Application of artificial intelligence for Euler solutions clustering

Valentine Mikhailov, Armand Galdeano, Michel Diament, Alexei Gvishiani,
Sergei Agayan, Shamil Bogoutdinov, Elena Graeva, Pascal Sailhac

HAL Id: insu-01355003

https://insu.hal.science/insu-01355003

Submitted on 22 Aug 2016

Application of artificial intelligence for Euler solutions clustering

Valentine Mikhailov*, Armand Galdeano‡, Michel Diament‡, Alexei Gvishiani*,
Sergei Agayan*, Shamil′ Bogoutdinov*, Elena Graeva*, and Pascal Sailhac‡

ABSTRACT

Results of Euler deconvolution strongly depend on the selection of viable solutions. Synthetic calculations using multiple causative sources show that Euler solutions cluster in the vicinity of causative bodies even when they do not group densely about the perimeter of the bodies. We have developed a clustering technique to serve as a tool for selecting appropriate solutions.

The clustering technique uses a methodology based on artificial intelligence, and it was originally designed to classify large data sets. It is based on a geometrical approach to study object concentration in a finite metric space of any dimension. The method uses a formal definition of cluster and includes free parameters that search for clusters of given properties.

Tests on synthetic and real data showed that the clustering technique successfully outlines causative bodies more accurately than other methods used to discriminate Euler solutions. In complex field cases, such as the magnetic field in the Gulf of Saint Malo region (Brittany, France), the method provides dense clusters, which more clearly outline possible causative sources. In particular, it allows one to trace offshore the main inland tectonic structures and to study their interrelationships in the Gulf of Saint Malo.

The clusters provide solutions associated with particular bodies, or parts of bodies, allowing the analysis of different clusters of Euler solutions separately. This may allow computation of average parameters for individual causative bodies. Those measurements of the anomalous field that yield clusters also form dense clusters themselves. Application of this clustering technique thus outlines areas where the influence of different causative sources is more prominent. This allows one to focus on these areas for more detailed study, using different window sizes, structural indices, etc.

INTRODUCTION

Euler deconvolution is a well-known method to determine the shape of causative bodies from potential field data. It is based on the approximation of the measured anomalous gravity or magnetic field in a running window by the field of a single elementary source of uniform density or magnetization, characterizing the position and depth of the nearest or largest causative source in the vicinity of the window. Hood (1965) uses this method for aeromagnetic data interpretation and demonstrates that the method is valid for point–pole and point–dipole sources. Thompson (1982) elaborates application of the method to 2D sources and derives structural indices for several elementary bodies. Reid et al. (1990) extends the method to three dimensions and discusses its applicability to gravity anomalies of finite steps and magnetic anomalies of thin dikes and sloping contacts. Keating (1998) applies the Euler approach to irregular 3D grids, using weights proportional to station accuracy and interstation distance. Zhang et al. (2000) use the method to interpret gravity gradient tensor measurements. Their paper also contains a comprehensive bibliography on the development and application of the method.

Euler deconvolution is a powerful method to gain preliminary information on position, shape, and depth of causative bodies from gravity and magnetic fields. It is especially effective for isolated compact bodies restricted by vertical sides. In such cases, Euler solutions cluster around the perimeter of the bodies in a horizontal plane and provide estimates of the depth of causative bodies. When anomalies are caused by multiple bodies (including structural interfaces and compact bodies of different size and depth), the method does

not always provide easily interpretable results: Euler solutions form broad clouds rather than dense clusters, making it difficult to outline side boundaries of causative sources. Often interpretation can be improved by rejecting solutions with low tolerance (Thompson, 1982), with large dispersion of depth estimates, or with depths that are unreasonably shallow or deep. Although effective for isolated anomalies, these criteria sometimes are inefficient in complicated areas because they are affected by noise, shallow sources, or neighboring bodies.

In this paper we present a new technique for selecting the best Euler solutions on the basis of artificial intelligence. Our goal is to automatically (or semiautomatically) determine clusters of Euler solutions that reliably outline the lateral extent of causative bodies and provide more reliable estimates of their depth (the latter also calls for correctly choosing the structural index; see, e.g., Barbosa et al., 1999).

First, we discuss Euler deconvolution and introduce a new method of cluster analysis based on a topological and geometrical approach to object concentration in finite metric space. Then we consider a synthetic example, focusing on the case where nearby bodies cause interfering anomalies. Finally, we apply the approach to magnetic anomalies from the Saint Malo region (Brittany, France) and use it to infer relationships between a Precambrian high-temperature belt and later linear and isometric magmatic intrusions.

THE METHOD OF EULER DECONVOLUTION

Euler deconvolution provides estimates of geometrical parameters for elementary causative bodies from gravity or magnetic anomalies and their horizontal and vertical derivatives (measured or calculated). This method assumes that the anomaly is a homogeneous function of spatial coordinates. By definition, a function is homogeneous of degree n when, for any t,

$$f(tx, ty, tz) = t^n f(x, y, z). \tag{1}$$

In this paper x, y, and z represent Cartesian coordinates with the z−axis directed downward and the x-axis directed to the north. For the 2D case, x is directed along the profile. Here we discuss only the magnetic case, but similar results hold for gravity.

Strictly speaking, the method is valid for magnetic or gravity anomalies caused by bodies with positions characterized by a single point—(x_0, y_0, z_0) in three dimensions or (x_0, z_0) in two. It can be point poles, point dipoles, lines of poles, and lines of dipoles. Several elementary bodies obey the Euler equation under specific conditions: for a dike (vertical or inclined), when its thickness is considerably smaller than its depth; for a finite step, when its step is considerably smaller than its depth; and so on. For all of these bodies the Euler equation can be written in the form

$$(x - x_0)\frac{\partial f}{\partial x} + (y - y_0)\frac{\partial f}{\partial y} + (z - z_0)\frac{\partial f}{\partial z}$$
$$= A - N \cdot f(x, y, z), \tag{2}$$

where x_0, y_0, z_0 are coordinates of the elementary source (the object of the Euler solution); x, y, z are coordinates of the ob-

servation point; N is the structural index, which depends on the shape of the body ($N = -n$) [structural indices for different simple equivalent sources are listed in Thompson (1982) and Reid et al. (1990)]; and A reflects a constant level in a measured field to be determined.

Euler deconvolution consists of determining four unknown parameters—x_0, y_0, z_0, and A—in running windows (of a size more than four field points) by solving a system of linear equations. This system assigns equation (2) to each window point. By solving this system in the mean-square sense, dispersion of unknown parameters can be also obtained.

The structural index is assigned a priori using additional information on the shape of causative bodies. Typically, several indices are used, and the one that provides the best fit to known superficial geological structure, seismic data, boreholes, etc., or the one having good clustering properties is accepted. [On a possibility to estimate structural index, see Slack et al. (1967), Steenland (1968), Barbosa et al. (1999), and Martelet et al. (2001)].

Even for synthetic examples computed for a single elementary body, when the anomalous field and its derivatives contain no errors, not all Euler solutions cluster around the perimeter of the causative body in the (x,y) plane. Standard Euler deconvolution (Thompson, 1982) uses different criteria to select the solutions. A solution can be rejected because of its low tolerance [$z_0/N \cdot \sigma_z < TOL$, where σ_z is the dispersion of the z_0 estimate (Thompson, 1982)], or because its dispersion is higher than a given value σ_{max}, or because it is unreasonably shallow ($z < z_{min}$) or deep ($z > z_{max}$). Our synthetic calculations show that the distance from an Euler solution to the center of the window from which the solution was obtained is also an important characteristic. We suggest rejecting solutions located at a distance several times larger than the window radius. (Depending on the structure of the anomalous field and the size of the window, this ratio ranges between 2 and 10.)

Even though 3D Euler deconvolution is in valid for 2D bodies, the method gives good results for a wide range of elongated bodies. Synthetic calculations for parallel dikes having a width up to 25 times less than their length with the window size equal to or less than the dike width showed that Euler solutions cluster well along the four side boundaries of the dikes. Parameters of this test were chosen similar to the magnetic field of the Saint Malo region, discussed later in this paper. In particular, the difference between the angle of dike strike and declination of magnetization was as small as 5°. When the window size is smaller than the dike width, Euler solutions outline both long sides of the dikes; when the window is wider, solutions cluster along the central line of each dike.

