algorithm in order to be able to deal with much larger catalogs. The relative location uncertainty is reported to be a few tens of meters [*Shearer et al.*, 2005], but we set $\Delta = 1$ km to perform a first coarse-scale inversion. Our justification for setting $\Delta = 1$ km is twofold. First, the absolute location errors are certainly larger than a few tens of meters. Second, using a value of Δ smaller than 1 km would require using a much larger catalog, with many more events to correctly sample small-scale features, hence increasing the computation time prohibitively. We detail in Figures 5a–5f the progression of the algorithm as the number of introduced faults increases, as we did in section 4.2 for the synthetic case. As the 3-D spatial structure of the sequence is quite complex, we show only the projection of the results onto the horizontal plane. Note that coordinates on all plots are given in UTM (Universal Transverse Mercator) zone 11 projection.

[39] Figure 5a shows the result of the fit with a single fault, for which we obtain $\lambda_3 = 2.9$ km. Since this value is larger than $\Delta = 1$ km, we introduce a second fault, and obtain the pattern shown in Figure 5b. The second fault is found to improve the fit mainly at the southern part of the sequence, where a large number of events are located. The largest λ_3 is still about 2.9 km. We thus introduce a third fault, and obtain the optimal three fault structure shown in Figure 5c. The new fault now helps to improve the quality of the fit in the northern end. This third fault is found to be horizontal, a signature of the competition between several strands, as observed in the synthetic test presented in Figure 3b. Indeed, one can observe visually that this northern region is characterized by at least three branches. The value λ_3 for this last fault is about 2.5 km. We thus introduce a fourth fault, and obtain the optimal structure shown in Figure 5d. Now, the northern end is fitted more satisfactorily, and the largest λ_3 value drops to 1.8 km. The algorithm thus continues to introduce new faults, and we proceed directly to the pattern we obtain with eight faults which is shown in Figure 5e. At this stage, the algorithm has dissected the central part of the data set, but still does not provide a good fit to the E-W branch located at 3.825×10^6 on the latitude axis. As the largest value λ_3 is 1.4 km, the algorithm still introduces more faults. Figure 5f shows the optimal fault structure fitting the data set of aftershock hypocenters with 12 faults. The northern part now appears to be fitted very nicely, while the southern part looks quite complex. As the largest value λ_3 is now 1.14 km, the algorithm requires more faults to fit the data with the threshold condition $\Delta = 1$ km. We find that the largest value of λ_3 drops below $\Delta = 1$ km for 16 faults, for which the computation stops and yields the final pattern shown in Figure 6. When interpreting Figure 6, it is important to realize that some fault planes are nearly vertical, which make them barely visible in the projection shown in Figure 6. White planes dip less than 45°.

[40] In order to describe the fault network, it is convenient to label the 16 faults from A to P, which allows us to discuss this pattern fault by fault. The parameters of the 16 fault planes (position, size and orientation) are given in Table 1. These fault planes will now be classified into three different categories, namely, (1) spurious planes (which have no apparent significance), (2) previously known planes (that correspond to mapped faults), and (3) unknown planes (that may correspond to blind faults or to otherwise structures unmapped for whatever reason).

4.3.2. Spurious Planes

[41] An inspection of Table 1 reveals that most proposed planes dip close to the vertical. Three planes, H, I, and N (plotted as white planes in Figure 6), have rather abnormal dips, which leads us to suspect that they are spurious. Indeed, H and I are near normal to plane G, which is located in a zone with rather diffuse seismicity in the direction normal to that plane. Introducing planes H and I is a convenient way to reduce the variance in that zone. It is likely that those planes have no tectonic significance, but have been found by the algorithm as the way to satisfy the criterion on λ_3 (we shall come back below to this argument). Plane N also seems to be introduced just to decrease the variance in a zone displaying fault branching. All other 13 fault planes out of the 16 have dips larger than 50, so that we have a priori no reason to doubt the reality of their existence. Figure 7 shows the 2-D map of the epicenters of the events associated with the three spurious faults and of those associated with the other 13 planes. Notice that the clusters associated with the spurious faults are significantly more diffuse in two dimensions than those associated with the 13 other faults. Note that the subset of seismicity associated with the planes H and I includes a dense subcluster which is compatible with the presence of fault G. Those events are indeed located at the intersections between the spurious planes and plane G. Future implementation of the method will require to better discriminate between diffuse background events and well-organized seismicity. Ways to improve the OADC method in this direction will be briefly discussed in section 5.

