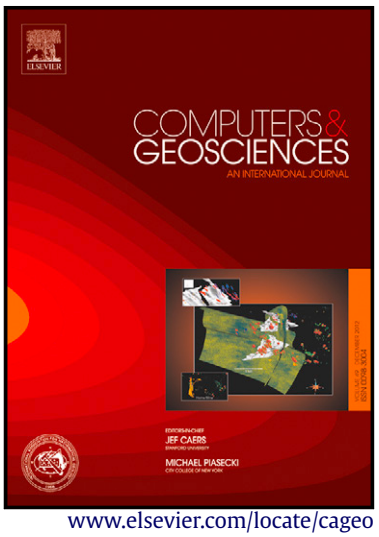


www.elsevier.com/locate/cageo

A new approach for semi-automatic rock mass joints recognition from 3D point clouds

A. Riquelme, A. Abellán, R. Tomás, M. Jaboyedoff

PII: S0098-3004(14)00074-0
DOI: http://dx.doi.org/10.1016/j.cageo.2014.03.014
Reference: CAGEO3355

To appear in: *Computers & Geosciences*

Received date: 12 October 2013
Revised date: 19 March 2014
Accepted date: 20 March 2014

Cite this article as: A. Riquelme, A. Abellán, R. Tomás, M. Jaboyedoff, A new approach for semi-automatic rock mass joints recognition from 3D point clouds, *Computers & Geosciences,* http://dx.doi.org/10.1016/j.cageo.2014.03.014

A new approach for semi-automatic rock mass joints recognition from 3D point clouds

A. Riquelme[1], A. Abellán[2], R. Tomás[1], M. Jaboyedoff[2]

(1) Departamento de Ingeniería Civil, Universidad de Alicante, Alicante, Spain

(2) Risk analysis group, Institut des sciences de la Terre (ISTE). Faculté des Géosciences et de l'Environnement. Université de Lausanne, Switzerland.

Abstract

Rock mass characterization requires a deep geometric understanding of the discontinuity sets affecting rock exposures. Recent advances in Light Detection and Ranging (LiDAR) instrumentation currently allow quick and accurate 3D data acquisition, yielding on the development of new methodologies for the automatic characterization of rock mass discontinuities. This paper presents a methodology for the identification and analysis of flat surfaces outcropping in a rocky slope using the 3D data obtained with LiDAR. This method identifies and defines the algebraic equations of the different planes of the rock slope surface by applying an analysis based on a neighbouring points coplanarity test, finding principal orientations by Kernel Density Estimation and identifying clusters by the Density-Based Scan Algorithm with Noise . Different sources of information —synthetic and 3D scanned data— were employed, performing a complete sensitivity analysis of the parameters in order to identify the optimal value of the variables of the proposed method. In addition, raw source files and obtained results are freely provided in order to allow to a more straightforward method comparison aiming to a more reproducible research.

Keyworkds: LiDAR, rock mass, discontinuities, semi-automatic detection, 3D point cloud, sensitivity analysis.

1. Introduction

Remote sensors such as Light Detection and Ranging (LiDAR) and Differential SAR Interferometry (DInSAR) have become an essential tool for the landslide analysis over the last decade (Abellán *et al.*, 2014; Jaboyedoff *et al.*, 2012; Oppikofer *et al.*, 2009; Rosser *et al.*, 2005; Viero *et al.*, 2010). LiDAR sensors, also known as laser scanners, allow the acquisition of high resolution (density of points up to 10^4 points/m^2) and high accuracy (std. dev. <1 cm at 100 m) three-dimensional information of the ground surface. Such systems allow obtaining the coordinates (X, Y, Z) of the points of a surface at high speed (~~up to~~ more than 222.000 measurements per second) from a considerable distance of acquisition (up to 6.000m). This sensor has revolutionized the acquisition of rock slope parameters that play a key role in the global and local stability including the orientation, spacing, persistence and roughness of ~~the~~ discontinuities. Not surprisingly, the number of publications dealing with the semi-automatic extraction of 3D features has exponentially grown in the last five years (García-Sellés *et al.*, 2011; Gigli and Casagli, 2011; Jaboyedoff *et al.*, 2007; Khoshelham *et al.*, 2011; Lato *et al.*, 2009; Lato *et al.*, 2010; Lato and Vöge, 2012; Olariu *et al.*, 2008; Slob *et al.*, 2005; Sturzenegger and Stead, 2009b; Sturzenegger *et al.*, 2011). Nevertheless~~, Nevertheless~~, in order to enable fast advancement in the application of the sensor in disciplines such as rock mechanics, geotechnics and earth sciences, development of new algorithms is needed (Abellán *et al.*, 2014).

