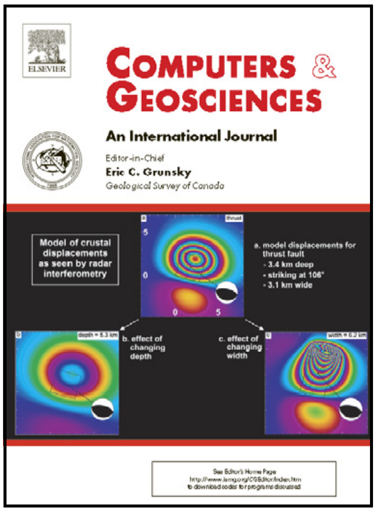
Author's Accepted Manuscript



www.elsevier.com/locate/cageo

Orientation domains: A mobile grid clustering algorithm with spherical corrections

Joana Mencos, Oscar Gratacós, Mercè Farré, Joan Escalante, Pau Arbués, Josep Anton Muñoz

PII: S0098-3004(12)00223-3
DOI: http://dx.doi.org/10.1016/j.cageo.2012.06.021
Reference: CAGEO2953

To appear in: *Computers & Geosciences*

Received date: 4 October 2011
Revised date: 25 June 2012
Accepted date: 26 June 2012

Cite this article as: Joana Mencos, Oscar Gratacós, Mercè Farré, Joan Escalante, Pau Arbués and Josep Anton Muñoz, Orientation domains: A mobile grid clustering algorithm with spherical corrections, *Computers & Geosciences,* http://dx.doi.org/10.1016/j.cageo.2012.06.021

Orientation domains: A mobile grid clustering algorithm with spherical corrections

Joana Mencos[1*]; Oscar Gratacós[1]; Mercè Farré[2]; Joan Escalante[2]; Pau Arbués[1]; Josep Anton Muñoz[1]

[*]corresponding author. GEOMODELS Research Institute, Department of Geodynamics and Geophysics, Universitat de Barcelona, Martí i Franquès s/n, 08028 Barcelona, Spain, phone number: +34 934 021 373, fax number: +34 934 021 340, *e-mail address*: jmencos@ub.edu (Joana Mencos)

[1] GEOMODELS Research Institute, Department of Geodynamics and Geophysics, University of Barcelona, Martí i Franquès s/n, 08028 Barcelona, Spain

[2] Grup de Recerca en Aplicacions i Models Matemàtics – GRAMM, Department of Mathematics, Faculty of Science, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain

Abstract

An algorithm has been designed and tested which was devised as a tool assisting the analysis of geological structures solely from orientation data. More specifically, the algorithm was intended for the analysis of geological structures that can be approached as planar and piecewise features, like many folded strata. Input orientation data is expressed as pairs of angles (azimuth and dip). The algorithm starts by considering the data in Cartesian coordinates. This is followed by a search for an initial clustering solution, which is achieved by comparing the results output from the systematic shift of a regular rigid grid over the data. This initial solution is *optimal* (achieves minimum square error) once the grid size and the shift increment are fixed. Finally, the algorithm corrects for the variable spread that is generally expected from the data type using a reshaped non-rigid grid. The algorithm is *size-oriented*, which implies the application of conditions over cluster size through all the process in contrast to density-oriented algorithms, also widely used when dealing with spatial data. Results are derived in few seconds and, when tested over synthetic examples, they were found to be consistent and reliable. This makes the algorithm a valuable alternative to the time-consuming traditional approaches available to geologists.

Highlights

- Structural data (azimuth/dip) classification into Orientation Domains
- Development of a grid-based proximity-oriented algorithm with square-error criterion
- Automatic cluster partition
- Corrections designed to improve clustering results and maintain geologic criterion

Keywords:

Structural analysis, bedding orientation, size-oriented clustering algorithm, shifting grid, square error criterion.

1. Introduction

1.1 Structural analysis from orientation data

The representation of the geometry of geological structures is central to several resource and environmental applications of geology, like the characterization of hydrocarbon traps, aquifers, waste disposal and CO_2 repositories. In this scenario, an accurate reconstruction is fundamental to reproduce the structures under study, and this can only be achieved following a 3D approach (Zanchi et al., 2009 and references within the same volume).

Orientation data analysis is a recognized fundamental step on the reconstruction of geological structures (Ramsay, 1967; Suppe, 1985; Groshong, 2006). One of the objectives of the analysis of orientation data is the discrimination of clusters. Clusters are data subsets that represent portions of the structure having a characteristic orientation (i.e. dip domains, Gill, 1953; Suppe, 1983; Fernández, 2004; Groshong, 2006). This approach can be useful where geological structures, and specially folded strata, can be represented as planar and piecewise features (Shaw et al., 2005; Groshong, 2006) and helpful in areas where structures are complex and/or under-sampled (Wise, 1992; Torrente, M. 2000; Fernández et al. 2004; Carrera et al 2009). Typically, clusters can be derived by following a semi manual approach (Cruden and Charlesworth, 1972; Fernández, 2004; Mencos, 2011) which summarizes as follows. The analyst selects a subgroup of orientation data from the entire data set, which is regarded as a candidate for cluster. This selection is based on spatial and geological considerations. This is followed by the inspection of scatter and density stereoplots and the study of the relationships between the eigenvectors and eigenvalues that result from the Principal Component Analysis. Stereographic projections have been traditionally used in geology to represent and analyze different types of data sets (i.e. lines or planes) at the same time and without considering their geographical position.

Two conditions need to be satisfied for the candidate to be accepted as a cluster:

a) In order to compare the eigenvalues in an objective way, Woodcock (1977) introduced a criterion, which is also referred to in Fernández (2005). The condition for a set of poles to perform the same cluster is that their first ordered eigenvalue (λ_1) is high enough in comparison with the second (λ_2) and the third (λ_3) ordered ones. This relationship can be expressed as:

$$\lambda_1 >> \lambda_2 > \approx \lambda_3 \approx 0 \qquad (1)$$

b) The second condition to be satisfied is that the subgroup of poles must lie within a range of orientations, such that:

$$u_{max}\text{-}u_{min} \leq u_0 \text{ and } v_{max}\text{-}v_{min} \leq v_0 \qquad (2)$$

where u_{max}-u_{min} denotes the range in azimuth of the data subset, v_{max}-v_{min} is the range in dip within the data subset, and [u_0,v_0] is a range or threshold defined by the analyst, which accounts for the variability that can be expected within orientation domains. This variability reflects instrumental error, geological roughness (e.g. lithology, bedding, texture, etc.) and sharpness (e.g. quality of exposure) of the measured feature (Cruden and Charlesworth, 1976).

If both conditions are met, then the cluster characterizes an orientation domain and can be subsequently enlarged with other measurements. If not, other subgroups have to be tested. In this way, clusters characteristic of different orientation domains are retrieved by a trial and error process driven by expertise. At the end of the process, a set of mean azimuth and dip values (from now on referred to as centroids) representative of each planar orientation domain are obtained.

This approach can yield different results depending upon the analyst expertise. It is also time-consuming, since it requires continued supervision and generally involves working with several types of software (for example CAD, database management systems, structural analysis and structural modelling applications, etc.).

Aimed to overcome these problems, the algorithm herein provides with a fast analysis tool reading from simple ASCII text files. Interaction with user is restricted to initial input parameters and the process remains essentially unsupervised. Output stands also simple, with initial orientation data grouped as clusters representing planar orientation domains. It is important to note that the algorithm does not consider the spatial distribution of the data, hence the results do not represent structural domains.

Subsequent analysis leading to structural domains can be achieved by loading the output data on a georeferenced 3-D visualization software. This is possible as the output data preserves their original XYZ location. Then, the user needs to manually select the structural domains.

It is noteworthy that this algorithm fits within a workflow developed for the reconstruction of geological structures in 3D (Fernández, 2004; Mencos, 2011). Thus, this workflow supports some aspects of the structural analysis that are not tackled by the algorithm, (e.g. the definition of structural domains on site).

1.2 Clustering methods

The term "clustering" regards to the unsupervised classification of elements into groups, called clusters. The existing standard methods for clustering can be divided into two main families: hierarchical and partitioning (non-hierarchical) methods. Hierarchical methods produce a nested series of partitions, while non-hierarchical methods produce only one partition. Several surveys on clustering analysis are available in the literature (Jain et al., 1999; Bock, H.H, 2002; Xu, R., 2005).

Hierarchical methods are strongly dependent on the first classification step and do not have a clear criterion for the final cluster partition. Moreover, when large amounts of data need to be classified, a typical method in hierarchical clustering such as the dendrogram visualization becomes unpractical.

Computationally efficient partitioning methods try to reach an optimal partition depending on a given criterion function, for instance minimizing the *square-error* function (i.e. the squared distances inside the clusters). The *k*-means (MacQueen, 1967), the simplest and most commonly used algorithm employing a square-error criterion, tends to work well with a number of isolated and compact clusters, but this condition is not guaranteed in orientation measurements. Moreover, most of the non-hierarchical methods require an a priori knowledge of the number of clusters to be obtained (e.g. Zhou and Maerz, 2002) and this condition is seldom met in geological studies.

In geological engineering, several studies exist that have been developed and used clustering algorithms to classify, group and/or characterize discontinuities. Zhou and Maerz (2002) and Tokhmechi et al. (2011) compare the application of some classical methods (Parzen classifiers, *k*-means, nearest neighbor, etc.). Jimenez-Rodriguez and Sitar (2006) develop a spectral clustering algorithm that combines the *k*-means method. Nevertheless, the above mentioned methods do not impose a size restriction to the cluster members, hence arbitrary cluster sizes are obtained (in contrast with the orientation domains here defined, see condition (2)).

Thus, the above described methods will not give analogous results to the classical procedure described in the previous section. For this reason an ad hoc tool for automated clustering has been designed. This tool lays within the framework of the grid-based clustering algorithms, although with some differences compared to others existing in the literature.

Central to grid-based methods is that individual measurements are converted to cell values. However, the existing methods merge initial calculated cells with surrounding ones in function of their density (i.e. number of individual measurements within each cell). These density-oriented methods are widely applied to spatial data and image processing, but they are not suitable for the geometric characterization of geological structures, in which the number of individual measurements does not necessarily constitute a criterion for cluster partition. In fact, since data distribution is not homogeneous, one orientation domain can be represented by a single data measurement. On the contrary, the developed algorithm has been designed specifically with geological considerations during cluster partition. Moreover, it represents a new approach that merges a square-error criterion function and a grid-based but size-oriented technique, as it will be detailed below.