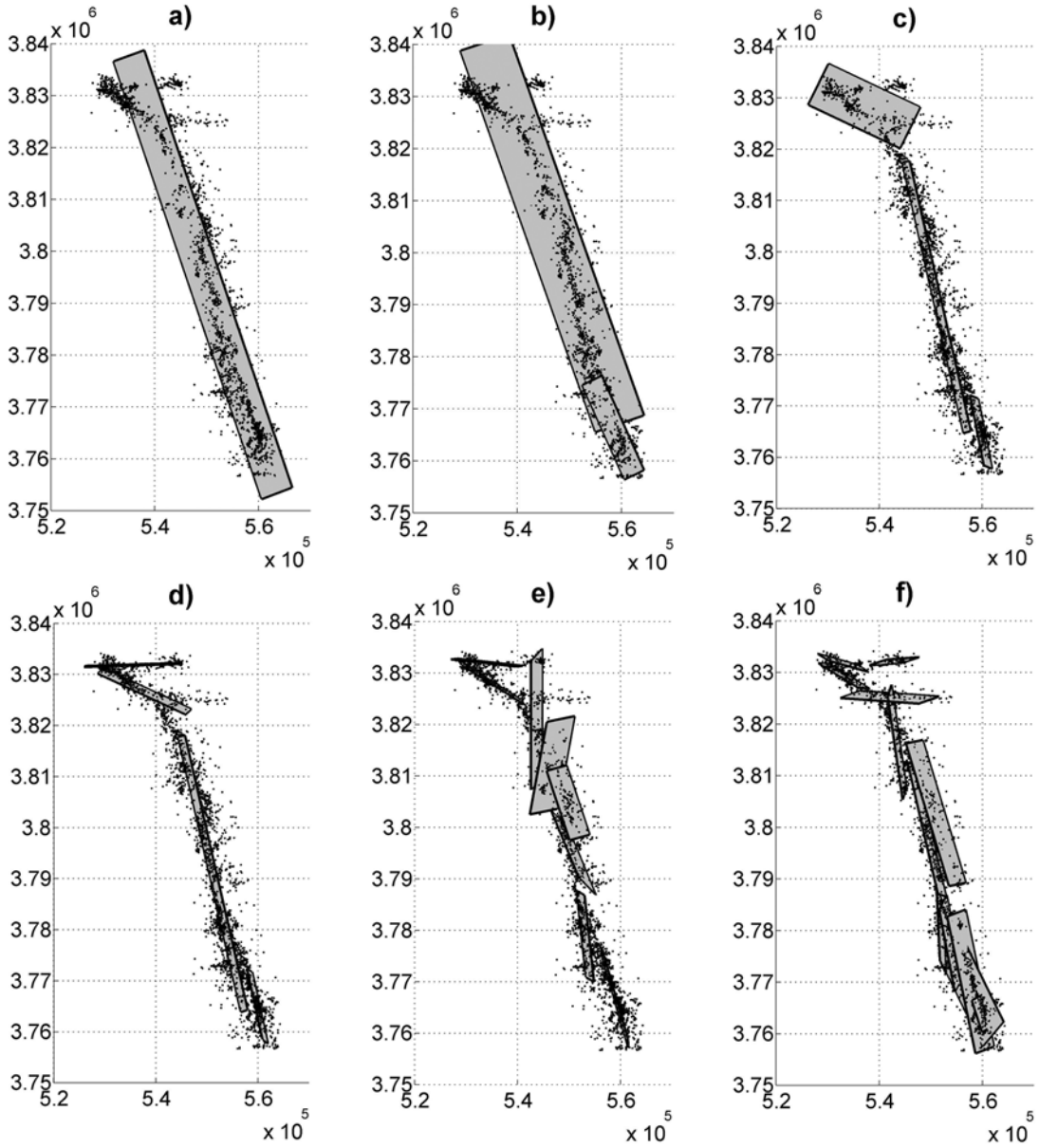

Figure 5. Progression of the OADC method when applied to a subset of the 1992 Landers, California, aftershocks sequence. (a) Fit of the sequence with one fault plane. (b) Fit with two fault planes. (c) Fit with three fault planes. (d) Fit with four fault planes. (e) Fit with eight fault planes. (f) Fit with 12 fault planes.

4.3.3. Previously Known Faults

[42] Because the fault planes have been obtained by fitting seismicity data, they do not crosscut the free surface. It is however instructive to compare the planes in Figure 6 to the faults mapped at the surface in the Landers area [see *Liu et al.*, 2003], and to underline possible correspondences. In order to facilitate such a comparison, we first plotted a map of surface-intersecting fault traces computed and reported in the Community Fault Model database (see Figure 8a). We then extrapolated all fault planes obtained with the OADC method and dipping more than 45 to the free surface, and computed their theoretical traces. The free surface was considered to coincide with the $z = 0$ surface, so that we did not take account of topographical features. Those traces are plotted in Figure 8b, together with their corresponding labels. From those plots, we conclude that planes C and E clearly correspond to the northern end of the Emerson fault. Plane G is a good candidate for the Homestead Valley fault. Continuing to the south, plane K seems to match with the southern end of the Johnson Valley fault. Plane L coincides with the northern end of the Burnt Mountain fault, and extends it further to the north. Plane P corresponds to the Eureka Peak fault. All those faults are known to have been activated during the Landers event. In the southern end, plane M fits well with the Pinto Mountain fault.