This paper proposes a new approach for the semi-automatic identification and extraction of rock slope planar features —i.e. the discontinuity sets affecting rock mass stability— using 3D point cloud data. The main novel contributions of the proposed method are: (a) the user-supervised removal of noisy points through the creation of a coplanarity test; (b) the semi-automatic identification of discontinuity sets using a Kernel Density Estimation (KDE) Analysis; (c) The automatic extraction of single discontinuities through a density-based clustering algorithm; (d) a complete sensitivity analysis of the parameters playing a key role in the method; and (e) the public availability of the complete 3D RAW and processed data sets used in this publication in order to provide method validation for other researchers in www.3d-landslide.com/projects/discontinuity/

1.1. Previous studies on discontinuity characterization from 3D point clouds.

Rock slope discontinuities play a key role in strength, permeability of rock masses and in the stability of surface and underground excavations (Harrison and Hudson, 2000; Hoek and Bray, 1981). Thus, a thorough understanding of the properties of discontinuities, included their orientation (i.e. dip and dip direction) is crucial in rock engineering applications.

In order to assess the global ~~s~~ quality of a rock mass, several authors proposed the use of geomechanical classifications more than twenty years ago. Rock mass classifications are means for the evaluation of the performance of rock masses based on their most important inherent and structural parameters (Pantelidis, 2009). In practice, a wide number of geomechanical classifications for slopes exist such as those proposed by Bieniawski (1989), Romana (1985), Hack et al. (2003) and Tomás (2007). These classifications require precise information of a series of slope parameters —such as discontinuities orientation, ~~length and persistence~~ —, which are classically obtained in tedious fieldwork campaigns using a geological compass. Some well-known techniques, such as the stereo photogrammetry, have allowed the measurement of orientations of individual discontinuities since the 1970s' ~~70's decade~~ (Rengers, 1967). In addition, basic photogrammetry principles and pattern recognition routines can be used to model surfaces in 3D, which can be very useful in the rock mechanics field. Unfortunately, these techniques require tedious and time consuming outlining of discontinuities (Slob *et al.*, 2005).

At the beginning of the XXI century, some authors suggested the possibility of accurately obtaining discontinuity orientation from 3D point clouds obtained by a total station (Feng *et al.*, 2001). Since then, and thanks to the wide accessibility of 3D sensors like LiDAR, different approaches were developed for obtaining the orientations of ~~discontinuity~~ discontinuities. Early studies proposed the use of least square method to a subset of points (Abellán *et al.*, 2006; Fernández, 2005; Sturzenegger and Stead, 2009a). Some other authors proposed the calculation of normal vectors to a series of 2.5D interpolated surfaces (Kemeny *et al.*, 2006a; Slob and Hack, 2004). Recently, the calculation of the normal vector associated to a subset of the 3D point cloud is widely accepted (Ferrero *et al.*, 2009; García-Sellés *et al.*, 2011; Gigli and Casagli, 2011;

Jaboyedoff *et al.*, 2007). More specifically, Jaboyedoff et al. (2007) proposed the calculation of the normal vector orientation for every point and its coplanar neighbours using the principal component analysis method (hereinafter PCA) This concept is also used to isolate multi-scale objects from LiDAR data (Ioannou, 2012). Other approaches calculate the orientation for each node in the TIN (Slob *et al.*, 2005; Vöge *et al.*, 2013) or are based on the searching of volumetric pixels (voxels) and subsequent calculation of the planar orientation (Gigli and Casagli, 2011). Remarkably, any of the above mentioned studies utilise kernels for the estimation of the density function, meaning that those points belonging to less sampled discontinuity sets can potentially be overlooked using commonly used methods.