Practical implementation of Euler deconvolution can pose additional problems. In large areas, the size of causative bodies may differ considerably. The window size will then appear to be too small for large bodies (providing small singular values and, as a consequence, large standard deviation of solutions) and too large for small bodies (failing to outline or even discern them). Because of interference of anomalies, Euler solutions do not cluster sharply around perimeters of causative bodies. This interference and high-frequency noise in the data can also hamper the use of criteria based on singular value or dispersion of z_0. Indeed, they increase the gradient of the anomalous field and its derivatives within the window and, as a result, they increase singular values and decrease the dispersion of

parameter estimates. Thus, a solution with good tolerance in fact can be ill posed.

To overcome these difficulties in selecting good Euler solutions, we developed and tested a new technique based on artificial intelligence. The results of our numerical calculations for synthetic examples using multiple bodies showed that even if Euler solutions form broad clouds, the density of the clouds is not uniform and more dense nuclei tend to outline the perimeter of causative bodies. This new method of cluster analysis helps us extract these nuclei.

RODIN—NEW METHOD OF CLUSTER ANALYSIS

The proposed approach to clustering objects follows dynamic pattern recognition ideas discussed by Gvishiani et al. (1995) and Dubois and Gvishiani (1998). The Rodin clustering algorithm has been designed by A. Gvishiani and S. Agayan (personal communication, 2002) to tackle problems of classifying large data sets in finite metric space. The main feature distinguishing this algorithm from the usual clustering methods is that Rodin investigates geometrical and topological features of an object's concentration. Correspondingly, it gives much less weight to analyzing of the object parameters. Therefore, Rodin is relevant for problems where calculated solutions (and not original objects that possess measured parameters) must be clustered. Following this line, here we apply Rodin to cluster synthetic and real calculated Euler solutions.

Rodin is based on a formal definition of cluster, which lets us construct an effective numerical algorithm. To introduce the method, we use an analogy with light propagation. Suppose that we have a lamp at each point of a given set C. To describe the light propagation, we introduce any nonnegative, descending function $\varphi(t)$ that has unit amplitude at the origin [i.e., $\varphi(t) \geq 0$ for all t, $\varphi(0) = 1$ and $\varphi(t_1) < \varphi(t_2)$ when $t_1 > t_2$]. Using this function, we determine the basic function $L(x)$ of the algorithm, which is the total illumination from a set A of any point x. Another important value that Rodin uses is average illumination $l(x)$ (as illumination divided by number of light sources). The latter lets us define cluster.

Definition of a cluster

A subset A in a given set of points C is considered to be a cluster if the average illumination of any point $x \in A$ produced by other points inside A is larger than the average illumination of x produced by points outside A. In other words, cluster A is a subset of C that produces the highest average illumination of any point $x \in A$ among all possible extensions $A + B$, where B is a subset of C having no common points with A. Thus, by adding new elements to A, we decrease the average illumination of point x. It follows from this definition that average illumination of all points in A increases when rejecting points that have the least average illumination, and this is the basis on which Rodin selects clusters A from a given set of points C. The formal detailed description of the algorithm is given in the appendix.

Herein, we search for r-clusters of Euler solutions (also described in the appendix). The idea of r-clustering deals with calculating average illumination from a given point x to other points of cluster A, taking into account only points of

cluster A with a distance from x less than a given radius r. Other clustering algorithms tend to find compact clusters of isometric shape. The r-algorithm, on the other hand, lets us find clusters of complicated irregular shape (e.g., elongate or toroidal with empty center).

The Rodin algorithm has two free parameters: α controls cluster density and their minimal average illumination, and r controls cluster shape. Running calculations for different values of free parameters enables one to find clusters which satisfy a priori information on their properties (e.g., their structure, size, and position) or to find clusters which fit best to geological and geophysical data. Sensitivity to the choice of free parameters is discussed below.

Rodin helps locate dense clusters using not only the position of points in the x,y plane (can be done visually) but also the density of their depth distribution. Another advantage of our clustering approach is that it separates Euler solutions originating from different causative bodies or from their parts (if causative bodies are large enough compared with the window size). Thus, Rodin lets us analyze separate clusters (determining their average depth, for example). Besides, points of an anomalous field that produce particular clusters also often form dense clusters themselves. Such a field cluster outlines an area where the influence of a particular causative body (or of its part if the body is large enough) is more prominent. We illustrate these statements with a synthetic example.
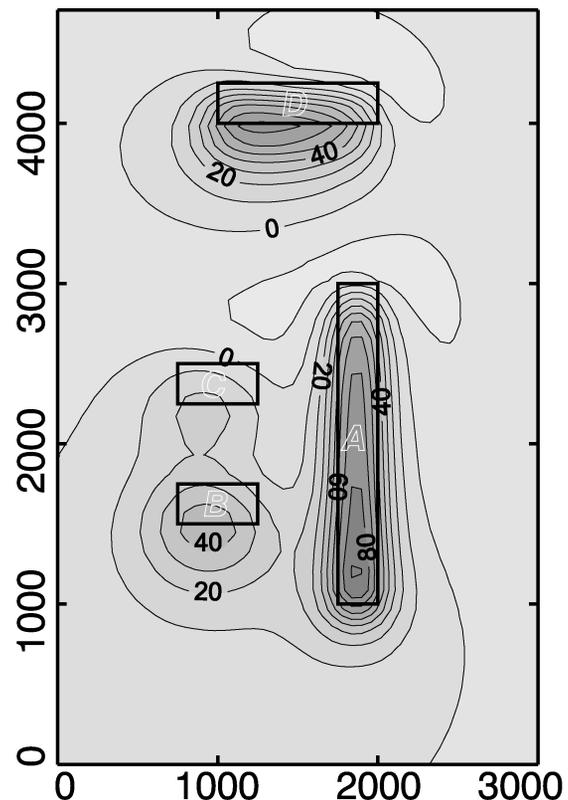

FIG. 1. Total magnetic field anomaly (nT) caused by four rectangular prisms denoted A–D. Prism A possesses magnetization parallel to the earth's field. The other prisms possess different magnetizations as shown in Table 1 (x-axis directed to the north). The distance along the axes is in meters.

We consider in the following synthetic example an anomaly of total magnetic field ΔT caused by four rectangular prisms, positioned in the x,y plane as shown on Figure 1 and with parameters listed in Table 1. The prisms have different sizes, depths to their tops, and magnetizations. The largest prism A has magnetization parallel to earth's field of 0.4 A/m. The other prisms possess magnetizations and inclinations as shown in Table 1. The resulting total-field anomaly (Figure 1) shows that prisms B and C do not produce strong anomalies. They

are smaller and have deeper upper boundaries in comparison with prisms A and D. Figure 1 also shows that anomalies from prisms B and C are disturbed by the anomaly from prism A.

The total field anomaly and its derivatives were computed on a regular grid measuring 6 by 3 km. The number of points along both axes was 50; so sample intervals were 122.4 and 61.2 m, respectively (we used a rectangular grid because it is an ordinary feature of aeromagnetic data). Euler solutions were calculated with running windows, five points on each side, and with a structural index $N = 3$ (point dipole). In comparison with smaller structural indices, a value $N = 3$ usually provides denser

Table 1.   Parameters of the four prisms used for the synthetic example.