4.3.4. Unknown Faults

[43] This last set contains planes A, B, D, F, J, M, and O. Planes A, B, D, and F are good candidates for blind faults,

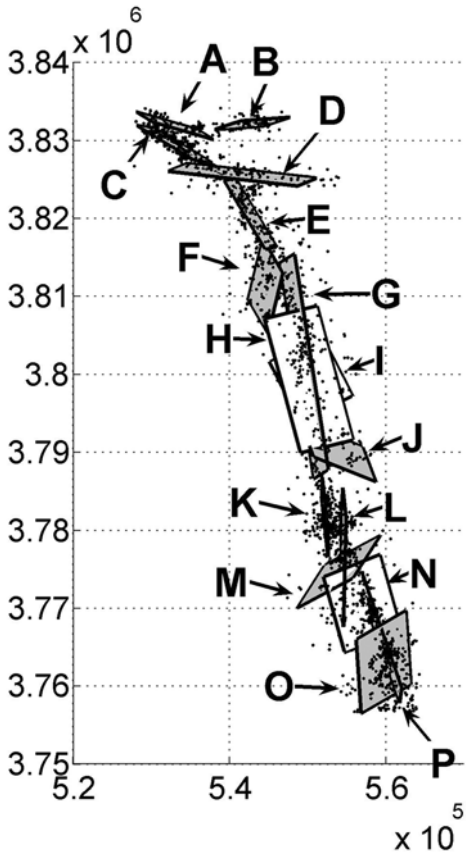

Figure 6. Fit with 16 fault planes of the 1992 Landers, California, aftershock data set. The value λ_3 of each cluster is smaller than the imposed resolution threshold $\Delta = 1$ km. White planes are considered as spurious (see text).

which do not appear in the Community Fault Model, but seem necessary to account for the clustering of seismicity in that region. This can be seen more clearly in Figure 9 which plots the full set of events found in *Shearer et al.*'s [2005] catalog in the same area, with and without the retrieved fault planes of Figure 6. Note that the fault pattern we retrieved from a quite limited data set agrees very well with the seismicity recorded on a much longer timescale, strengthening our confidence in the potential of this new method. Plane J is located in a zone of complex faulting and may feature events belonging to different structures, including the Homestead Valley fault. Plane O would be a genuine blind fault, that the OADC method has discovered.

4.3.5. Alternate Solutions

[44] The previous set of planes is the result of a single inversion. It is important to keep in mind that inversions contain a random component due to the random splitting of the cluster with the largest variance. To test for the robustness of our results, we processed the same data set repeatedly, using different initial values for the random number generator seed. Figure 10 displays the traces of the fault patterns obtained for nine different runs (in addition to the solution discussed previously). The number of retrieved faults varies from 13 (12) to 22 (20) (numbers within brackets refer to faults that dip 45° and more). Note that runs in Figures 10a and 10h converged exactly to the same solution as the one shown on Figure 8b (which is the reason why we chose to compare its structure with the CFM fault trace network). More runs are needed to check if a majority of inversions converge to that same solution (see section 5). Some of the solutions could be eliminated as they provide a too large number of crossing or too closely overlapping faults (Figures 10d, 10e, 10g and 10i, for example), which do not look like realistic fault patterns. If we accept that criterion, it turns out that the pattern shown on Figure 8b is the solution obtained in half of the runs.

5. Discussion and Conclusion

[45] We have introduced a new powerful method, the 3-D Optimal Anisotropic Dynamic Clustering (OADC), that allows us to partition a 3-D set of earthquake hypocenters into distinct anisotropic clusters which can be interpreted as fault planes. The method was first applied to synthetic data

Table 1. Correspondence Between Fitting Planes of the Landers, California, Sequence Obtained With the OAWC Method and Known Faults in the Same Area[a]

Label	Name	Longitude	Latitude	Depth, km	Strike	Dip	Length, km	Width, km	λ_3, km	N
A	blind	−116.532	34.632	6.3	84.3	85.2	9.84	6.71	2.53	99
B	blind	−116.640	34.630	5.8	288.7	84.4	8.75	6.33	2.36	133
C	Emerson	−116.640	34.608	6.4	299.3	83.4	8.99	7.50	2.15	345
D	blind	−116.546	34.573	6.8	275.4	83.6	17.02	11.25	3.07	178
E	Emerson	−116.535	34.527	6.7	329.3	82.9	11.39	10.19	2.87	210
F	Homestead Valley	−116.514	34.444	6.6	186.8	72.4	10.78	8.58	3.46	194
G	Homestead Valley	−116.461	34.351	6.8	172.1	80.8	28.18	12.59	3.20	446
H	spurious	−116.455	34.336	3.4	343.2	34.9	17.86	8.43	2.72	193
I	spurious	−116.452	34.336	9.9	324.9	17.9	8.57	7.84	3.43	90
J	Homestead Valley	−116.415	34.252	4.9	299.5	70.8	10.10	9.12	3.11	59
K	Johnson Valley	−116.433	34.176	7.3	174.6	88.4	12.82	7.39	2.84	280
L	Burnt Mountain	−116.409	34.129	6.5	180.7	87.4	15.05	10.82	2.54	306
M	Pinto Mountain	−116.415	34.113	7.2	225.4	75.0	11.80	8.41	2.63	108
N	spurious	−116.383	34.076	6.5	58.7	39.2	12.78	7.83	2.70	117
O	blind	−116.354	34.008	4.0	26.5	51.8	11.31	8.99	3.44	226
P	Eureka Peak	−116.359	34.041	4.2	164.1	81.8	17.55	5.80	2.63	519

[a]Each fitting plane is named with the label given in Figure 6. The name of the corresponding natural fault (if it is known) is given. Spurious indicates that the plane has no tectonic significance, while blind means that the natural fault does not intersect the free surface and was not known before this study. Each plane is characterized by the latitude, longitude, and depth of its center, as well as by its strike, dip, and dimensions in km. The values are given of λ_3 of the corresponding cluster and the total number of events in that cluster.