2. The mobile rectangular grid algorithm with spherical correction

2.1 Principles and notations

Denoted by (***u***, ***v***) is a pair of orientation angles (azimuth and dip respectively) with spherical coordinates, where azimuth ***u*** (or dip direction measured from North in a clockwise direction) takes values in [0, 360º] and dip ***v*** in [0, 90º] measured downward from horizontal. Taking unitary radius, the orientation pairs (***u***, ***v***) correspond to the following Cartesian 3D-coordinates *(Xs, Ys, Zs):*

$$Xs = \sin u \sin v, \qquad Ys = \cos u \sin v, \qquad Zs = \cos v \qquad (3)$$

These normalized direction cosinus represent unitary vectors on the sphere, and the domains separation must respect the inherent spherical geometry. Several statistical techniques exist to specifically treat the distributional properties of spherical data (Fisher et al., 1987), although they are not suitable when working on regular grids on sphere.

The algorithm is based on the planar representation of orientation data considering their Cartesian coordinates (azimuth against dip in a 2-axis Cartesian plot or ***u-v*** plot). It is well known that the representation of oriented data on a ***u-v*** plot introduces a distortion that is more accentuated towards the horizontal values (maximum distortion tends to a singularity in horizontal dips) and invalidates the results of the clustering analysis. For example, in a stereographic projection, the poles of the subhorizontal planes appear clustered around the centre of the sphere, leading to the interpretation of a single orientation domain. On the contrary, in the ***u-v*** plot data close to the horizontal appear scattered in the lower part resulting in an overrepresented classification (Figure 1 A). Without losing sight of the distortion problem (that can be corrected a posteriori as it will be explained afterwards), the advantages of using a ***u-v*** plot are:

a) it facilitates the definition of a regular mesh that takes into account the range in azimuth (***u_0***) and dip (***v_0***) observed within orientation domains (and explained in the previous section 1.1). This regular mesh divides the orientations space in ***n*** x ***m*** regular cells or isometric areas defining orientation domains characterized by the [***u_0,v_0***] range (Figure 1 B);

b) it is manageable, from a computational point of view, as opposite to spherical representations;

c) a rigid shift of an initial grid can be easily implemented.

With these assumptions in mind, the proposed clustering process uses the cylindrical projection (identifying continuity between 0º and 360º in azimuth, Figure 1 C, D and E). In this projection it superimposes a family of regular grids in order to find out which grid in that family best separates the orientation data. This gives a first clustering classification that is corrected later on in order to avoid singularities and correct the distortion.

Before going further, given below are some details about the distortion. The distortion can be numerically evaluated as follows:

Assuming a small rectangular [***u_0,v_0***] cell in the planar representation with centre (U,V). The area of its spherical image can be approximated by

$$u_0 \cdot \sin(V) \cdot v_0 \quad (4)$$

which is smaller than ***u_0*** x ***v_0*** unless ***v***=90º. Thus, in order to guaranty spherically isometric domains, the rectangular cells should be locally corrected taking

$$u_0\,(V) \times v_0, \quad \text{where} \quad u_0\,(V) = u_0 / \sin V \qquad (5)$$

where ***V*** is the dip mean value for all the measurements in the cell. Thus, the size of the cells corresponding to subhorizontal dips is enlarged, while the cells for vertical dips remain practically unchanged. The first clustering classification is then modified by merging the domains that fit all together into a new (enlarged) cell, this new cell being centred in the redefined common centroid point. The corrections ensure that, at the end, the clusters cells are approximately of equal spherical area. These local modifications adapt the final solution to the configuration of the orientation measures and break the rigidity of the initial mesh too.

2.2 The algorithm

2.2.1 Part I: Rigid shifting grid-based method

The first part of the algorithm determines an initial partition of the orientation angles ($\boldsymbol{u}$,$\boldsymbol{v}$) into clusters. At the end, all the angle pairs within the same cluster will be close enough one to each other to satisfy condition (2). Solution is approached by applying a rigid shifting grid-based method to find a kind of optimal fitting. Each step is listed below in detail (Figure 2 a to e):

a) The algorithm reads from an ASCII file consisting in *n* pairs of orientation angles $(u_1,v_1),\ldots,(u_n,v_n)$. Additional information in the ASCII file are geographical coordinates in UTM format *(x, y, z)* and polarity (defined by *N* as normal; *I* as reversed).

b) User is required to type the tolerance accepted within an orientation domain (grid width $\boldsymbol{u_0}$; grid height $\boldsymbol{v_0}$) and a grid mobility increment parameter ($\boldsymbol{p}$). This parameter $\boldsymbol{p}$ will determine the shifting of the regular mesh at later steps.

c) The algorithm searches for horizontal data within the file (v=0). If horizontal data are found they are omitted in the cluster calculation and printed in a separate file as a single horizontal domain. This step prevents any division by 0 (see equation 5).

d) The algorithm generates a regular grid with grid spacing ($\boldsymbol{u_0},\boldsymbol{v_0}$) and grid vertex (the lower left point of the lower left cell) anchored in the origin of the coordinate system.

This grid separates the data into the grid cells. Given two orientation measurements (u_i,v_i) and (u_j,v_j), if they are in the same cell, then they satisfy

$$|u_i-u_j| \leq u_0 \quad \text{and} \quad |v_i-v_j| \leq v_0 \qquad (6)$$

At this stage, all the measurements in a cell perform a cluster. For any cluster partition, the usual $\boldsymbol{R^2}$ (R-square) statistic index is computed,

$$R^2 = 1\text{-}(\textit{variability within clusters})/(\textit{total variability}) \qquad (7)$$

where the variability is computed as the sum of the squared distances of the measures with respect to the corresponding centroid. This index is a quality criterion of fit to the particular partition. It is computed in terms of $(\boldsymbol{u},\boldsymbol{v})$, i.e. the cylindrical representation, but it works locally well on the sphere because of the small cells size.

e) The algorithm looks for an optimum cluster classification (based on $\boldsymbol{R^2}$ criterion). This is performed by moving rigidly the grid vertex (anchoring point) of that initial grid, both horizontally and vertically and by tiny increments of $\boldsymbol{p}$ size (user defined increment parameter). There will exist as many grid configurations as points fit within the lower left grid cell of the initial grid, depending on $\boldsymbol{p}$ parameter. Each of these new generated grids satisfies condition (2). The optimum, in this case depending on $\boldsymbol{p}$, is reached when the rigid rectangular grid best fits the set of nodes, i.e. maximises $\boldsymbol{R^2}$. Notice that highest $\boldsymbol{R^2}$ is equivalent to a minimum *square-error criterion function* (Jain et al., 1999). The idea of shifting a grid structure has been used by several authors for shape recognition (Ma and Chow, 2004; Chang et al., 2009).

2.2.2 The algorithm. Part II: spherical and unrigidity corrections

The second part of the method consists in applying a correction to the initial cluster distribution (Figure 2 e), aiming to reduce the distortion and improve the obtained results. The correction consists in two operations that are done simultaneously (Figure 2 f to h):

f) Spherical adaptation: This step is necessary to adapt the grid partition to the spherical geometry of data (Figure 3). It consists in converting the initial ($\boldsymbol{u_0}$,$\boldsymbol{v_0}$) cells (isometric on the Cartesian plane) to pseudo-isometric clusters on the sphere (Figure 3 A). As it has been pointed out previously, to correct areal distortion, the resulting cell size can be rewritten as (5).

The new rectangular cells which size is defined by (5) are wider as they approach the zero dip area. The spherically adapted clusters are not isometric any more in the plane but they approximately are on the sphere. In this way, orientation domains (highly) horizontal will admit a strong variation in the $\boldsymbol{u}$ component (Figure 1 A). Ideally, subhorizontal nodes will be part of a single orientation domain despite of their azimuth attitude.

g) Unrigidity correction: A pair of nodes can be close enough one to each other to be part of the same orientation domain (i.e. accomplish condition (2)), but the partition obtained after step (e) separates them into different clusters. This situation is related to the rigidity of the mesh that cannot be adapted to the entire set of nodes. This mesh rigidity can be improved through the application of a proximity criterion, so as to regroup some domains originally separated during the initial calculation (Figure 3). This operation is performed by searching for all the initial clusters that, even separated, fit entirely within a spherically adapted cell. The search is done in an organized way starting by the closest centroid pair.

h) Final results: Output file is an ASCII data file including the location (x,y,z) of the original data point; its orientation (azimuth, dip); identification number of the orientation domain to which it belongs to; number of nodes within that orientation domain; calculated cluster azimuth (centroid azimuth); calculated cluster dip (centroid dip); distance in azimuth between the data point and the calculated cluster centroid; distance in dip between the data point and the calculated cluster centroid; range in azimuth within calculated orientation domain; range in dip within calculated orientation domain; identification number relating the data points to the position of these data points in the original file; and finally the polarity, using 0 as a normal or 1 as an inverse (Table 1).

As a summary, the designed algorithm subdivides a set of orientation data into constant orientation domains, using user-defined tolerance thresholds that account for variability within orientation domains. The designed approach does not require a prior knowledge of the number of clusters to identify as well as their geographic location. The program has been implemented in C.

3. Sample synthetic experiments

A set of experiments has been designed to test the capability of the algorithm. These experiments consist in different synthetic geological structures, each one representing a fold with a specific structural configuration. The first experiment is used to illustrate the algorithm behaviour in detail (Figure 4 B and C, see Figure 5, and 6 for results). The other experiments (Figure 7 and 8) represent more complex structures and are used to test the algorithm response in front of different ideal situations. The objective is to illustrate the relationship between the output of the algorithm and the synthetic structures, i.e. the capability of the algorithm to identify representative orientation domains at convenience.

The experiments set up has been done as follows: each structure was generated in a 3D reconstruction program by creating a folded surface (with a scale of hundreds of meters). After that, some roughness was added to the surface in order to mimic the variability that accounts for instrumental error, natural roughness and sharpness (except experiment 3, Figure 7 B). This was done using a random function (urand(0.5, -0.5) in meters) applied to the Z value of each node of the initial surface (Fig 4 A). As a result, the final orientation of the surface triangles could vary up to +-10º. Finally, a set of discrete orientation values was randomly picked on the surface, aiming to represent a realistic field data acquisition (Figure 4 B). Thus, the final result is a set of scattered points, each of them having a particular location and orientation (x, y, z, azimuth, dip). After that, the experiment set up was ready for conventional structural analysis (Figure 4 C).