Prism	x_0, x_1 (km)	y_0, y_1 (km)	z_0 (m)	$|J|$ (A/m)	$I^{(\circ)}$	$D^{(\circ)}$
A	1.0–3.0	1.75–2.0	150	0.4	70	5
B	1.5–1.75	0.75–1.25	400	1.0	30	55
C	2.25–2.5	0.75–1.25	450	1.0	50	55
D	4.0–4.25	1.0–2.0	200	0.5	50	55

(x_0, x_1)—coordinates along the x-axis to the north
(y_0, y_1)—coordinates along the y-axis to the east
z_0—depth to the top boundary (depth to the bottom for all prisms was 1000 m)
$|J|$—modulus of magnetization vector
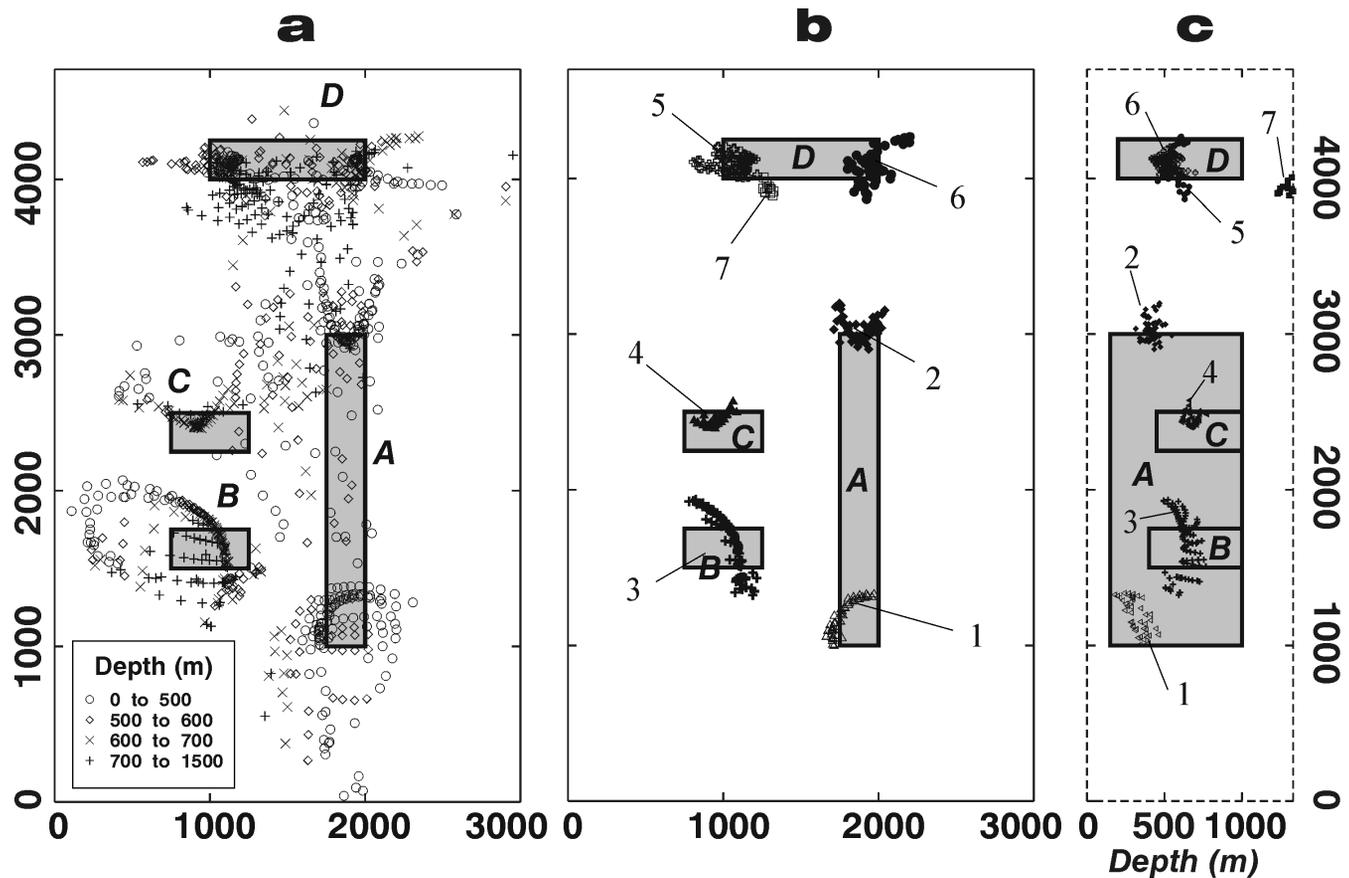I—inclination
D—declination

FIG. 2. Results of Euler deconvolution. (a) Euler solutions having singular values more than 0.01. (b) Results of clustering Euler solutions using the Rodin algorithm. (c) Depth distribution of clustered Euler solutions. To construct Figure 2c, points from Figure 2b were projected on the x,z plane (viewed from the east). Cluster 7 is an artifact.

clusters of points (Fairhead et al., 1994) grouped mainly at the depth between the top boundary and the center of a prism. To demonstrate an advantage of the clustering approach, we used a very complex synthetic example: the distance between body A and bodies B and C is 0.5 km, while the width of bodies B and C is also 0.5 km, a value close to the depth to the top of bodies B (0.4 km) and C (0.45 km). In this particular case calculations with structural index $N < 3$ provide absolutely no interpretable results. By the same reason we did not add any noise for the synthetic magnetic field. The given grid generated $46 \times 46 = 2116$ Euler solutions. This number was reduced to 985 by rejecting solutions with very low minimal singular values. Horizontal locations of these solutions are shown on Figure 2a.

For prism A, having magnetization directed approximately northward ($D = 5°$), Euler solutions cluster around the northern and southern boundaries. Because the horizontal derivative $\partial T/\partial y$ in the central part of body A is close to zero, few solutions have nonzero singular values along the western and eastern boundaries of A. Euler solutions outline the southern boundary of prism A and are shifted to the north from its northern boundary due to the influence of prism D. Shallow points ($z_0 < 600$ m) coincide better with the northern and southern boundaries. If prism D were an isolated body, the calculations with the window size being two times larger than dimension of the prism along the x-direction would provide an elongate cluster aligned along the axis of the prism from west to east. Here, neighboring bodies disturb the true picture. Thus, Euler solutions of prism D are shifted southward. Figure 2a also shows that the distance of shifting varies: robust solutions located close to the eastern and western boundaries are less shifted than weak solutions, which are located far from the axis of the prism. Distribution of the solutions in the vicinity of prisms B and C gives no way to outline these bodies.

Classical criteria for selecting solutions improve the resolution of solutions in the vicinities of southern and northern boundaries of prism A and in the vicinities of western and eastern boundaries of prism D. However, having higher dispersion and lower tolerance, all points around bodies B and C were rejected and, as a result, these bodies were not discerned.

Figure 2b shows the result obtained by applying the r-cluster Rodin algorithm dealing with three coordinates (x_0, y_0, z_0) with $\alpha = 0.02$ and $r = 0.4$ km (see the appendix). Seven clusters have been identified (Figure 2b). Clusters 1 and 2 clearly outline northern and southern boundaries of body A, while clusters 5 and 6 group densely around the eastern and western boundaries of prism D. Cluster 3 is less dense, marking an area larger than prism B. Nevertheless, this cluster also provides some information on the position of prism B. Cluster 4 marks prism C, being shifted slightly to the north from its center. Cluster 7 appears to be an artifact. It exhibits low tolerance in comparison to other clusters; so we reject it, applying convenient criteria or taking into account that every solution of cluster 7 is situated far from the center of the windows which produced these solutions. We made several calculations to find the values of α and r, which provide more dense and localized clusters. Results of clusterization appeared to have little sensitivity to these parameters, being similar when parameter α ranged between 0.018 and 0.022 and parameter r ranged between 0.35 and 0.45 km. It is worth noting that the search of the optimal

values of parameters α and r in the Rodin algorithm can be done quite efficiently because clusterization takes a very short time (about 3–4 s for the synthetic example using a standard personal computer).

Figure 2c shows the depth distribution of the dipoles. To construct this figure, we projected all points on the x,z plane (viewed from the east). The clusters provide correct relative depths for four causative bodies. Clusters 1 and 2 clearly show that prism A is shallower than other prisms. Prism D also appears to be shallower than bodies B and C (clusters 3 and 4, respectively).