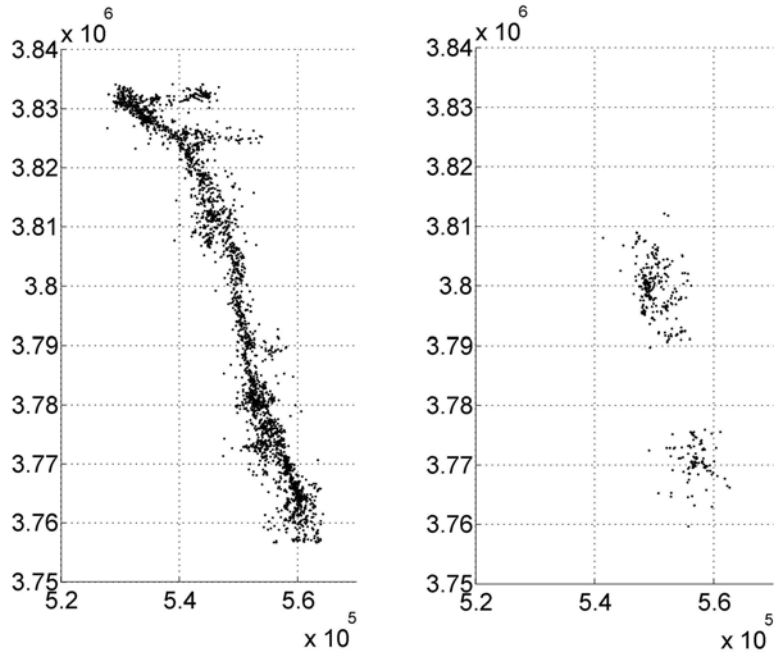

Figure 7. Epicentral maps of events of the Landers aftershock sequence associated (left) with nonspurious faults and (right) with spurious faults H, I, and N.

sets, and these tests confirmed its ability to recover correctly with high accuracy the existing faults. We then applied the OADC method to the sequence of aftershocks following the 1992 Landers event in California. In this later case, we were able to recognize the faults that were already known by surface mapping. In addition, the OADC method allowed us to identify some additional blind faults. These faults appear to make sense when looking at the long-term epicenter maps or 3-D plots of seismicity. One could argue that the keen eyes of a geologist would have come up with similar results. However, the advantage of our method is that it finds automatically the characteristics of the fault planes as well as the hypocenters that belong to each of them, without any need for the operator to pick them manually. The method provides the orientation of the faults, as well as the size of their active part during the earthquake sequence used to

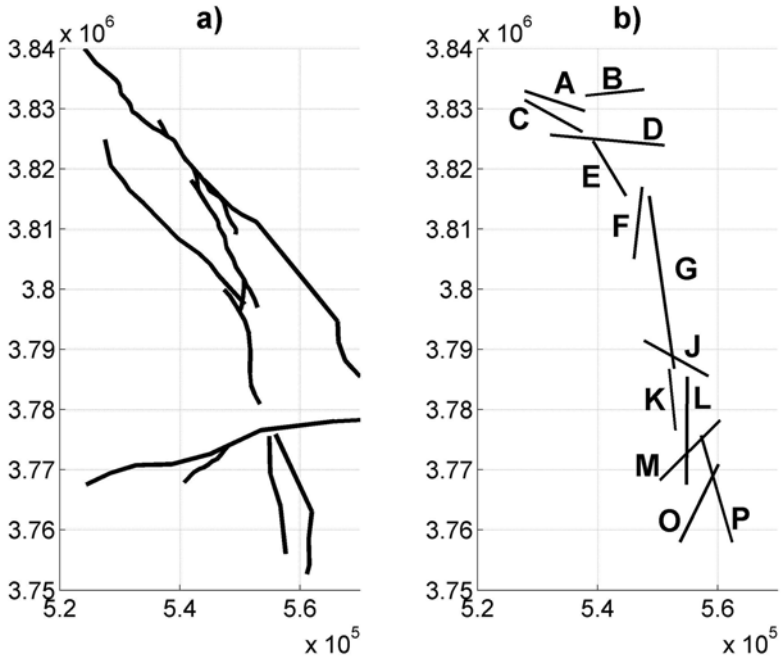

Figure 8. (left) Observed fault traces taken from the Community Fault Model and (right) extrapolation of nonspurious retrieved fault planes of Figure 6 to the free surface.

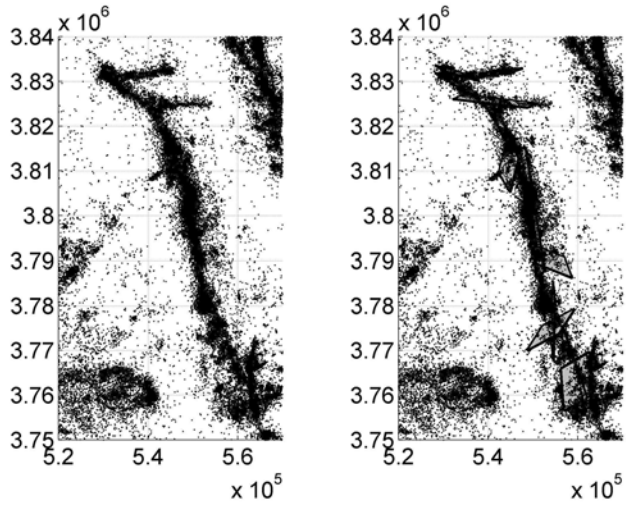

Figure 9. Epicentral view of the full catalog available for the Landers area (left) without and (right) with superimposition of the non spurious retrieved faults of Figure 6.

identify them. We will see below that the next improvements of the method can take account of different types of quantitative data, as well as provide a description of the fault network in probabilistic terms.