In the first experiment, the generated structure corresponds to a kink-type fold including six planar regions separated by sharp hinges (Figure 4 B). Fold geometry is cylindrical with horizontal axis. Each of the planar regions has a characteristic orientation (dip/azimuth value). The results of the conventional structural analysis are shown in Figure 4 C, and are consistent with the six-region structure.

In the second experiment, the created structure is also a kink-type fold constituted by five planar regions, within which azimuth and dip remain more or less constant (Figure 7 A.1). In this case fold geometry is conical with horizontal axis.

In the third experiment, the created structure is a smooth folded surface that represents a conical fold with continuous curvature (Figure 7 B.1), i.e. it can be defined as constituted by an infinite number of planar regions. In this case, dip and azimuth show a progressive change that is more pronounced close to the cone apex.

The fourth experiment represents also a kink-type fold with cylindrical fold geometry and horizontal axis. In this case, the fold has three cylindrical domains with two structural trends (Figure 7 C.1).

Two additional experiments have been designed using the above described bi-axial kink-type cylindrical fold: The fifth experiment, consisting in the selection of a data subset considering only one cylindrical domain of the fourth experiment (Figure 8 A). Finally, the sixth experiment, consisting in a random selection of data extracted from the experiment 4 (Figure 8 B).

3.1. Test results

Experiment 1: The algorithm has been run nine times (T1 to T9) with different values of dip range (v_0). Azimuth threshold has been maintained constant through all tests (u_0=10º), as changes in strike are negligible in the designed synthetic structure. v_0 ranges from 5 to 45 degrees, with an incremental value of 5 degrees in each run. Results are summarized in figures 5 and 6.

Experiment 2: The algorithm has been run two times varying dip range (v_0 = 5º, 15º) and maintaining azimuth constant (u_0=10º) to enhance the dip influence in cluster identification. The second run (u_0, v_0 = 10º,15º) solved the five planar regions of the synthetic structure (Figure 7 A.2 and A.3).

Experiment 3: The algorithm has been run two times maintaining dip range constant (v_0=15º) and varying azimuth range (u_0= 10º, 45º). The algorithm tends to separate the data into narrower orientation domains as [u_0, v_0] decreases. Subvertical limbs do not show significant differences between runs due to very low azimuth variability in these areas; in contrast, subhorizontal domains are larger as azimuth increases, due to a greater variability in azimuth (Figure 7 B.2 and B.3).

Experiment 4: The algorithm has been run two times ($\boldsymbol{u_0, v_0}$ = 15º,15º and 30º,15º). In this case, the algorithm identified orientation domains despite their geographical position (e.g. orientation domains within cylindrical domain number 2, Figure 7 C.2). Greater azimuth ranges ($\boldsymbol{u_0}$=30º) give a fewer number of clusters in the hinge area, where azimuth variability is higher (compared to the fold limbs).

Experiment 5: The algorithm has been run once, using the same tolerance thresholds of the fourth experiment ($\boldsymbol{u_0, v_0}$ = 30º,15º) to compare the results. Fewer clusters were found near the hinge area compared to experiment 4, due to a lower variability in azimuth and dip (e.g. compare the circled areas in Figure 8 A, each one containing a single cluster, in contrast with the same areas of Figure 7 C.3). This lower variability in azimuth and dip can be related to a fewer amount of data.

Experiment 6: The algorithm has been run once with the same tolerance thresholds than experiment 4 ($\boldsymbol{u_0, v_0}$ = 30º,15º). The use of fewer data implies les azimuth and dip variability, and therefore fewer orientation domains are obtained (Figure 8 B). However, the structure is well defined.

4. Discussion

The synthetic experiments allowed exploring the capability of the program to solve the given structures with different resolutions, by varying one or both of the user-defined initial parameters (dip or azimuth values). Some of the results are discussed below.

In the first experiment, considering that the synthetic cylindrical structure has been built using six planar domains, the best solution is when the six expected domains are distinguished ($\boldsymbol{u_0, v_0}$ = 10º,5º and $\boldsymbol{u_0, v_0}$ = 10º,10º). Dip ranges greater than these values produce an under-sampled structure. If the structure under study is a kink-type fold conical structure (experiment 2), the application of the algorithm can also identify the expected five planar regions.

In a real case, where the geometry of the structure under study is generally unknown, there is not a unique best-fit solution, as all possible solutions would be computationally correct. The best-fit orientation-domain discrimination will depend on the available data, the desired resolution and the geological properties of the materials under study (e.g. lithology, bedding or texture, among others) (Figure 9). Moreover, if the structure could be defined as continuous, as for example in experiment 3, then the desired resolution is definitely a key-factor for cluster partition. In such a case, there would be as many clusters as initial data exist, because the structure is defined as smooth and continuous (Figure 7 B). Ideally, there are not a finite number of planar domains that define the geometry of the structure, so that the final solution depends on the analyst.

Initially, no geographic position is required to identify planar regions considering their orientation, hence orientation data with similar values can be grouped into the same cluster even if they are geographically separated (experiment 4, Figure 7.C). This can give relevant information about the fold geometry and/or evolution when it is framed within a reconstruction process. As the reconstruction process goes forward, it could be necessary to spatially select data subsets in order to refine the results (Figure 8 A)

The extraction of a data set from experiment 4 by area or randomly (experiments 5 and 6, respectively, Figure 8) leads to similar results with small differences since initial data are different. However, these differences do not prevent to obtain a correct geometry of the analyzed structure.

A lower threshold of azimuth and/or dip range can be established, below which the orientation-domain configuration will not describe the geological geometry of the structure under study (Figure 10). The orientation-domain configuration below this threshold would be biased by the instrumental error, geological roughness and sharpness (Figure 9).

As the azimuth and/or dip range increases, the geometry depicted by orientation domains has lower resolution, and the number of identified planar regions decreases (Figure 5 and 6). At the end, there is an upper tolerance threshold such that all the available data will belong to a single orientation domain (Figure 9).

Compared to the semi-manual approach, the designed algorithm is fast for the orientation domain definition, as well as it gives objective results. This last fact is due to an automatic grouping of the original data considering only the user-established initial thresholds.

A potential alternative to our approach could be based on *quaternions* (lying in the hypersphere S^3), which provide a computationally efficient way to store and rotate 3-dimensional vectors (e.g. Karney, C., 2007). In Karney, C., 2007, the quaternions algebra is used to solve several problems in the orientations space. In particular, they describe the projection of a cubical regular grid (defined on a *tesseract* or 4-dimensional cube), over the hypersphere. This results in a distorted grid with maximum distortion in the corners. The implementation of this method to our context would imply projections of a regular grid onto a classical sphere (S^2) causing a distortion too. The regular shift of the grid would cause unchecked distortions that would not add significant improvement to our implementation.

5. Conclusions

An algorithm is presented to automatically obtain constant orientation domains. It is based on a shifting rectangular grid clustering algorithm. Three main requirements led to the use of a grid-based algorithm: unknowing the number of clusters to be obtained, omitting the geographic location of data during process and obtaining clusters composed of close orientations.

The algorithm first generates data clusters from a set of orientation data. These initial clusters are subsequently improved by making a deformation of the grid to adapt it to the spherical geometry inherent to the orientation measurements.

The resulting domain classification is based on *preserve a size* criterion for the output cluster orientation domains. It starts working in the angular space and then corrects the distortion to be almost isomorphic on the spherical representation given by the director cosinus.

It depends on the user parameter specifications. The accurate definition of the threshold parameters is a fundamental task for the analyst. The effects of measurements accuracy, the work scale, the lithology, the structural style of deformation, etc., must be taken into account when defining the parameter thresholds. By testing different thresholds, the computed partitions can be improved and this is controlled by the quality criteria of fit. Nevertheless, before a definitive final orientation domain assignment, the output domains should be plotted on a 3D representation of the terrain, where other variables can be taken into account (i.e. geographic proximity, stratigraphic position or lithology, among others).

The performed experiments conclude that the algorithm gives acceptable results on the selected tolerances with respect to the data distribution, for the selected geometries.

The introduction of the “mobile grid algorithm with spherical adaptation and unrigidity correction” speeds up the process of structural analysis and improves the existent workflow for the reconstruction of geological structures (Fernández, 2004). The obtaining of any output result is fast compared to the manual approach, so that the algorithm can be applied multiple times with different input parameters. With such a procedure, multiple possible solutions can be explored in a short amount of time, until an adequate result is obtained.

Acknowledgements

This work has been carried out with the financial support of the Inversión Positiva de Estructuras de Tectónica Salina (CGL2010-21968-C02-01) and Modelización Estructural 4D (CGL2007-66431-C02-01BTE) projects. We wish to acknowledge Grup de Geodinàmica i Anàlisi de Conques (2009SRG-1198), GEOMODELS Research Institute and Grup de Recerca en Aplicacions i Models Matemàtics. We also wish to acknowledge Paradigm for providing gOcad software, which has been used for data management and visualization. The original manuscript has been improved thanks to the valuable comments of two anonymous reviewers and editor Jef Caers.

6 References

Bock, H.H., 2002. Clustering methods: from classical models to new approaches. Statistics in transition 5(5), 725-758.

Carrera, N. and Muñoz, J. A. and Roca, E., 2009. 3D reconstruction of geological surfaces by the equivalent dip-domain method: An example from field data of the Cerro Bayo Anticline (Cordillera Oriental, NW Argentine Andes), Journal of Structural Geology 31(12), 1573-1585.

Chang, C-I., Lin, N.P., Jan, N-Y., 2009. An Axis-Shifted Grid-Clustering Algorithm. Tamkang Journal of Science and Engineering 12(2), 183-192.

Cruden, D.M., Charlesworth, H.A.K., 1972. Observations on the numerical determination of axes of cylindrical and conical folds. GSA Bulletin 83, 2019-2024.

Cruden, D.M., Charlesworth, H.A.K., 1976. Errors in strike and dip measurements, GSA Bulletin 87, 977-980.

Fernández, O., 2004. Reconstruction of geological surfaces in 3D. An example from the southern Pyrenees. PhD Dissertation, University of Barcelona, Barcelona, Spain, 321 pp.

Fernández, O., 2005. Obtaining a best fitting plane through 3D georeferenced data. Journal of structural geology 27, 855-858.