As we mentioned above, cluster analysis enables one to separate solutions linked with different bodies. This provides estimates of the average depth of every cluster. Clusters 1 and 2 have an average depth $z_{av} = 320$ and 420 m, which is between the depth to the top of prism A ($z_1 = 150$) and the depth to its center ($z_c = 575$ m). Clusters 5 and 6 show that prism D is deeper than prism A ($z_{av} = 523$ and 545 m when $z_1 = 200$ and $z_c = 600$ m, respectively). The depth of clusters 5 and 6 is noticeably larger because when prism size is comparable to the size of the running window, Euler solutions cluster closer to the center of a prism. For prism B, $z_{av} = 620$, $z_1 = 400$, and $z_c = 700$ m. For prism C, $z_{av} = 674$, $z_1 = 450$, and $z_c = 725$ m. Because the structural index is 3, both estimates are within the
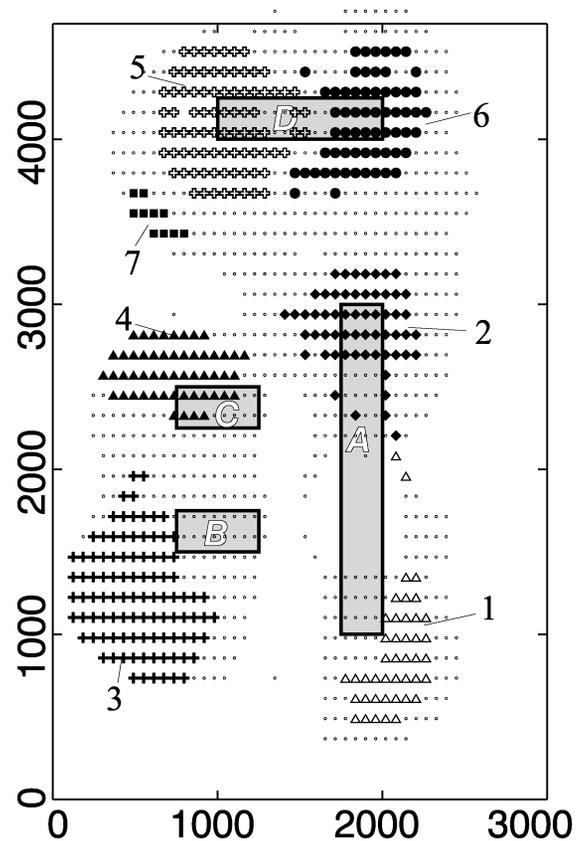

FIG. 3. Areas of an anomalous field for which the influence of a particular body or parts of it (if the body is large enough in comparison with window size) is more prominent. Symbols at grid points indicate which cluster an Euler solution belonged to when the center of the running window was on this grid data point. Data from these areas can be analyzed separately using different window sizes and structural indices.

interval from the top to the center of the prisms, being closer to the center since prisms are small in comparison to the window size. Cluster 7 has an average depth as large as 1280 m; this is another indicator of its poor reliability.

Now we proceed with the question of which data points of the measured field ΔT were actually used in forming each cluster. Figure 3 shows the data points (from the original grid) with different symbols. These symbols indicate to which cluster an Euler solution belongs when the center of the running window was on this grid data point. Remarkably, window centers also form dense clusters. Thus, application of our cluster analysis allows one to outline areas of the measured fields for which the influence of a particular body or parts of it (if the body is large enough in comparison to the window size) is more pronounced and clear. We can speculate that this opens new ways in separating anomalous fields and finding parameters of causative bodies. It allows one to subdivide the original anomalous field into smaller areas that may be specifically studied, e.g., using differ-

ent window sizes or structural indices. However, this statement calls for more detailed investigation.

APPLICATION TO REAL FIELD DATA

We used the method to analyze a portion of the aeromagnetic map of the Armorican massif in France (Galdeano et al., 2001). The study area includes the eastern part of the Saint Malo Gulf and the inland area to the south. Figure 4 shows the main tectonic units of the region (Chantraine et al., 2001): the Guingamp and Saint Malo units. They form a high-temperature belt bounded by the main Cadomian contact (labeled MCC on Figure 4) to the north and by the Cancale–Plouer fault (labeled PC on Figure 4) to the south (Chantraine et al., 1988; Brun and Balé, 1990). According to seismic data (Bitri et al., 1997), MCC and PC faults dip to the north. Thrusting at these faults was accompanied by left lateral movements.
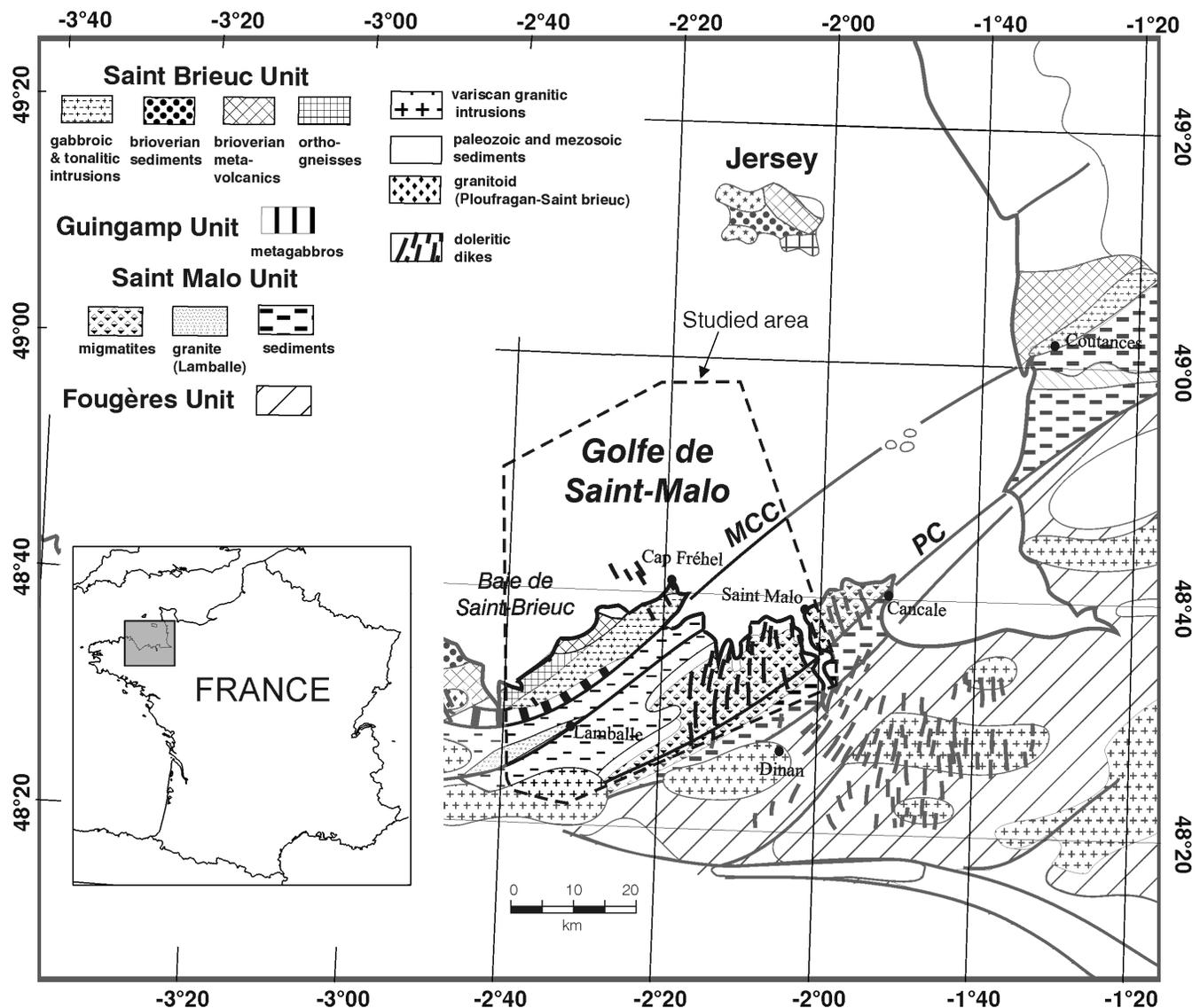

FIG. 4. Tectonic map of the Gulf of Saint Malo region (Chantraine et al., 2000). MCC—Main Cadomian Contact. PC—Cancale–Plouer fault.

The evolution of this region had been mainly controlled by subduction in the process of the Celtic ocean closure. The Guingamp–St. Malo high-temperature belt was formed 600–540 Ma during the Cadomian orogeny, a late Precambrian tectonic event which resulted here in obduction of a back arc basin over a continental margin (Balé and Brun, 1989). Later this belt was moderately affected by Hercynian deformations and metamorphism. The belt is comprised of micaschists, paragneisses, and migmatites. In the Saint Malo unit migmatites occupy the core of an asymmetric metamorphic dome, on which anatectic granites were dated at 541 ± 5 Ma (Brun and Balé, 1990).