[46] The main drawback of our new method is that it can propose fault planes that are likely spurious (error of type II), as they lack any tectonic meaningful interpretation. This problem is due to the presence of diffuse seismicity, for which the OADC method forces the introduction of fault planes in the diffuse zone in order to reduce the global variance of the distances from earthquakes to the proposed fault planes. In the case of the Landers aftershock sequence, we were able to sort them out using their anomalous dip parameters, compared with all known fault planes in the Landers area which are nearly vertical, as also confirmed by focal mechanisms. This postprocessing step will become much more subjective when trying to sort out spurious planes in more complex tectonic settings. We thus believe that the OADC method should be enriched by additional information available in earthquake catalogs besides the 3-D hypocenter locations. For example, focal mechanism characteristics are now determined for a large number of events, so that the event/cluster/fault association should also use a criterion of compatibility between the focal mechanism of each event with the orientation of the best fitting plane of the cluster. The idea is to make the algorithm converge to a set of anisotropic clusters over which the distribution of focal mechanisms is approximately uniform. This extension of the method will be developed and tested elsewhere.

[47] In the same vein, we propose another future extension of the OADC method, consisting in using the information on waveforms contained in the southern California earthquake catalog of *Shearer et al.* [2005]. The method used to build this catalog consists in grouping events according to the similarity of their waveforms. This first step thus yields a protoclustering of events. Time delays of wave arrival times for events belonging to the same protocluster yield relative locations of those grouped events. Events belonging to the same protocluster display similar waveforms and are thought to be characterized by similar rupture mechanisms, and may thus reveal successive ruptures on the same fault. *Shearer et al.* [2005] used this idea and performed principal component analyses on each protocluster of the Northridge, California, aftershock sequence. This idea is similar to ours, but nothing proves that a single protocluster samples only one fault segment. For example, when looking at the seismicity in the Imperial fault area, *Shearer* [2002] has been able to identify some very fine scale features in earthquake clouds that otherwise looked very fuzzy. In this last example, a protocluster of events was found to be clearly composed of two parallel lineaments, with a 500 m offset between them. However, the operator still has to use a ruler and a protractor to compute the dimension and orientation of each lineament, and his eyes to count them. Using a method such as ours on that specific protocluster would naturally partition it into two (sub)clusters, and provide their associated dimensions and orientations. Thus a natural future application of our method would be to apply it on each of the clusters resulting from *Shearer et al.*'s [2005] protopartition. The advantage is that the number of events in each cluster is small, so that the algorithm will converge very quickly, and that events will be already presorted according to the similarity of their rupture mechanism. The fact that all events within protoclusters are located with an accuracy of the order of a few tens of meters should allow one to provide a very detailed fit of the 3-D spatial structure of the catalog.

[48] Another drawback of the present version of the OADC method is that each earthquake is assigned to a given fault plane with certainty. In other words, each earthquake is with probability $P = 1$ attached to a given fault plane and thus with zero probability to any other fault plane. *Wesson et al.* [2003] used a Bayesian inference approach to compute the probability that a given earthquake has been triggered on a given fault. Their method also allows one to compute the probability that an event belongs to the background seismicity (i.e., that it does not belong to any fault), but it requires the a priori knowledge of the full fault network to compute such probabilities. As most of this network is unknown, especially at depth, the method of *Wesson et al.* [2003] assigns to the background activity some subsets of events that clearly display organized, fault-like structures on seismicity maps. A natural idea we propose is to couple their approach and ours, performing Bayesian inference at each dynamic clustering step and stopping the procedure when the probabilistic partition of events into faults and background converges to a fixed point. Note that Bayesian inference will lead naturally to introduce earthquake location uncertainties that may vary from event to event (whereas it is considered as constant in the current version of the method, an oversimplification of

Figure 10. Extrapolation of nonspurious retrieved fault planes to the free surface for the Landers aftershock sequence. (a) to (i) Maps corresponding to a different OADC processing of the Landers data set, using different random generator seeds. See the main text for a discussion their respective likelihoods.