Fisher, N.I., Lewis, T., Embleton, B.J.J., 1987. Statistical Analysis of Spherical Data. Cambridge University Press, UK, 344pp.

Gill, W.D., 1953. Construction of geological sections of folds with steeplimb attenuation, AAPG Bulletin 37(10), 2389-2406.

Groshong, R.H.Jr., 2006. 3D Structural Geology: A Practical Guide to Quantitative Surface and Subsurface Map Interpretation. Springer-Verlag, DE, 400 pp.

Jain, A.K., Murty, M. N., Flynn, P.J., 1999. Data clustering: A review. ACM Computing Surveys 31(3), 264–323.

Jimenez-Rodriguez, R., Sitar, N., 2006. A spectral method for clustering of rock discontinuity sets. International Journal of Rock Mechanics and Mining Sciences 43 (7), 1052-1061.

Karney, C., 2007. Quaternions in molecular modeling. Journal of Molecular Graphics and Modelling 25 (5), 595-604.

Ma, E.W.M., Chow, T.W.S., 2004. A new shifting grid clustering algorithm. Pattern recognition 27, 503-514.

MacQueen, J. B., 1967. Some Methods for classification and Analysis of Multivariate Observations. In: Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, US, California Press 1, 281-297.

Mencos, J., 2011. Metodologies de modelització i reconstrucció 3D d'estructures geològiques: anticlinal de Sant Corneli – Bóixols (Pirineus Centrals). PhD Dissertation, University of Barcelona, Barcelona, Spain, 277 pp.

Ramsay, J.G., 1967. Folding and Fracturing of Rocks. McGraw Hill, US, 568 pp.

Shaw, J.H., Connors, C., Suppe, J., 2005. Seismic Interpretation of Contractional Fault-Related Folds: An AAPG Seismic Atlas. American Association of Petroleum Geologists, Studies in Geology 53.

Suppe, J.,1983. Geometry and kinematics of fault-bend folding, American Journal of Science 283, 684-721.

Suppe, J., 1985. Principles of Structural Geology. Prentice-Hall, US, 537 pp.

Tokhmechi, B., Memarian, H., Moshiri, B., Rasouli, V., Noubari, H. A., 2011. Investigating the validity of conventional joint set clustering methods. Engineering Geology 118 (3-49), 75-81.

Torrente, M.M., Civile, D., Martino, C., Milia, A., 2000. Assetto strutturale ed evoluzione tettonica dell'area di Monte Vesole-Monte Chianello (Cilento, Appennino meridionale), Bollettino della Società Geologica Italiana 119, 733-747.

Wise, D.U., 1992. Dip Domain method applied to the Mesozoic Connecticut valley rift basins, Tectonics 11(6), 1357-1368.

Woodcock, N.H., 1977. Specification of fabric shapes using an eigenvalue method. Geological Society of America Bulletin 88, 1231-1236.

Xu, R., 2005. Survey of clustering algorithms. IEEE Transactions on neural networks 16(3), 645-678.

Zanchi, A., de Donatis, M., Gibbs, A., Mallet, J.L., 2009. Imaging geology in 3D, Computers & Geosciences 35 (1), 1-3.

Zhou, W., Maerz, N.H., 2002. Implementation of multivariate clustering methods for characterizing discontinuities data from scanlines and oriented boreholes. Computers and Geosciences 28 (7), 827-839.

Figure Captions

Figure 1. Stereographic (equal-area lower hemisphere stereoplots) versus Cartesian representation of orientation data. A. Differences between stereographic and ***u-v*** plot representations of a subhorizontal data set. B. Regular mesh superimposed to the ***u-v*** plot and defined by ***n*** x ***m*** cells in function of the defined u_0 and v_0 values. White nodes are azimuth-dip angle pairs. Black node is the mean azimuth and dip value (centroid). C. Stereographic projection of a set of nodes. D. ***u-v*** planar plot representation of the same data set. E. ***u-v*** cylindrical representation with 0º and 360º identification.

Figure 2. Flow-chart describing the procedure followed by the algorithm. See text for a more detailed explanation of **a** to **h** steps.

Figure 3. Corrections (gray lines) applied to the initial cluster distribution (black lines). Black dots denote the initial clusters centroids without corrections. Grey dots denote the new clusters centroids after corrections. Sphericity correction: note that the influence of this correction is important in lower dips (A) and small in higher ones (B), because when $v \approx 90º$, then $\sin(v) \approx 1$. Rigidity correction: Performed to join a pair of nodes close enough to be part of the same orientation domain. This correction affects in the same way all considered nodes, independently of their position in the Cartesian plot (A and B).

Figure 4. Initial setup for the first sample synthetic experiment. A. Illustration of how roughness is applied to the original folded surface using a random function. This function modifies Z values of the surface nodes and consequently the orientation of the surface triangles. B. Six-region kink-type fold geometry and the set of points randomly picked on the surface (represented as oriented disks). C. Stereographic representation (equal-area lower hemisphere stereoplot) and associated statistics of the data set

showing a cylindrical distribution and a six-region structure. E1, E2 and E3 denote the resulting eingenvectors (E1 representing the highest one and E3 the lowest).

Figure 5. Test results on the synthetic data set for the first experiment: ***u-v*** plot (left) and perspective view (right). To the left, blue corresponds to data points and red corresponds to centroids. To the right, the coloured disks correspond to the orientation points picked on the surface, coloured in function of the cluster assignment (note that colour is assigned randomly in each run). C1 to C6 indicate the orientation domains. Test results separated from A to D in function of the given tolerance thresholds: T1: v_0 = 5º; T2: v_0 =10º; T3: v_0 =15º; T4: v_0 =20º; T5: v_0 = 25º; T6: v_0 = 30º; T7: v_0 = 35º; T8: v_0 = 40º; T9: v_0 = 45º.

Figure 6. Summary of the obtained results. Setup parameters are: T1: v_0 = 5º; T2: v_0 =10º; T3: v_0 =15º; T4: v_0 =20º; T5: v_0 = 25º; T6: v_0 = 30º; T7: v_0 = 35º; T8: v_0 = 40º; T9: v_0 = 45º. Obtained orientation domains are: T1-T2; 6 clusters; T3-T6: 4 clusters; T7: 3 clusters; T8-T9: 2 clusters.

Figure 7. Set up configuration and test results for experiments 2, 3 and 4, corresponding respectively to a conical kink-type fold (A), conical smooth fold (B) and bi-axial cylindrical kink-fold (C). For each experiment, 1 shows the stereographic projection of initial data, 2 shows the test results in a 3D perspective view (disks are coloured in function of cluster assignment), 3 shows the test results plotted on a ***u-v*** plot (with initial data points and clusters centroids).

Figure 8. Test results for experiments 5 and 6. A. Experiment 5: Cylindrical kink-type fold extracted from one sector of the experiment 4. B. Experiment 6: Data subset randomly selected from experiment 4 configuration (Bi-axial kink-type fold geometry is preserved). See text for more detailed explanations.

Figure 9. Plot of the number of obtained clusters in function of $\boldsymbol{v_0}$. The number of identified planar regions decreases for larger $\boldsymbol{v_0}$ thresholds. Note that the orientation ranges below a certain threshold can be attributed to an inherent error in the orientation domain separation.

Figure 10. Orientation domains identified for experiment 1 when using a tolerance threshold below the resolution of the designed experiment. A. Cluster distribution in a perspective view (v_0 = 1). Disks are coloured in function of cluster assignment. B. Number of planar domains obtained using small dip thresholds.

Table Captions

Table 1. Example of output results. The table represents part of an output ASCII file showing 26 initial data grouped into two orientation domains (the first one with 11 data and the second one with 15) and the given related parameters for each point: x, y and z coordinate, azimuth and dip values, orientation domain assignation, number of points included in the domain, azimuth and dip of the calculated centroid, distance in orientation between the point and the corresponding centroid, orientation domain range, identification number for the point and polarity.

x coord	y coord	z coord	Original Azimuth	Original Dip	Orientation domain nr.	Nr. of data in the domain	Cluster Azimuth	Cluster Dip	Difference with Az	Difference with Dip	Az range in domain	Dip range in domain	Id nr.	Polarity
13052.57	978.91	0.99	120.0	34.0	1	11	119.91	34.51	0.084	0.510	3.10	3.15	1	0
13108.65	1070.17	1.76	120.3	33.6	1	11	119.91	34.51	0.431	0.878	3.10	3.15	3	0
13011.77	902.78	4.43	119.4	32.9	1	11	119.91	34.51	0.500	1.600	3.10	3.15	4	0
13144.60	1131.14	5.20	121.0	33.0	1	11	119.91	34.51	1.088	1.510	3.10	3.15	5	0
13044.95	971.58	1.07	120.9	33.4	1	11	119.91	34.51	1.023	1.102	3.10	3.15	6	0
12998.95	887.07	3.70	120.0	35.0	1	11	119.91	34.51	0.086	0.489	3.10	3.15	0	0
13067.70	1008.26	1.67	119.6	36.0	1	11	119.91	34.51	0.279	1.537	3.10	3.15	2	0
13090.45	1035.17	10.63	119.6	35.5	1	11	119.91	34.51	0.267	0.943	3.10	3.15	7	0
13003.62	885.65	10.98	117.9	36.1	1	11	119.91	34.51	2.009	1.551	3.10	3.15	8	0
13060.47	991.69	5.37	120.2	34.5	1	11	119.91	34.51	0.288	0.021	3.10	3.15	9	0
13124.01	1099.41	6.32	120.0	35.6	1	11	119.91	34.51	0.055	1.059	3.10	3.15	10	0
13012.43	960.04	23.05	119.0	54.3	2	15	120.50	55.03	1.501	0.764	4	3	22	0
12991.01	910.55	24.46	123.0	53.0	2	15	120.50	55.03	2.499	2.028	4	3	14	0
13026.41	949.38	8.36	120.1	54.8	2	15	120.50	55.03	0.407	0.203	4	3	11	0
13129.09	1140.78	18.37	120.1	56.0	2	15	120.50	55.03	0.430	0.972	4	3	13	0
13094.94	1094.37	23.57	120.0	54.6	2	15	120.50	55.03	0.490	0.393	4	3	15	0
13081.93	1065.88	16.65	119.0	55.8	2	15	120.50	55.03	1.501	0.732	4	3	16	0
13102.51	1099.76	11.21	120.2	54.9	2	15	120.50	55.03	0.349	0.137	4	3	17	0
13050.96	1006.83	12.98	120.0	56.0	2	15	120.50	55.03	0.481	0.972	4	3	20	0
13063.53	1021.47	4.81	119.9	54.8	2	15	120.50	55.03	0.562	0.272	4	3	23	0
13009.72	922.46	4.55	120.2	55.9	2	15	120.50	55.03	0.272	0.828	4	3	25	0
13111.80	1109.74	13.44	121.0	54.8	2	15	120.50	55.03	0.499	0.245	4	3	12	0
13055.29	1025.48	20.66	121.0	55.0	2	15	120.50	55.03	0.499	0.022	4	3	18	0
13080.15	1047.18	0.73	121.0	54.6	2	15	120.50	55.03	0.499	0.384	4	3	19	0
13117.86	1108.04	3.78	122.0	55.7	2	15	120.50	55.03	1.499	0.644	4	3	21	0
13014.93	937.44	8.91	121.0	55.3	2	15	120.50	55.03	0.499	0.298	4	3	24	0