A distinctive feature of the geological structure of the region is a swarm of doleritic dikes. The dikes were intruded in Precambrian basement and later were truncated and metamorphosed by Hercynian granites (Vidal, 1980). According to

Lahaye et al. (1995), the Saint Malo dikes are typical continental tholeites. Emplacement of the dikes corresponds to an extensional (pull-apart) tectonic phase. Paleomagnetic and K/Ar data give an age of dikes of 330 ± 10 Ma, i.e., of the Lower Carboniferous age (Perroud et al., 1986). Onshore dikes have a small thickness (5 m on average), and their strike ranges from north–south to north-northeast–south-southwest. Linear magnetic anomalies in this area are considerably wider, probably indicating that dikes fan from wider, deeper magmatic bodies.

Aeromagnetic measurements were carried out at a flight altitude of 350 m. Flight lines were oriented 48°N with 500 m line spacing, which let us interpolate data on a north–south grid with step 250×250 m in both directions (Galdeano et al., 2001). This aeromagnetic map demonstrates a good correlation
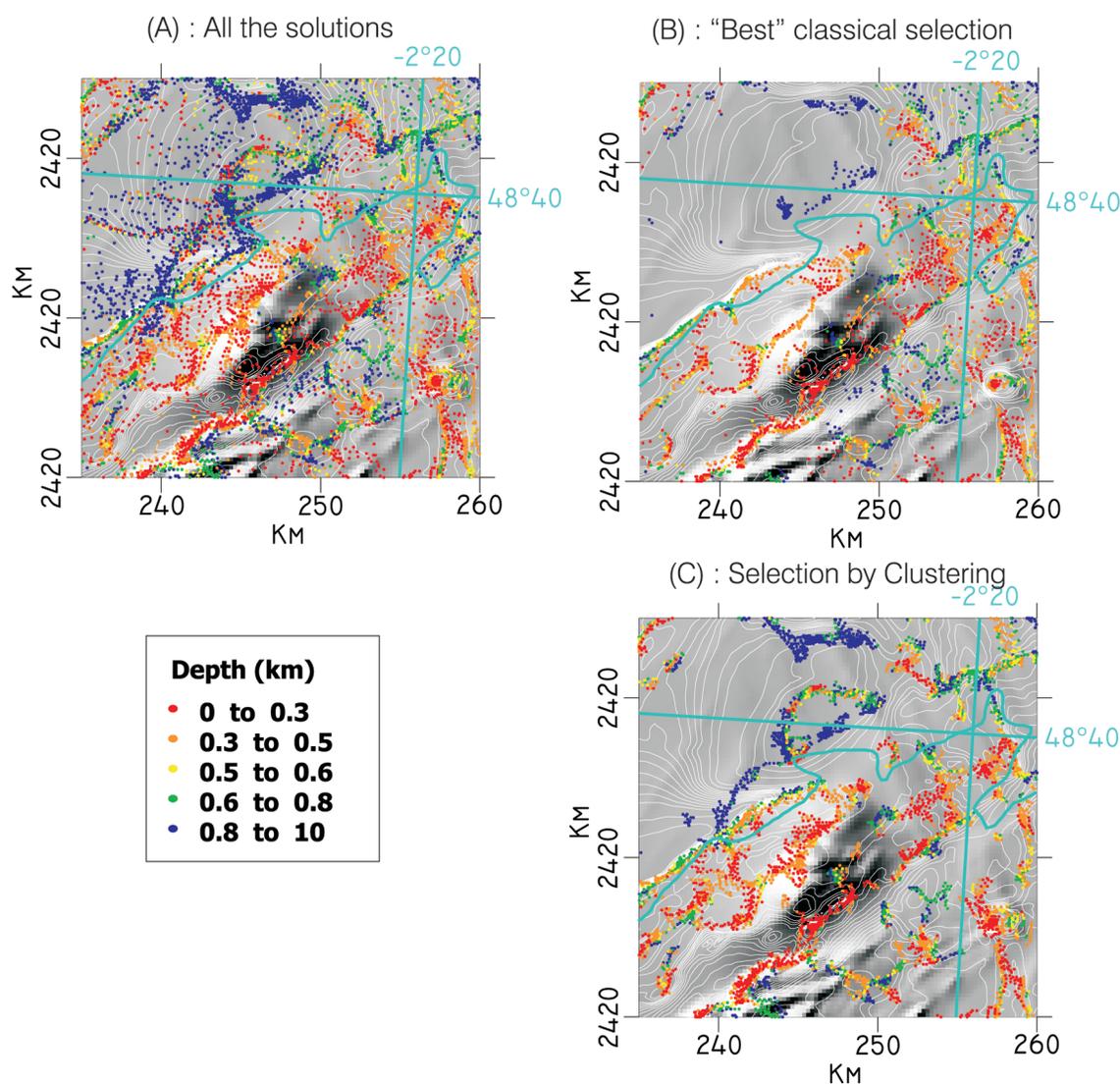

FIG. 5. Results of Euler deconvolution for the central part of the study area (its position in the study area is shown by a box on Figure 6). Shaded relief and isolines show amplitude of magnetic anomaly. (a) All points used in the analysis. (b) Results of selection using standard criteria (Thompson, 1982). (c) Results of clustering Euler solutions. Note that southwest–northeast linear clusters in the bottom right side of the figure are better established on this plot. Clusters on this plot also better outline possible sources of isometric anomaly at the top of the map, showing that depth of sources increases to the north.

with known geological structures. The long-wavelength anomalies outline the arcuate structure of the Cadomian belt. Short-wavelength anomalies in the Saint Malo region are mainly associated with dikes exposed onshore and extending offshore. In the Saint Malo region, magnetic anomalies are oriented north–south, whereas in the Gulf of Saint Malo their strike changes to northwest–southeast. These anomalies exhibit quasi-linear structure. Their relationship with the Guingamp–Saint Malo belt is unclear, since the structure of the magnetic anomalies close to their intersection is complicated. The main purpose of the application of Euler deconvolution and cluster analysis was to investigate the structure of quasi-linear magnetic anomalies onshore and offshore and their relationship with the Guingamp–Saint Malo high-temperature belt.

Figure 5a shows the total magnetic anomaly for a small part of the region by shaded relief and isolines, as well as Euler solutions obtained with window size 9×9 points (2×2 km) and structural index $N = 3$. About 40% of Euler solutions were rejected because of their extremely small singular values. Obviously, it is too difficult to use these results to analyze the geological structure of the region.

We first selected Euler solutions using different standard criteria. Solutions with low tolerance, situated at depths >2 km and located at a distance five times larger than the window size from the center of window, were rejected. Results are shown in Figure 5b. In comparison with Figure 4, the Euler solutions now outline isometric and linear bodies. However, the position of possible causative sources remains unclear.

Figure 5c shows the result of clustering. The Rodin algorithm was applied to the original set of Euler solutions using $\alpha = 0.8$ and $r = 0.3$ km (see the appendix). The result is stable in the sense of possible changes of these parameters, i.e., close pictures were obtained for α ranging from 0.78 to 0.82 and for r ranging from 0.25 to 0.35 km. The algorithm found dense clusters that more clearly outline possible causative sources. Indeed, an isometric cluster is evident in the northern part of the map, as are linear clusters stretching from southwest to northeast in the lower half of the map, marking the Cadomian high-temperature belt.

Figure 6 presents the results of clustering of Euler solutions for the Saint Malo region with the same parameters as Figure 5c. The initial set of Euler solutions included 34 500 points. The clustering algorithm rejected almost 7000 points, finding dense clusters that outlined isometric and linear structures of the region. In particular, isometric clusters outlined the lateral extent of possible causative bodies in the central parts of isometric anomalies marked on Figure 6 by numbers 1, 2, and 3. The linear clusters extending southwest–northeast to the south of body 1 can be considered the eastward continuation of the Main Cadomian fault, since tracing the eastward prolongation of the Cancale–Plouer fault is problematic.

A distinctive feature of the Euler solutions in Figure 6 is the linear clusters lineated north–south in the southern part of the map and northwest–southeast in its northern part. Inland, they coincide with the strike of doleritic dikes. Clustering solutions in the vicinity of the Main Cadomian fault show that besides north–south trending dikes, there is probably another dike swarm striking approximately northwest–southeast. Using Euler solutions, the later dikes can be followed across the fault, and further they can be correlated with an offshore dike swarm of northwest–southeast direction. At the intersection with the Main Cadomian fault, the structure of the possible causative sources is unclear, even though linear clusters crossing the fault show that the dikes are younger than this thrust zone. Less regular linear structure of clusters in the vicinity of the fault can indicate a rejuvenation of the Main Cadomian contact thrust zone after doleritic dyke emplacement. Most of the Euler solutions to the south of the fault show shallow depth of the possible causative sources. Depth of the solutions in the northwest–southeast linear clusters to the north of the main Cadomian contact is not uniform: in the central part of the area solutions are situated at shallow depth, becoming deeper when moving southwest to body 1 and northeast to bodies 2 and 3.