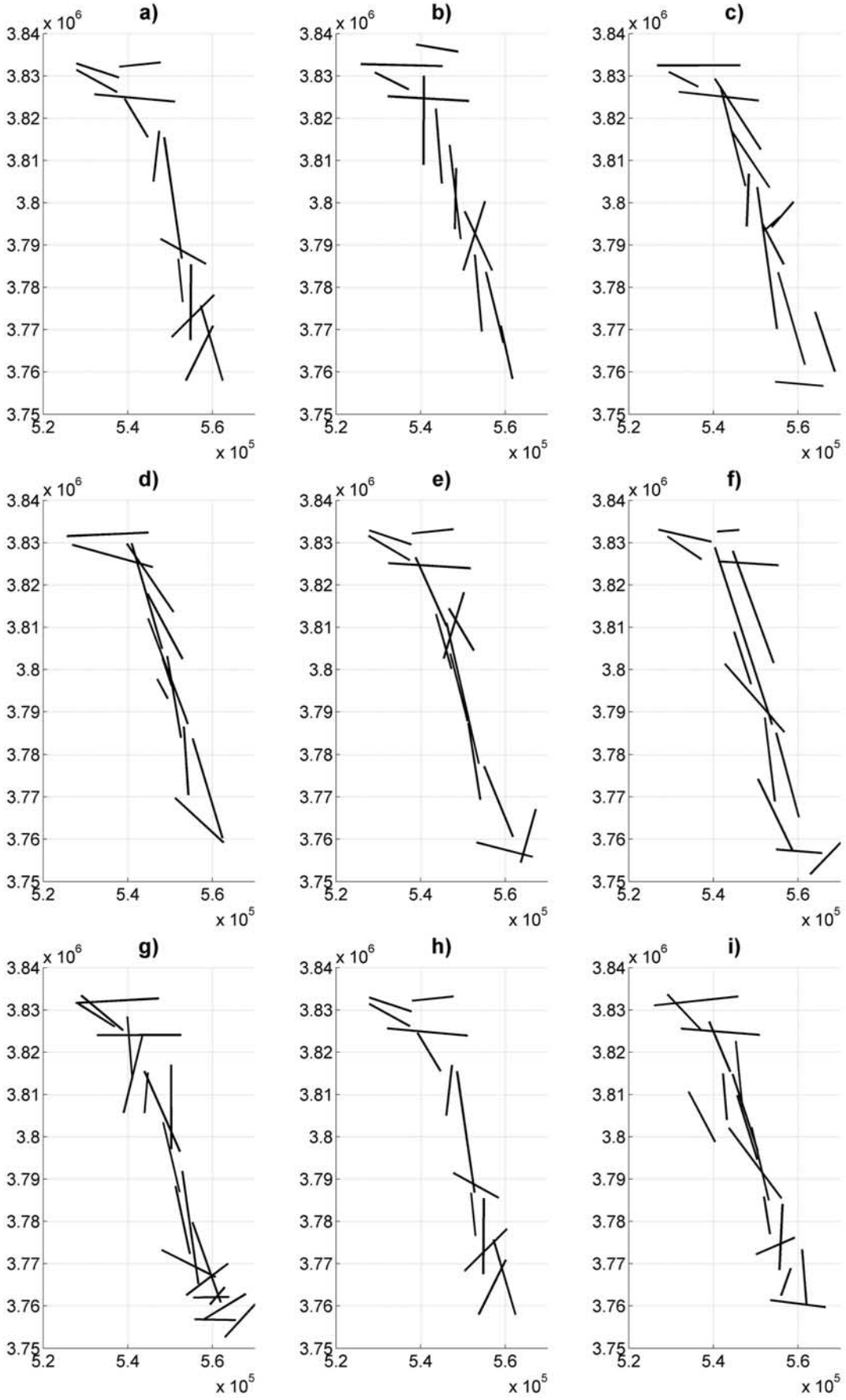
a)
b)
c)
d)
e)
f)
g)
h)
i)
x 10^6
3.84
3.83
3.82
3.81
3.8
3.79
3.78
3.77
3.76
3.75
5.2
5.4
5.6
x 10^5

Figure 10

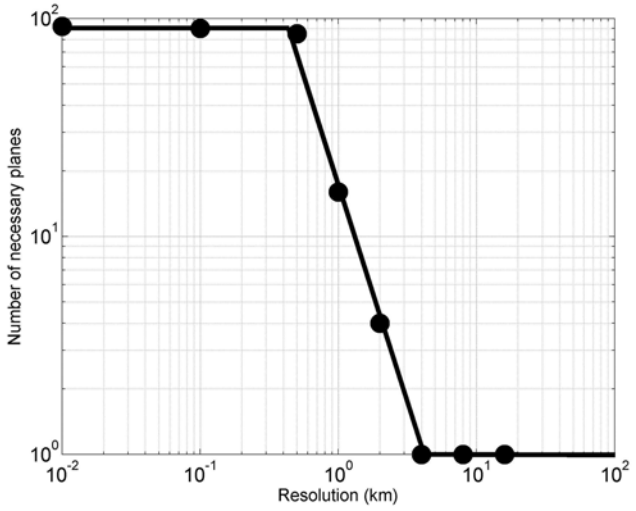

Figure 11. Number of planes necessary to fit the set of aftershocks of the Landers earthquake, as a function of the resolution Δ. Note the finite range of scales over which the power law behavior $N(\Delta) \simeq \Delta^{-2}$ is valid.

reality). As the final partition may not be a global minimum in terms of variance, we also plan to plug the k harmonic means method of *Zhang et al.* [1999] (which uses harmonic averages of distances instead of the true variance) into the OADC method, as it is much less sensitive to initial conditions than the classical k means method. A natural starting fault network configuration could then be the Community Fault Model, that the OADC method would update and refine iteration by iteration.