...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Figure 1

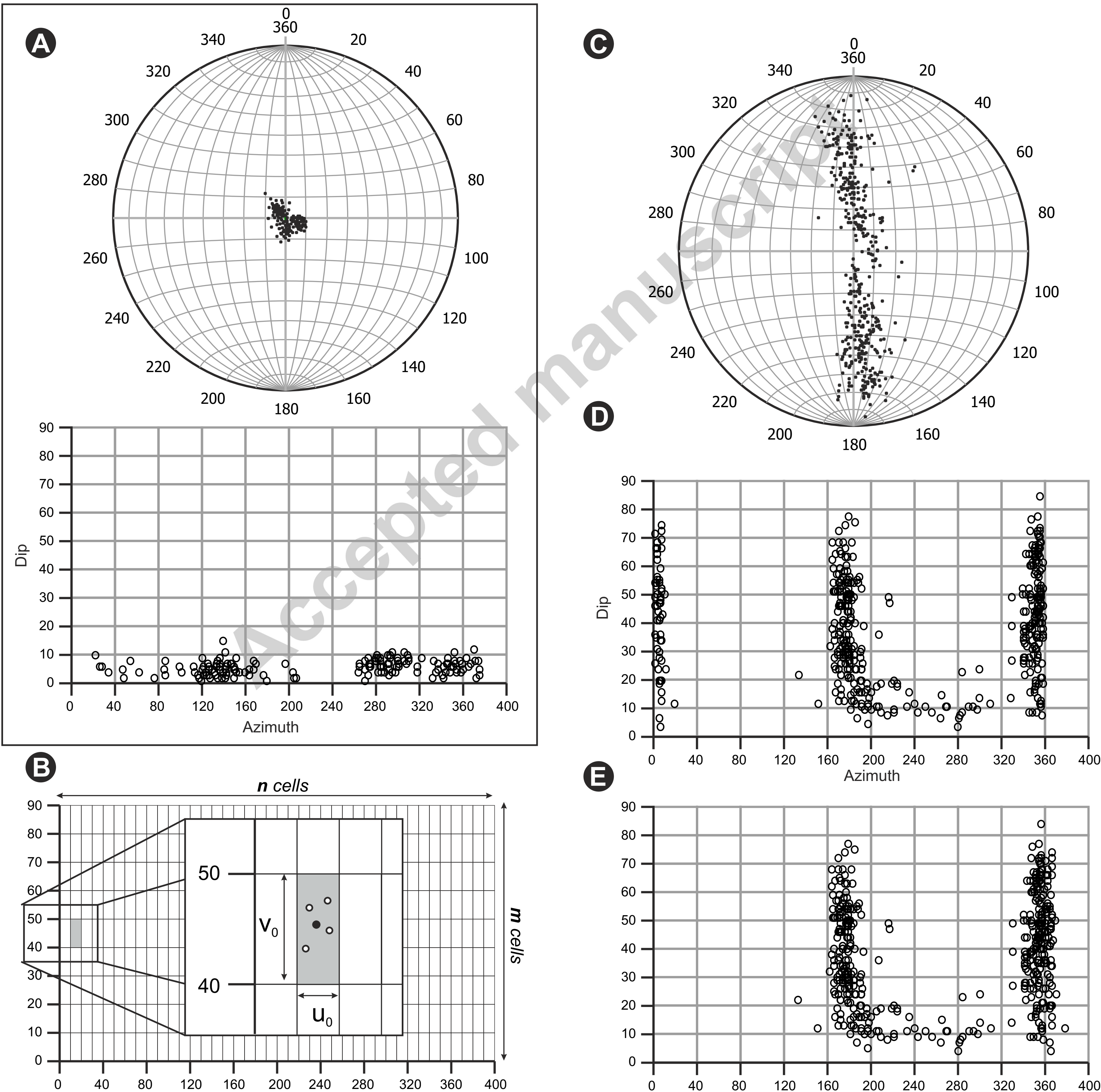


Figure 2

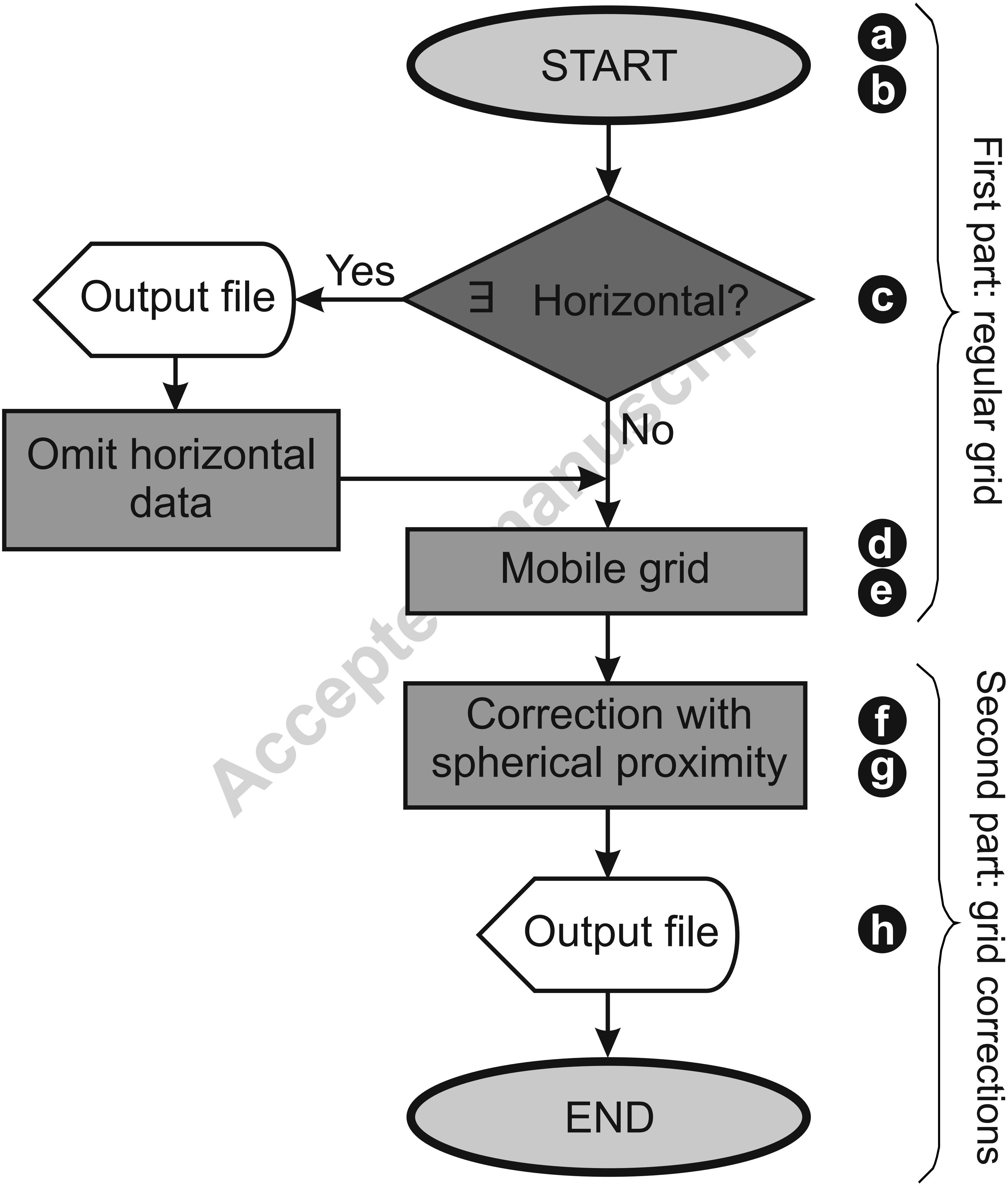

START
∃ Horizontal?
Yes
Output file
Omit horizontal data
No
Mobile grid
Correction with spherical proximity
Output file
END
a
b
c
d
e
f
g
h
First part: regular grid
Second part: grid corrections