A more detailed analysis of the results of our application to the data in the Saint Malo region is beyond the scope of this paper and is the subject of a separate one. Comparison of Figure 5b and Figure 5c shows that on real data our clustering technique was more efficient than standard approaches for selecting Euler solutions.

CONCLUSION

Results of Euler deconvolution strongly depend on the quality of the solution selection. A clustering technique based on topological analysis of solution concentration can efficiently extend methods routinely used for this purpose. Our synthetic calculations based on multiple causative sources showed that, even where Euler solutions do not group densely around the lateral extent of the bodies, the density of their distribution appears to be higher in the vicinity of the causative bodies, suggesting that clustering techniques hold promise.

We applied a new method of clustering called Rodin to select dense clusters of Euler solutions. The method is based on a formal definition of cluster, which enables one to construct an effective clustering algorithm. Thus, we proceed from informal to formal selection of Euler solutions. This is an important step toward automating the selection procedure.

On synthetic and real data, Euler solutions from our clustering technique appeared to successfully outline causative bodies. An advantage of the clustering method application is that clusters thus determined provide solutions associated with particular bodies or with their parts, making it possible to analyze different clusters of Euler solutions separately. This may allow, for example, computing average parameters for individual causative bodies. Another advantage is that data points of anomalous fields responsible for different clusters also form dense clusters themselves. Thus, the clustering technique enables one to outline areas where the influence of different causative sources is more prominent. This opens the possibility for more detailed interpretation of the data in areas using different window size and structural indices.
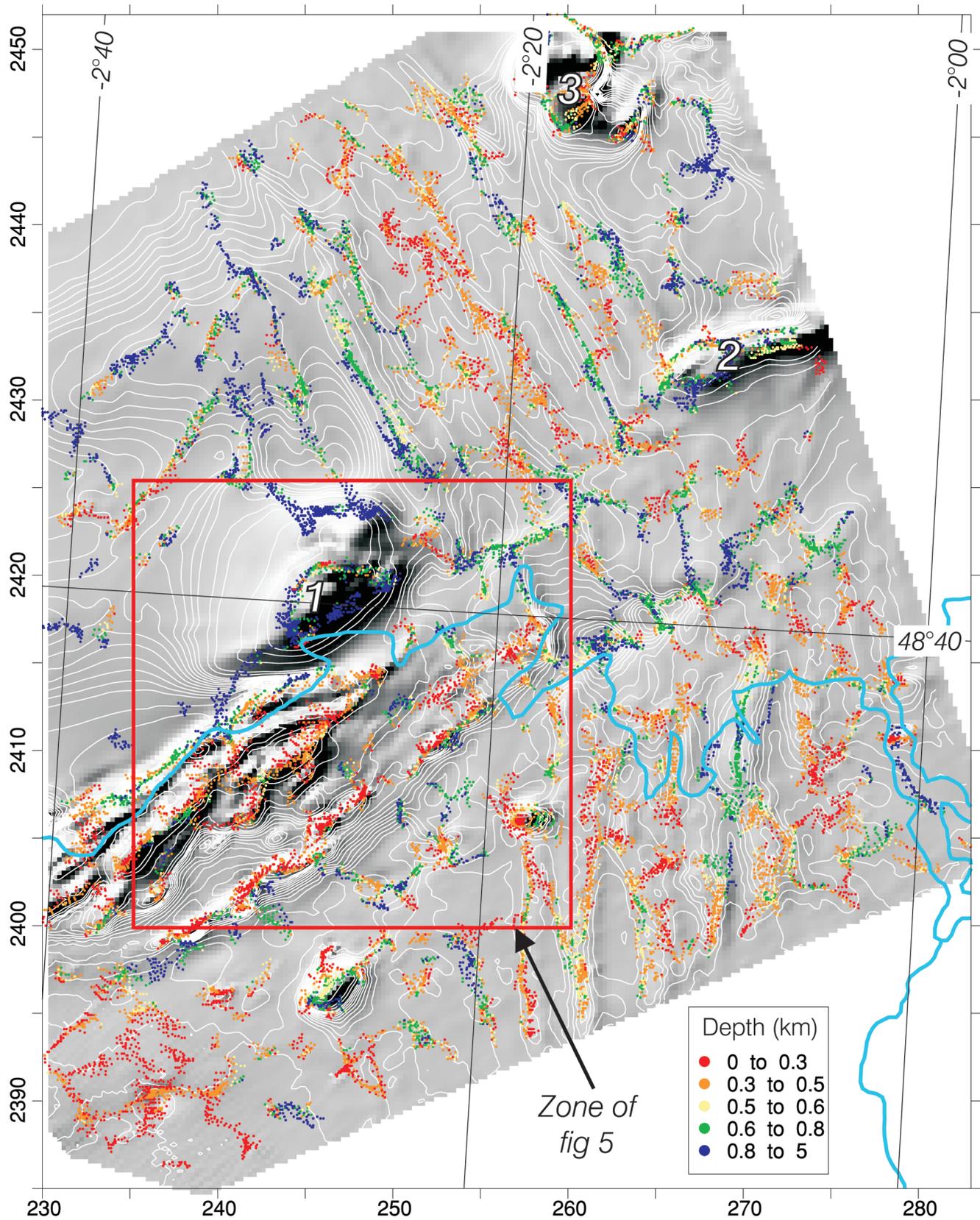
FIG. 6. Results of clustering Euler solutions for the Gulf of Saint Malo area. The rectangular box marks the position of the test area used to estimate parameters (Figure 5). Isometric anomalies marked by numbers 1, 2, 3 are clearly outlined by clusters of Euler solutions. Clusters striking southwest–northeast to the south of body 1 show a possible extension of the Main Cadomian contact. Linear clusters striking north–south in the southern part of the map and northwest–southeast in its central and northern parts were interpreted as swarms of doleritic dikes of different age. Shaded relief and isolines show the amplitude of magnetic anomalies.

who carefully read the manuscript and made many valuable suggestions. The very detailed and useful reviews of V. Barbosa, P. Keating, and associate editor A. Saad are also acknowledged. This paper is IPGP contribution N 1768.

REFERENCES

Balé, P., and Brun, J-P., 1989, Late Precambrian thrust and wrench zones in northern Brittany (France): J. Struc. Geol., **11**, 391–405.

Barbosa, V. C. F., Silva, J. B. C., and Medeiros, W. E., 1999, Stability analysis and improvement of structural index estimation in Euler deconvolution: Geophysics, **64**, 48–60.

Bitri, A., Brun, J-P., Chantraine, J., Guennoc, P., Marquis, G., Marthelot, J.-M., Pivot, F., and Truffert, C., 1997, Structure crustale du bloc cadomien de Bretagne Nord (France): Séismique réflexion verticale et sondage magnétotelurique (project Géofrance 3D-Armor): C. R. Acad. Sci., Paris, **325**, 171–177.

Brun, J. P., and Balé, P., 1990, Cadomian tectonics in northern Brittany: Geol. Soc. Spec. Publ., **51**, 95–114.

Chantraine, J., Chauvel, J.-J., Bale, P., Denis, E., and Rabu, D., 1988, Le Briovérien (Protérozoïque supérieur à terminal) et l'orogenése cadomienne en Bretagne (France): Bull. Géol. Soc. France, **4**, 815–829.

Chantraine, J., Egal, E., Thiéblemont, D., Le Goff, E., Guerrot, C., Ballère, M., and Guennoc, P., 2001, The Cadomian active margin (North Armorican massif, France): A segment of the North Atlantic Panafrican belt: Tectonophysics, **331**, 1–18.

Dubois, J. O., and Gvishiani, A. D., 1998, Dynamic systems and dynamic classification problems in geophysical applications: Springer-Verlag.

Galdeano, A., Asfirane, F., Truffert, C., Egal, E., and Debeglia, N., 2001, The aeromagnetic map of the French Cadomian belt: Tectonophysics, **331**, 99–108.

Fairhead, J. D., Bennett, K. J., Gordon, R. N., and Huang, D., 1994, Euler: Beyond the "Black Box": 64th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 422–424.