[49] We would also like to stress that the 3-D Optimal Anisotropic Dynamic Clustering method provides a natural tool for understanding the multiscale structure of fault zones. *Liu et al.* [2003] showed that the spatial dispersion of aftershock epicenters normal to the direction of the Landers main rupture could not be explained by location errors alone, revealing the complex mechanical behavior of this fault zone. In the examples presented above, we considered that the resolution parameter Δ controlling the number of faults necessary to explain the seismic data could be mapped in a natural way to the uncertainty of the spatial location of the events. While this is a natural choice, more generally, Δ can take any arbitrary value assigned by the user and, in such a case, it must be interpreted as the spatial resolution at which the user needs to approximate the anisotropic fault structure defined by the catalog of events. For example, if we consider the choice $\Delta = 10$ km, the output is the one presented in Figure 5a, as only one fault is necessary to fit the data at this coarse resolution. As Δ decreases, the number of necessary faults increases. Figure 11 shows that the number of fault planes necessary to fit the Landers' data varies with Δ as

$$N(\Delta) \simeq \Delta^{-\mu}, \quad \text{with} \quad \mu = 2, \tag{4}$$

over a limited range (note that we processed data only once for each value of Δ, whereas future investigations will average over many inversions). For very large Δ, it is obvious that just one fault is necessary. For small Δ, the number of necessary planes reaches a limit as we need at least three data points to define each plane. If there are N_d data points, then the upper limit for the maximum number of planes will be of the order of $N_d/3$. Note that in the case of the Landers sequence, the maximum number of planes is much lower than this limit, which is due to the nonrandom organization of hypocenters and maybe to the fact that different events are located at the same place. This power law behavior is of course reminiscent of a scale-invariant geometry, but we still have to clarify a few points before drawing any conclusion. The first one is a theoretical one: usual methods to compute the fractal dimension of a given object consider scale-dependent approximations of that object with isotropic balls (for example, box counting uses isotropic balls with the L_1 norm, while the Hausdorff method uses isotropic balls with the L_2 norm). In the present case, we approximate our data set by highly anisotropic "balls," which in addition do not have the same size. It thus seems inappropriate to interpret μ as a genuine fractal dimension. The second is a pragmatic one: For a given scale-invariant distribution of points, the measured fractal dimension depends on its cardinal number. The reason is that, the larger the cardinal, the better the distribution is sampled, and the better is its statistical characterization. If the cardinal is low, the multiscale structure of the distribution is not properly sampled and the measured fractal dimension is biased (see *Ouillon et al.* [1995] or *Ouillon and Sornette* [1996]). As we still do not know the effect of the bias that may affect the determination of the exponent μ, much more work is needed in that direction.

[50] Considering earthquake forecasting and seismic hazard assessment, the knowledge of the fault network architecture in the vicinity of large events will also help to test different competing hypothesis on stress transfer mechanisms. Until now, two strategies have been used in order to check for the consistency between the geometry of the main shock rupture and the spatial location of aftershocks. The first one considers that all rupture planes of aftershocks are parallel to the main rupture. The second one postulates that the rupture planes are optimally oriented. None of those hypotheses turn out to be true [see *Steacy et al.*, 2005]. Therefore we think that imaging accurately the 3-D structure of a fault network constitutes one of the most important steps in order to decipher static or dynamic earthquake triggering, as well as to understand the mechanics of faulting on timescales up to a few decades. The 3-D Optimal Anisotropic Dynamic Clustering provides a first significant step in this direction.

[51] **Acknowledgments.** The authors would like to thank Heidi Houston and Peter Shearer for precious information about the Landers catalog, as well as Andreas Plesch as a referee and Frederik J. Simons, Associate Editor, for their very constructive comments on the first version of the manuscript.

References

Bhattacharya, P., H. Liu, A. Rosenfeld, and S. Thompson (2000), Hough transform detection of lines in 3D space, *Pattern Recog. Lett.*, *47*, 65–73.

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

Bottou, L., and Y. Bengio (1995), *Convergence Properties of the K-Means Algorithm, Adv. Neural Inf. Processing Syst.*, vol. 7, MIT Press, Denver.

Bradley, P., and U. M. Fayyad (1998), Refining Initial Points for K-Means Clustering, in *Proceedings of the Fifteenth International Conference on Machine Learning (ICML '98)*, edited by J. Shavlik, pp. 91–99, Morgan Kaufmann, San Francisco, Calif.

Cohen, A., I. Daubechies, and P. Vial (1993), Wavelets on the interval and fast wavelet transforms, *Appl. Comput. Harmon. Anal.*, *1*, 54–81.

Darrozes, J., P. Gaillot, and P. Courjault-Radé (1998), 2D propagation of a sequence of aftershocks combining anisotropic wavelet transform and GIS, *Phys. Chem. Earth*, *23*, 303–308.

Duda, R. O., and P. E. Hart (1972), Use of the Hough transformation to detect lines and curves in pictures, *Commun. ACM*, *15*, 11–15.

Duda, R. O., P. E. Hart, and D. G. Stork (2001), *Pattern Classification*, 2nd ed., John Wiley, New York.

Faulkner, D. R., A. C. Lewis, and E. H. Rutter (2003), On the internal structure and mechanics of large strike-slip fault zones: Field observations of the Carboneras fault in southeastern Spain, *Tectonophysics*, *367*, 235–251.

Gaillot, P., J. Darrozes, M. De Saint Blanquat, and G. Ouillon (1997), The normalised optimised anisotropic wavelet coefficient (NOAWC) method: an image processing tool for multi-scale analysis of rock fabric, *Geophys. Res. Lett.*, *24*, 1819–1822.