Figure 3

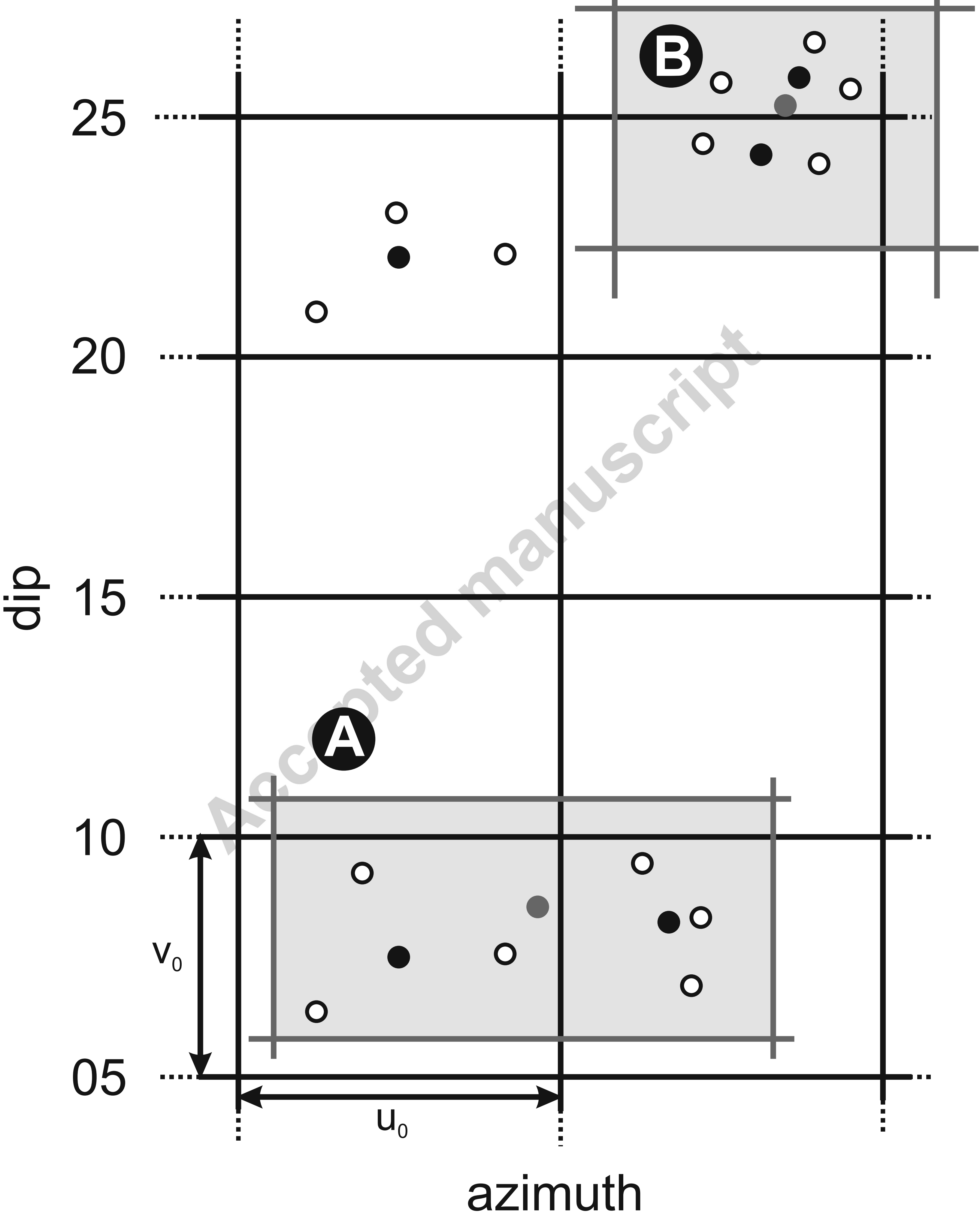
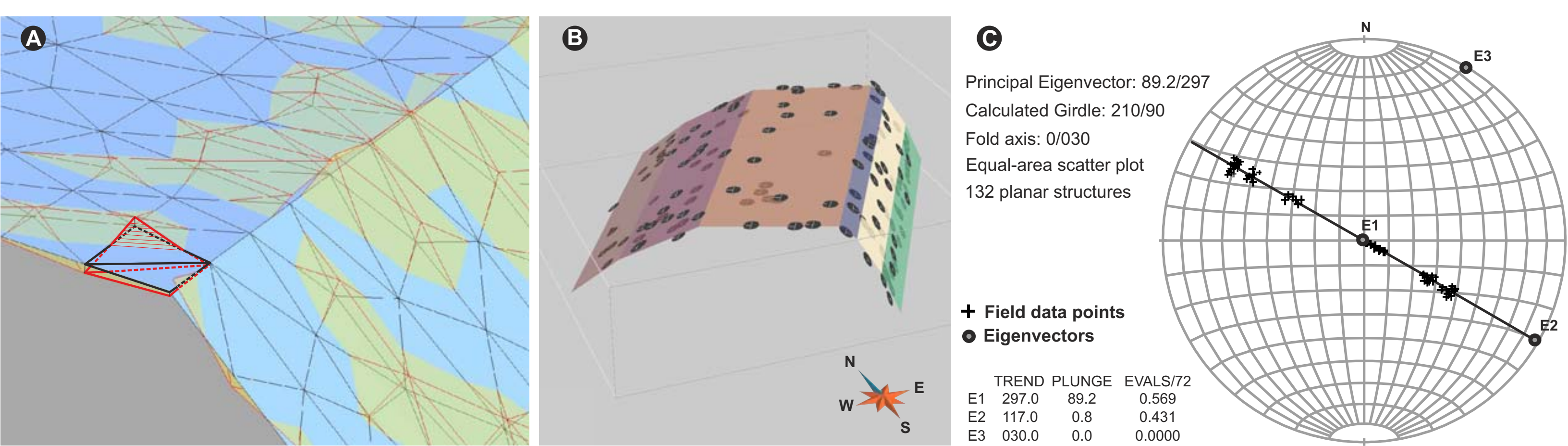
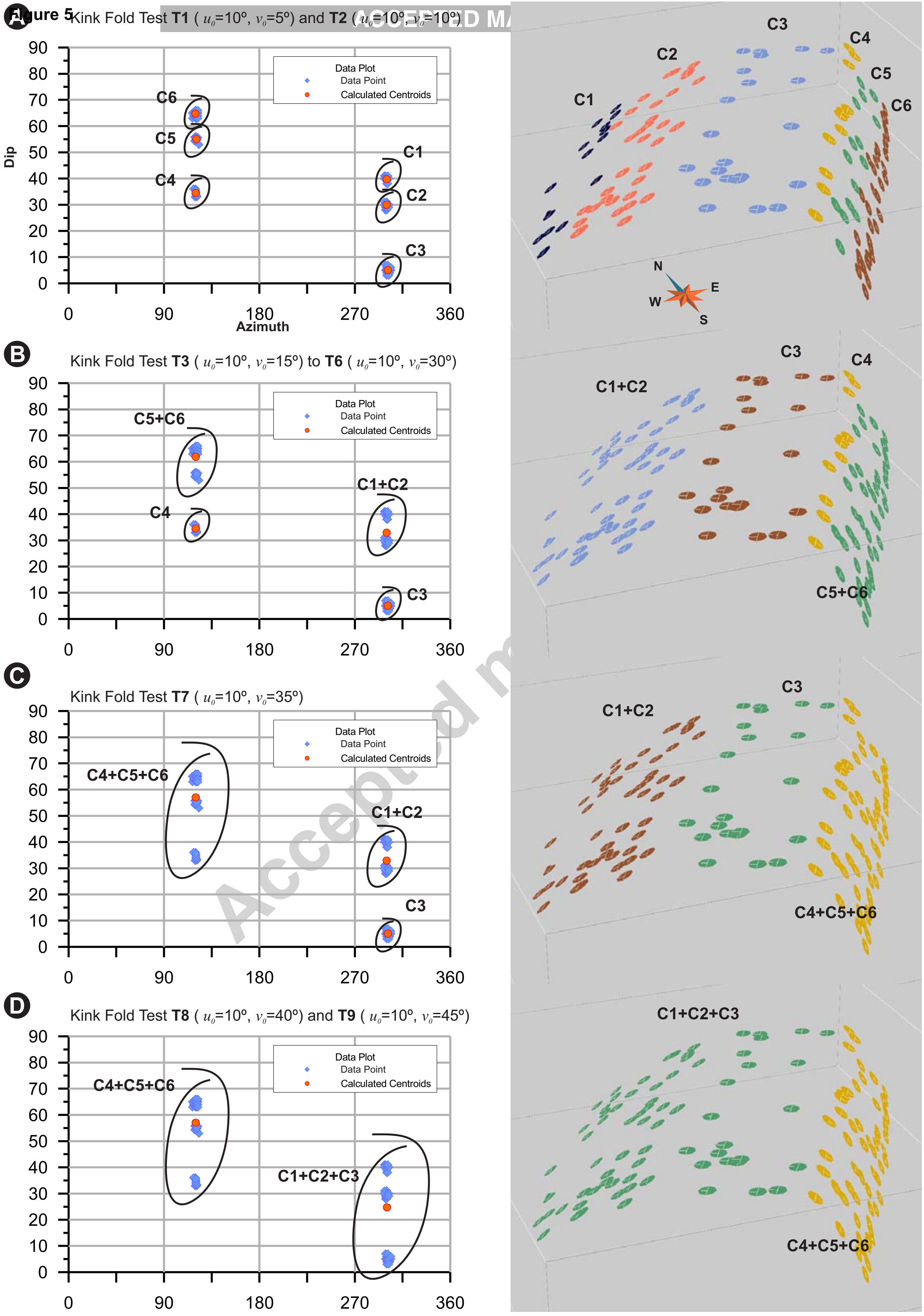


Figure 4 colour



	TREND	PLUNGE	EVALS/72
E1	297.0	89.2	0.569
E2	117.0	0.8	0.431
E3	030.0	0.0	0.0000

Figure 5



A
Kink Fold Test T1 (u_0=10º, v_0=5º) and T2 (u_0=10º, v_0=10º)
Data Plot
Data Point
Calculated Centroids
Dip
Azimuth
0
90
180
270
360
0
10
20
30
40
50
60
70
80
90
C1
C2
C3
C4
C5
C6
N
E
W
S
B
Kink Fold Test T3 (u_0=10º, v_0=15º) to T6 (u_0=10º, v_0=30º)
C5+C6
C4
C1+C2
C3
C
Kink Fold Test T7 (u_0=10º, v_0=35º)
C4+C5+C6
C1+C2
C3
D
Kink Fold Test T8 (u_0=10º, v_0=40º) and T9 (u_0=10º, v_0=45º)
C4+C5+C6
C1+C2+C3