Gvishiani, A. G., Zhizhin, M. N., Mikoyan, A. N., Bonnin, J., and Mohammadioun, B., 1995, Syntactic analysis of waveforms from the world-wide strong motion database, *in* Elnashai, A., Ed., European seismic design practice: Balkema Publ. 557–564.

Hood, P., 1965, Gradient measurements in aeromagnetic surveying: Geophysics, **30**, 891–902.

Keating, P. B., 1998, Weighted Euler deconvolution of gravity data: Geophysics, **63**, 1595–1603.

Kirillov, A. A., and Gvishiani, A. G., 1983, Theorems and problems of functional analysis: Springer-Verlag.

Lahaye, Y., Blais, S., Auvray, B., and Ruffet, G., 1995, Le volcanisme fissural paléozoïque du domaine nord-armoricain: Bull. Géol. Soc. France, **166**, 601–612.

Martelet, G., Sailhac, P., Moreau, F., and Diament, M., 2001, Characterization of geological boundaries using 1-D wavelet transform on gravity data: Theory and application to the Himalayas: Geophysics, **66**, 1116–1129.

Perroud, H., Auvray, B., Bonhommet, N., Macé, J., and van Voo, R., 1986, Paleomagnetizm and K-Ar dating of Lower Carboniferous dolerites dykes from northern Brittany: Geophys. J. Roy. Astr. Soc., **87**, 143–153.

Reid, A. B., Allsop, J. M., Grancer, H., Millett, A. J., and Somerton, I. W., 1990, Magnetic interpretation in three dimensions using Euler deconvolution: Geophysics, **55**, 80–91.

Slack, H. A., Lynch, V. M., and Langan, L., 1967, The geomagnetic gradiometer: Geophysics, **32**, 877–892.

Steenland, N. C., 1968, Discussion on: "The geomagnetic gradiometer" by Slack, H. A., Lynch, V. M., Langan, L., 1967: Geophysics, **32**, 877–892: Geophysics, **33**, 681–683.

Thompson, D. T., 1982, EULDPH: A new technique for making computer-assisted depth estimates from magnetic data: Geophysics, **47**, 31–37.

Vidal, P., 1980, L'évolution polyorogénique du Massif Armoricain: Rapport de la géochronologie et de la géochimie isotopique du strontium: Mém. Soc. Géol. Miner. Bretagne, **21**.

Zhang, C., Mushayandebvu, M. F., Reid, A. B., Fairhead, J. D., and Odegard, M. E., 2000, Euler deconvolution of gravity tensor gradient data: Geophysics, **65**, 512–520.

APPENDIX

DESCRIPTION OF THE RODIN ALGORITHM

To describe the method we use a simple physical analogy with light propagation. Suppose that in every point of a metric space X there is a source emitting light. (Below we follow basic notations of functional analysis listed in Table A-1, e.g., Kirillov and Gvishiani, 1983). We call by metric space a pair (X, d), where X is a set and d is a function, such as $d(x, y) = d(y, x)$; $d(x, y) \geq 0$; $d(x, y) = 0$ if and only if $x = y$; and $d(x, z) + d(z, y) \geq d(x, y)$ for $\forall x, y, z \subset X$. To describe the decrease of the light intensity away from the source, we can choose any arbitrary descending on $[0, +\infty)$ nonnegative functions $\varphi(t)$ of the unit amplitude [i.e., $\varphi(t_1) > \varphi(t_2) \geq 0$ for all $t_1 < t_2$, $\varphi(0) = 1$].

To describe how a given point is illuminated by all other points, we define the illumination of a point x by a point y as a function $\delta_x(y) = \varphi[d(x, y)]$. Here, $d(x, y)$ is the distance between the two points. In a metric space, by reciprocity $\delta_x(y) = \delta_y(x)$.

Two natural and often used examples of the functions $\varphi(t)$ and $\delta_x(y)$ are

$$\varphi(t) = \frac{1}{1+t} \quad \text{and} \quad \delta_x(y) = \frac{1}{1 + d(x, y)} \quad \text{(A-1)}$$

$$\varphi(t) = e^{-t/r} \quad \text{and} \quad \delta_x(y) = e^{-d(x,y)/r}. \quad \text{(A-2)}$$

In our method, we use this illumination as the major characteristic of a given point $x \in X$: the most highly illuminated points will be more active and important in our analysis, while poorly illuminated points will not be active, being isolated in X.

The method

Now we can introduce illumination and averaged illumination of a point x by a set of light sources A. We define A_x as a set A without element x if it belongs to A, i.e.,

$$A_x = \begin{cases} A & \text{if } x \notin A \\ A - x & \text{if } x \in A \end{cases}. \quad \text{(A-3)}$$

Definition 1.—We define the illumination of the point x by the set A as $L_A(x) = \sum_{y \in A_x} \delta_y(x)$.

Definition 2.—We define the averaged illumination of the point x by the set A as $l_A(x) = L_A(x)/|A_x|$. In the case when A contains only one element x, $l_A(x) = 0$.

It is clear from the definition that by adding elements to the set A, we enhance the illumination $L_A(x)$ (additional elements give additional light). But this is not necessarily true for the averaged illumination, because it is not directly dependent of number of elements $|A_x|$. In fact, the important property of the averaged illumination is that it is quasi-linear. In other words, if $A \subseteq X$ and $B \subseteq X$, then for all $x \in X$ the following equation

is valid:

$$l_{A \cup B}(x) = \frac{L_{A \cup B}(x)}{|A \cup B|_x} = \frac{|A|_x}{|A \cup B|_x} l_A(x) + \frac{|B|_x}{|A \cup B|_x} l_B(x)$$
$$- \frac{|A \cap B|_x}{|A \cup B|_x} l_{A \cap B}(x). \qquad (A\text{-}4)$$

In particular, if $A \cap B = \emptyset$, then

$$l_{A+B}(x) = \frac{|A|_x}{|A|_x + |B|_x} l_A(x) + \frac{|B|_x}{|A|_x + |B|_x} l_B(x). \quad (A\text{-}5)$$

In addition, when $x \in A$,

$$l_{A+B}(x) = \frac{|A| - 1}{|A| + |B| - 1} l_A(x) + \frac{|B|}{|A| + |B| - 1} l_B(x). \qquad (A\text{-}6)$$

Equation (A-6) provides the relation between the averaged illumination in the case when $x \in A \subset C \subseteq X$:

$$l_C(x) = \frac{|A| - 1}{|C| - 1} l_A(x) + \frac{|C| - |A|}{|C| - 1} l_{C-A}(x). \qquad (A\text{-}7)$$

It follows from equation (A-7) that $l_C(x)$ is always situated on the segment between $l_A(x)$ and $l_{C-A}(x)$ because $\min(l_A(x), l_{C-A}(x)) \le l_C(x) \le \max(l_A(x), l_{C-A}(x))$. Equation (A-7) provides the necessary and sufficient condition for an inequality $l_C(x) > l_A(x)$ or $l_C(x) < l_A(x)$.

Statement 1.—Let $x \in A \subset C \subseteq X$; then $l_C(x) > l_A(x)$ when $l_{C-A}(x) > l_A(x)$ and $l_C(x) < l_A(x)$ when $l_{C-A}(x) < l_A(x)$.

Thus, an inequality $l_C(x) > l_A(x)$ means that the set C contains points that are closely located to x and do not belong to A(Figure A-1).

Definition of a cluster

If the condition $l_C(x) < l_A(x)$ is satisfied for all subsets $C \subseteq X$ that contain A ($\forall C \supset A$) [i.e., if $l_A(x) \ge l_{A+B}(x)$ for $\forall B$, $B \cap A = \emptyset$], then the averaged illumination of any point $x \in A$ from the subset A is not less than its averaged illumination from the

Table A-1.   Symbols and definitions.

Symbol	Definition
$x \in A$	Element x belongs to set A
$x \notin A$	Element x does not belong to set A
$A \subset B$	Subset A is included in set B but is not equal to set B
$A \subseteq B$	Subset A is included or equal to set B
$\emptyset$	Empty set
$A \cap B$	Intersection of sets A and B is equal to the set of elements, which belong to both sets A and B at the same time
$A \cup B$ or $A + B$	Union of sets A and B is equal to the set of elements, which belong to at least one of the sets A or B
$|A|$	A number of elements in set A
$\forall a$	For every element the same condition takes place
$\arg \min_{x \in A} s(x)$	Equal to those values of argument y at which function $s(x)$ reaches its minimal value $\forall x \in A$

whole set C. In other words, there are no elements close to x in the subset $C - A$.