Gaillot, P., J. Darrozes, and J. L. Bouchez (1999), Wavelet transforms: The future of rock shape fabric analysis?, *J. Struct. Geol.*, *21*, 1615–1621.

Gaillot, P., J. Darrozes, P. Courjault-Rad, 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.

Grégoire, V., J. Darrozes, P. Gaillot, P. Launeau, and A. Nédélec (1998), Magnetite grain shape fabric and distribution anisotropy versus rock magnetic fabric: A 3D-case study, *J. Struct. Geol.*, *20*, 937–944.

Guzofski, C. A., J. H. Shaw, G. Lin, and P. M. Shearer (2007), Seismically active wedge structure beneath the Coalinga anticline, San Joaquin basin, California, *J. Geophys. Res.*, *112*, B03S05, doi:10.1029/2006JB004465.

Hamerly, G., and C. Elkan (2003), Learning the k in k-means, paper presented at the Seventeenth Annual Conference on Neural Information Processing Systems (NIPS), Neural Inf. Process. Syst., Vancouver, B.C., Canada.

Jones, R. H., and R. C. Stewart (1997), A method for determining significant structures in a cloud of earthquakes, *J. Geophys. Res.*, *102*, 8245–8254.

Kanungo, T., D. M. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A. Y. Wu (2004), A local Search Approximation Algorithm for k-Means Clustering, *Comput. Geometry Theory Appl.*, *28*, 89–112.

Likas, A., N. Vlassis, and J. J. Verbeek (2003), The global k-means clustering algorithm, *Pattern Recog.*, *36*(2), 451–461.

Liu, J., K. Sieh, and E. Hauksson (2003), A structural interpretation of the aftershock "cloud" of the 1992 M_w 7.3 Landers earthquake, *Bull. Seismol. Soc. Am.*, *93*, 1333–1344.

MacQueen, J. (1967), Some methods for classification and analysis of multivariate observations, in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, edtied by L. M. Le Cam and J. Neyman, pp. 281–297, Univ. of Calif. Press, Berkeley.

Nicholson, T., M. Sambridge, and Ó. Gudmundsson (2000), On entropy and clustering in earthquake hypocentre distributions, *Geophys. J. Int.*, *142*, 37–51.

Ouillon, G. (1995), Application de l'analyse multifractale et de la transformée en ondelettes anisotropes à la caractérisation géométrique multiéchelle des réseaux de failles et de fractures, *Doc. BRGM*, *246*.

Ouillon, G., and D. Sornette (1996), Unbiased Multifractal Analysis of Fault Patterns, *Geophys. Res. Lett.*, *23*, 3409–3412.

Ouillon, G., C. Castaing, and D. Sornette (1995), Hierarchical Geometry of Faulting, *J. Geophys. Res.*, *101*, 5477–5487.

Pelleg, D., and A. Moore (2000), X-means: Extending k-means with efficient estimation of the number of clusters, in *Proceedings of the Seventeenth International Conference on Machine Learning*, pp. 727–734, Morgan Kaufmann, San Francisco, Calif.

Plesch, A., and J. H. Shaw (2002), SCEC 3D Community Fault Model for Southern California, Eos Trans., *AGU*, *83*(47), Fall Meet. Suppl.Abstract S21A-0966.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992), *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd ed., Cambridge Univ. Press, New York.

Sarti, A., and S. Tubaro (2002), Detection and characterization of planar fractures using a 3D Hough transform, *Signal Processing*, *82*, 1269, doi:10.1016/S0165-1684(02)00249-9.

Scholz, C. (2002), *The Mechanics of Earthquakes and Faulting*, Cambridge Univ. Press, New York.

Shearer, P. M. (2002), Parallel fault strands at 9-km depth resolved on the Imperial Fault, southern California, *Geophys. Res. Lett.*, *29*(14), 1674, doi:10.1029/2002GL015302.

Shearer, P., E. Hauksson, and G. Lin (2005), Southern California hypocenter relocation with waveform cross-correlation, part 2: Results using source-specific station terms and cluster analysis, *Bull. Seismol. Soc. Am.*, *95*, 904–915.

Steacy, S., S. S. Nalbant, J. McCloskey, C. Nostro, O. Scotti, and D. Baumont (2005), Onto what planes should Coulomb stress perturbations be resolved?, *J. Geophys. Res.*, *110*, B05S15, doi:10.1029/2004JB003356.

Wesson, R. L., W. H. Bakunm, and D. M. Perkins (2003), Association of earthquakes and faults in the San Francisco Bay area using Bayesian inference, *Bull. Seismol. Soc. Am.*, *93*, 1306–1332.

Zhang, B., M. Hsu, and U. Dayal (1999), K-harmonic means—A data clustering algorithm Tech. Rep. HPL-1999-124, Hewlett-Packard Labs, Palo Alto, Calif.

C. Ducorbier, Géosciences Azur, UMR 6526 CNRS-UNSA-IRD, Mecatec Unit, 250 rue Albert Einstein, Les Lucioles I, Sophia Antipolis, F-06560 Valbonne, France. (ducorbier@geoazur.unice.fr)

G. Ouillon, Lithophyse, 4 rue de l'Ancien Sénat, F-06300 Nice, France. (lithophyse@neuf.fr)

D. Sornette, D-MTEC, ETH Zürich, Kreuzplatz 5, CH-8032 Zürich, Switzerland. (dsornette@ethz.ch)