Figure 6

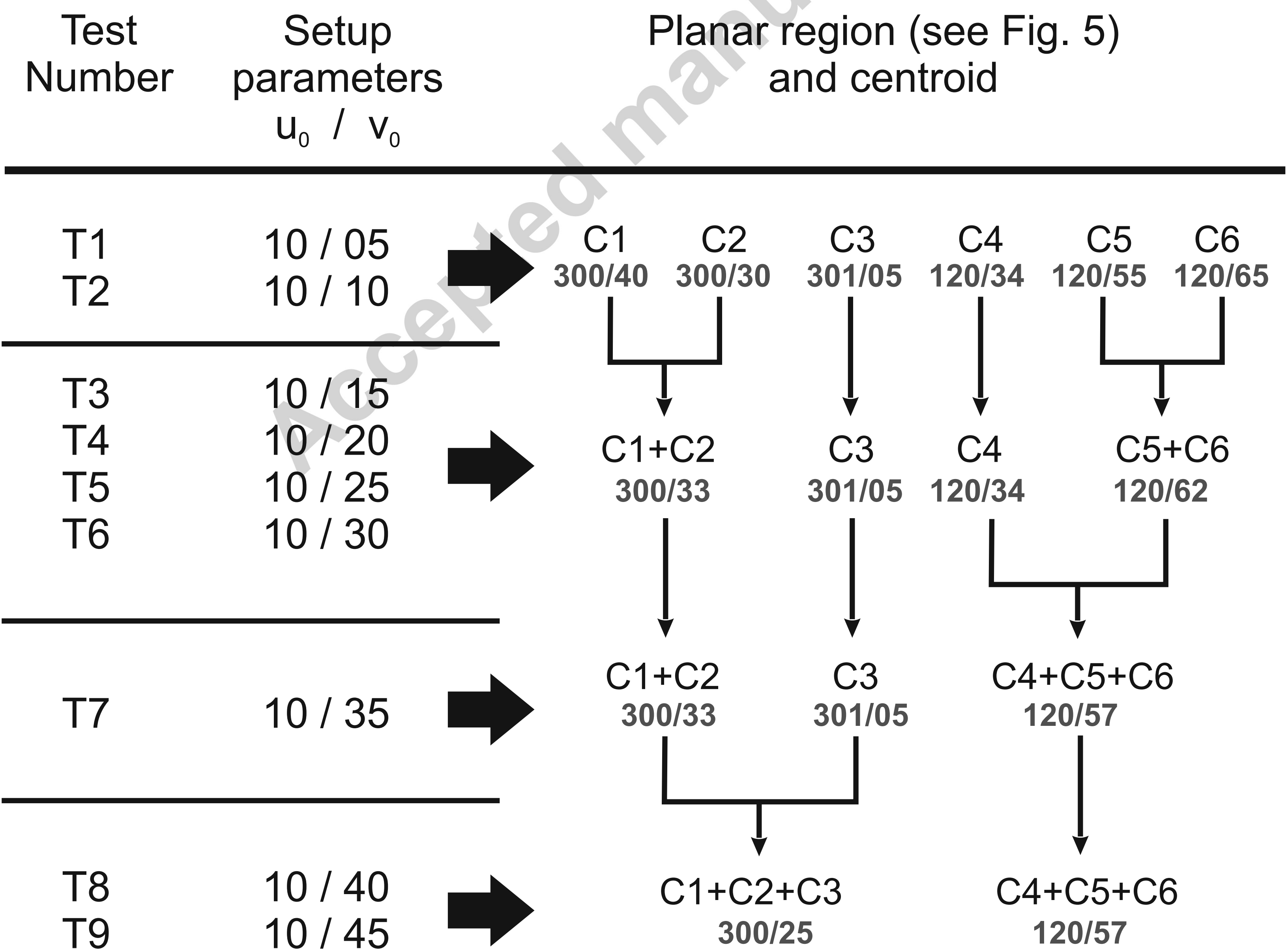
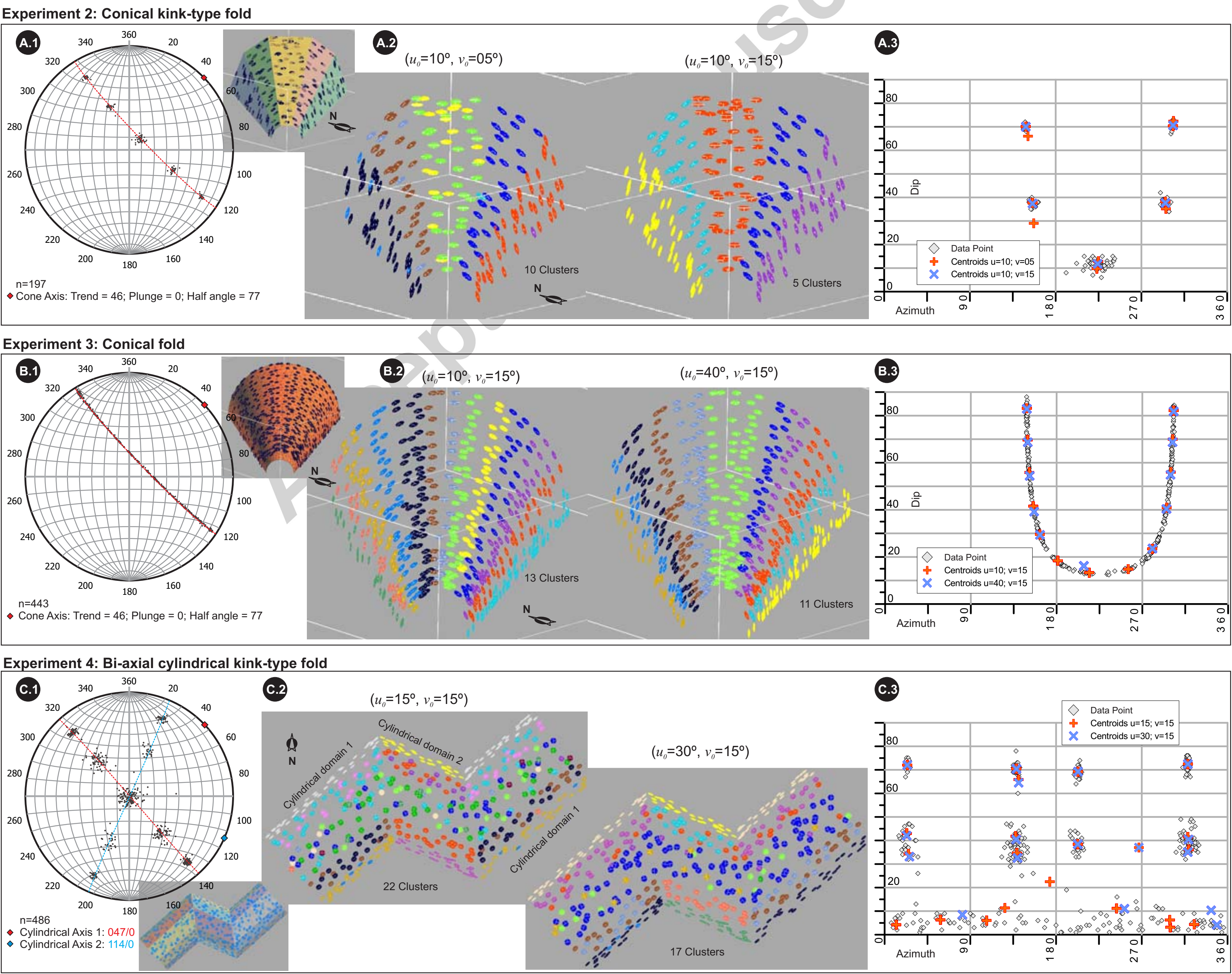

Test Number
Setup parameters
u_0 / v_0
Planar region (see Fig. 5) and centroid
T1
T2
10 / 05
10 / 10
C1
300/40
C2
300/30
C3
301/05
C4
120/34
C5
120/55
C6
120/65
T3
T4
T5
T6
10 / 15
10 / 20
10 / 25
10 / 30
C1+C2
300/33
C3
301/05
C4
120/34
C5+C6
120/62
T7
10 / 35
C1+C2
300/33
C3
301/05
C4+C5+C6
120/57
T8
T9
10 / 40
10 / 45
C1+C2+C3
300/25
C4+C5+C6
120/57

Figure 7



Experiment 2: Conical kink-type fold
A.1
n=197
Cone Axis: Trend = 46; Plunge = 0; Half angle = 77
A.2
(u_0=10º, v_0=05º)
10 Clusters
(u_0=10º, v_0=15º)
5 Clusters
A.3
Dip
Azimuth
Data Point
Centroids u=10; v=05
Centroids u=10; v=15
Experiment 3: Conical fold
B.1
n=443
Cone Axis: Trend = 46; Plunge = 0; Half angle = 77
B.2
(u_0=10º, v_0=15º)
13 Clusters
(u_0=40º, v_0=15º)
11 Clusters
B.3
Dip
Azimuth
Data Point
Centroids u=10; v=15
Centroids u=40; v=15
Experiment 4: Bi-axial cylindrical kink-type fold
C.1
n=486
Cylindrical Axis 1: 047/0
Cylindrical Axis 2: 114/0
C.2
(u_0=15º, v_0=15º)
Cylindrical domain 1
Cylindrical domain 2
Cylindrical domain 1
22 Clusters
(u_0=30º, v_0=15º)
17 Clusters
C.3
Data Point
Centroids u=15; v=15
Centroids u=30; v=15
Azimuth

Figure 8

A **Experiment 5:**
Cylindrical kink-type fold subset
(u_0=30º, v_0=15º)

B **Experiment 6:**
Bi-axial cylindrical kink-type fold subset
(u_0=30º, v_0=15º)