Definition 3.—A subset $A \subseteq X$ is a cluster in X if

$$\forall x \in A, \ \forall B \subseteq X, \ B \cap A = \emptyset \ \Rightarrow \ l_A(x) \ge l_{A+B}(x).$$
$$(A\text{-}8)$$

In other words, A is a cluster if A is a subset of higher illumination than any of its possible extensions $A + B$.

If subset B contains only one element $B = \{y\}$, then taking into account the quasi-linearity of the averaged illumination [equation (A-6)], we arrive at the following inequality that is equivalent to equation (A-8):

$$\forall x \in A, \ \forall y \in X - A \qquad l_A(x) \ge \delta_y(x). \qquad (A\text{-}9)$$

Inequality (A-9) allows us to describe a cluster as a subset of X where any point has greater inner averaged illumination than the illumination from any outer source coming to this point. It is expressed by the following formula:

$$\forall x \in A, \ \forall y \in X - A \ \Rightarrow \ \frac{1}{|A| - 1} \sum_{z \in A_x} \delta_z(x) \ge \delta_y(x).$$
$$(A\text{-}10)$$

By definition, a cluster contains at least two points.

To construct the algorithm we introduce the notion of a cluster quality, defined as definition 4.

Definition 4.—Let A be a cluster in X and $x \in A \subset C$. We define the quality of the cluster A as

$$N(A) = \min_{x \in A} l_A(x). \qquad (A\text{-}11)$$

Let's suppose that $x \in A$, $|A| > 2$ and $\varphi(d(x, y))$ is a potential function (for example, equation (A-1) or (A-2)). Then the following two statements are true.

Statement 2.—$l_A(x)$ is a uniformly continuous function on the set A.

Statement 3.—If $x \in A$, then $\forall y \in A_x$; $l_A(y) < l_{A_x}(y)$ when $\delta_x(y) < l_A(y)$; and $l_A(y) > l_{A_x}(y)$ when $\delta_x(y) > l_A(y)$. These statements are true for any descending nonnegative functions $\varphi(d(x, y))$. In particular, statement 3 can be easily verified for functions given by equations (A-1) and (A-2). Also, from statement 3, it follows that the averaged illumination of all
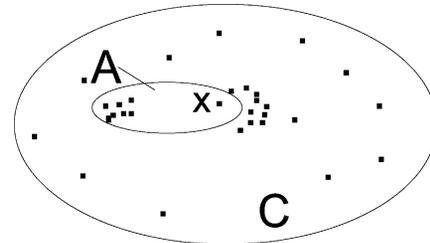

FIG. A-1. The definition of cluster. A is not cluster because $l_{C-A}(X) > l_A(X)$.

points in A [and, as a consequence, the quality $N(A)$] increases when the points with the least averaged illumination are removed from A.

Description of the algorithm

Statements 2 and 3 provide a constructive algorithm for rejecting points that do not obey the definition of cluster. Thus, we can remove from a cluster all the excess points just as a sculptor cuts off the excess parts from a stone to create a sculpture. That is why we named our algorithm after Rodin.

The flowchart of the Rodin algorithm is shown in Figure A-2, where A_k denotes the current version of the searched cluster A.

r-clusters

The basic algorithm described above has several extensions that make it possible to find clusters with some special properties. For example, to find clusters with some sophisticated shape (linear, for example) we introduce the concept of r-clusters.

Let us consider a subset A and the function φ given by formula (A-1). We define the subset $S_A(x, r) = \{y \in A,\, d(x, y) \le r\}$ and define the local r-averaged illumination of a point $x \in X$ as $l_{A,r}(x) = l_{S_A(x,r)}(x)$. For a given subset A and a point x, the function $l_{A,r}(x)$ decreases when r increases. Therefore, if A is a cluster in X, $A \subset B \subseteq X$, and $x \in A$, then $l_{A,r}(x) \ge l_{B,r}(x)$. It is easy to verify that the latter statement is valid for a function φ given by formula (A-1). The former statement enables us to introduce definitions of a local r-cluster.

Definition 5.—$A \subseteq X$ is a local r-cluster if

1) $l_{A,r}(x) \ge l_{B,r}(x)\ \forall x \in A$ and $\forall B \supset A$ and
2) $\forall x \in A\quad S_A(x, r)$ is a cluster in $S_X(x, r)$.

The algorithm of the r-cluster Rodin can be obtained from the algorithm by replacing the local average illumination by the r-local one. The results presented in this paper were obtained using the r-cluster Rodin algorithm with the function φ given by equation (A-1).
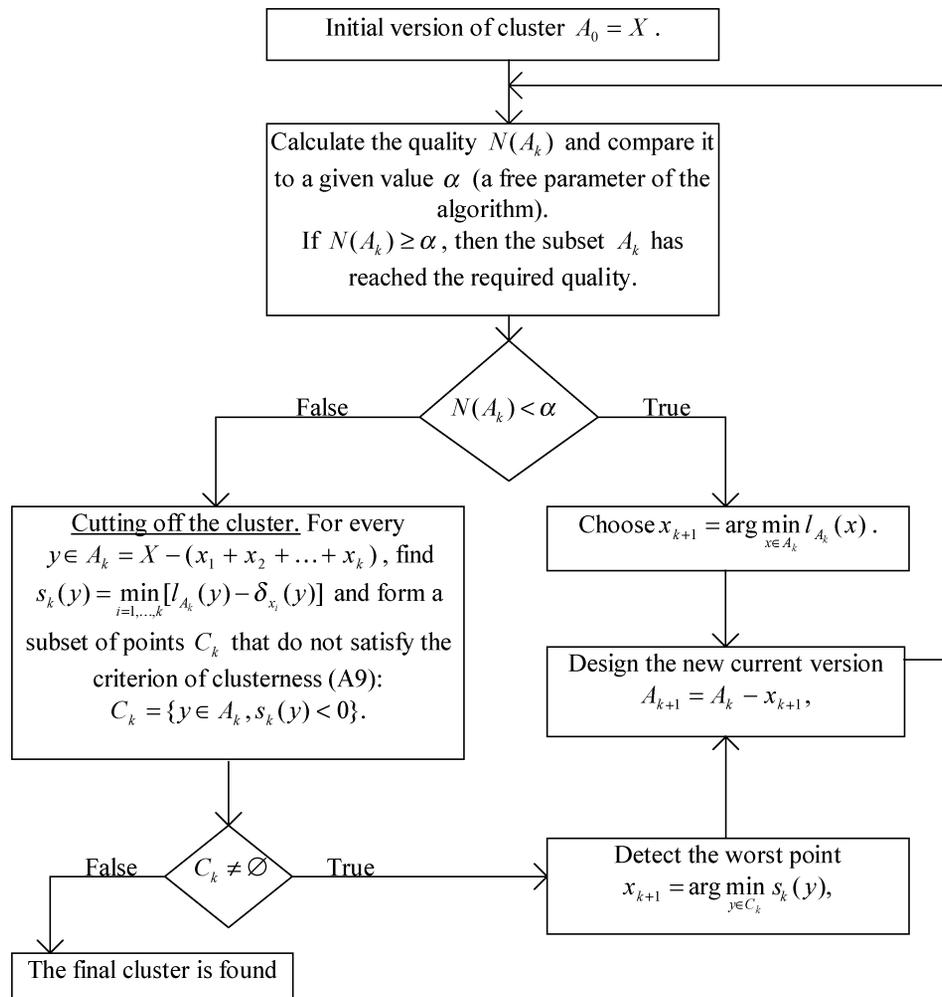

FIG. A-2. Flow diagram of the Rodin algorithm.

Choice of free parameters

As are most pattern recognition algorithms, Rodin is an automated (but not an automatic) procedure. Therefore, the choice of its free parameters (α and r) must be made by the algorithm's users. However, there are different systems of necessary conditions, following which concrete values of free parameters can be established. In this study we required clustering stability. Indeed, for the chosen values α and r in synthetic and real examples (Figures 2, 5, and 6), obtained pictures of clustering remain stable for parameter variations in the range of $\pm 10\%$.