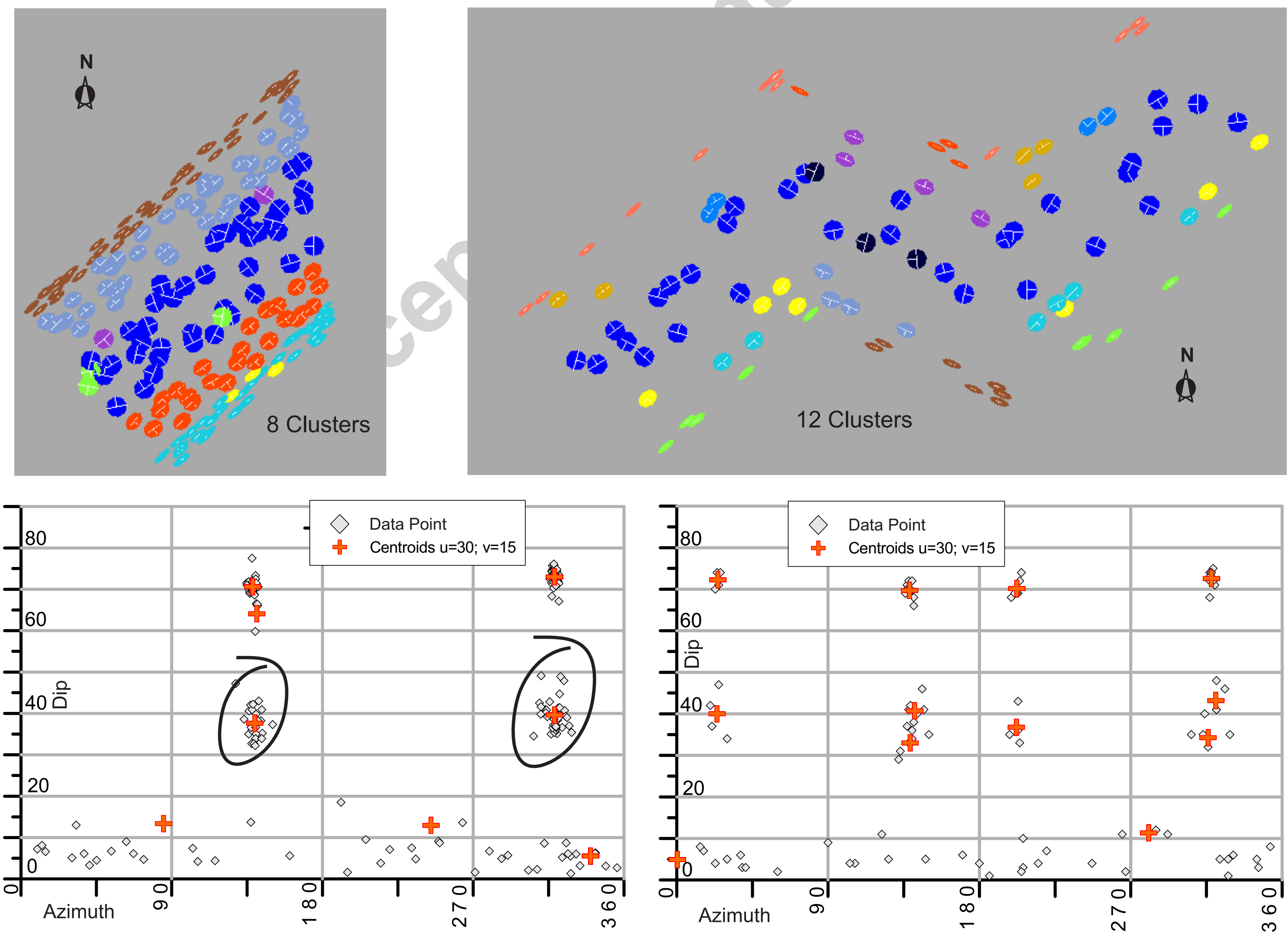


Figure 9

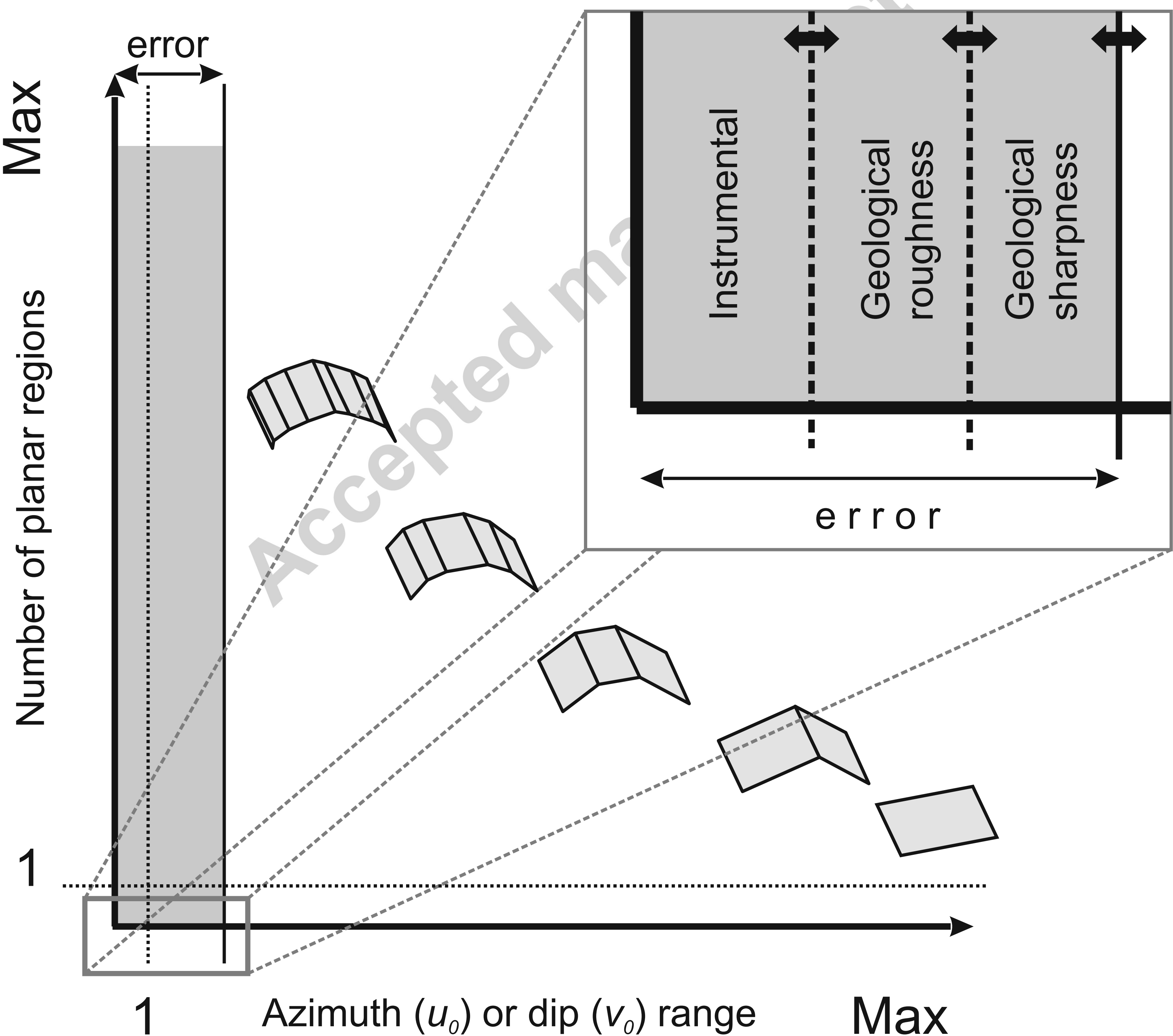
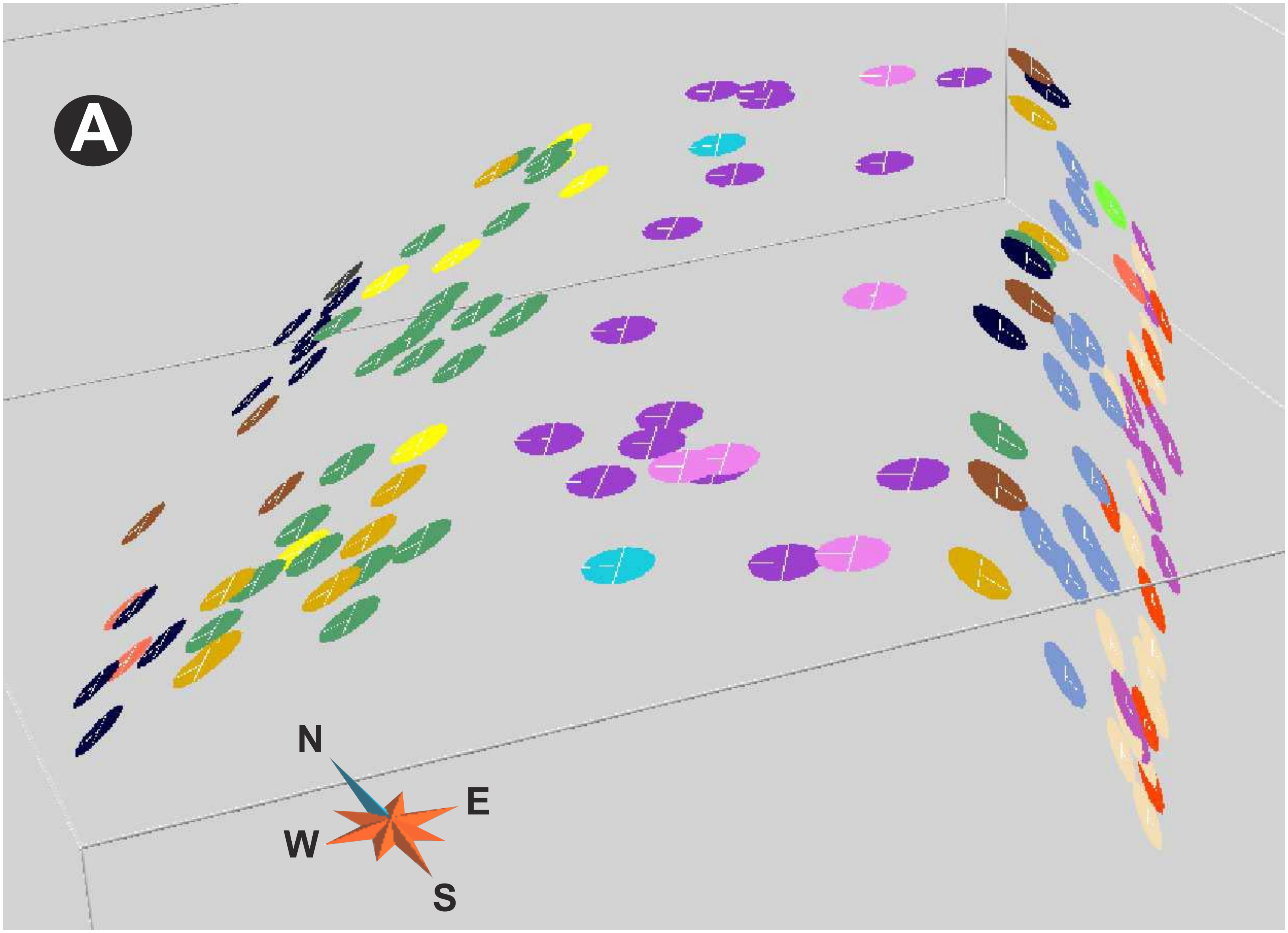

error
Max
Number of planar regions
1
1
Azimuth (u_0) or dip (v_0) range
Max
Instrumental
Geological roughness
Geological sharpness
e r r o r

Figure 10 colour



B

Setup parameters u_0 / v_0	Number of Planar Domains
10 / 1	21
10 / 0.5	29
10 / 0.1	39