~~The calculation of the normal vector orientation requires a previous set of points definition.~~ Most of the current discontinuity detection methods use triangulated irregular network (TIN) to simplify the surface (Gigli and Casagli, 2011; Lato *et al.*, 2009; Slob *et al.*, 2007). ~~Reversely~~ Conversely, our proposal uses ~~each 3D point of the point cloud~~ real 3D information contained in every point and its corresponding neighbours to see the local differences in ~~identify the different sets controlling~~ the geometry of the slope.

Some authors offer ~~a~~ commercial software packages, such as ~~the pioneer~~ Split-FX (Slob *et al.*, 2005) and Coltop-3D (Jaboyedoff *et al.*, 2007). Some recent studies include the use of a Graphic User Interface (GUI) in Matlab environment such as the recently developed DiAna (Gigli and Casagli, 2011) or PlaneDetect (Vöge *et al.*, 2013), but the use of these software is not publicly available. Other applications for the geomechanical classifications include: (a) the automatic detection of discontinuity spacing (Slob and Hack, 2004; Slob *et al.*, 2005), which is based on the cluster analysis of sets of discontinuities (Roncella and Forlani, 2005; Turner *et al.*, 2006); (b) the removal of objects characterized by chaotic shapes—such as vegetation—together with the calculation of other parameters of the geomechanical classifications —such as spacing/frequency and persistence—which can also be (potentially) achieved using tools such as 3D-Veros (Brodu and Lague, 2012) and DiAna (Gigli and Casagli, 2011) ~~softwares~~. Unfortunately, only a limited number of benchmarks is publicly available — such as the Rockbech common repository described in Lato et al. (2013), so there is a need for a comparative performance analysis of the existing algorithms mentioned in this manuscript.

The paper is organized as follows: (a) an introduction to LiDAR techniques and their application to discontinuity extraction is presented in section 1; (b) the methodology for discontinuity extraction and the presentation of the case studies used in this paper are presented in section 2; (c) Section 3 shows a sensitivity analysis of the method using simple geometries (case study A); Section 4 shows the application of our method to a more complex scenario (road cut slope, case study B). In addition, the methods' parameters are calibrated and then their processing parameters values are proposed. Finally, section 5 discusses and summarizes the results and explores the future lines of research.

2. Methodology

The proposed method aims to detect ~~planes that form~~ ~~the~~ structural discontinuities ~~in~~ $\mathbb{R}^3$ ~~–~~using 3D point clouds than can be typically obtained from LiDAR sensors, 3D digitizers, etc. Unlike other methodologies, our proposal uses, throughout the workflow, ~~along the workflow~~ the "true" 3D information contained on the LiDAR point cloud, instead of using interpolated 2.5D mesh surface. ~~The discontinuities are identified and defined as planes. In addition, each single point is assigned to a discontinuity or plane so all the LiDAR information is maintained.~~ Given ~~Thereby, given~~ the set of raw data points (X, Y, Z) from the observed scene (hereinafter 'P'), if the slope surface is mostly defined by discontinuities, the outcrop points can be appropriately ordered into sets which define planes. These planes define the discontinuity sets.

The method basically performs a compass data acquisition for each point, but only if it is surrounded by other coplanar points. Therefore, there is an obvious advantage: it is possible to obtain millions of virtual compass measurements ~~lectures~~ in a few minutes, even in otherwise ~~in~~ non-accessible areas.

The proposed methodology is developed through three main steps (Figure 1):

a) PART A - Local curvature calculation: ~~Consisting in~~ this consists of a nearest neighbour search~~ing~~ and ~~in~~ the determination of the discontinuity orientation in every point. This task is described in section 2.2.
b) PART B - Statistical analysis of the planes: this consists of ~~consisting in~~ the determination of the principal orientations, ~~main orientation of~~ which represent

~~are~~ the different discontinuities sets ~~affecting~~ that affect to the rock mass. The next step is ~~and in~~ the identification of those points that belong to a common discontinuity set. This part, developed in section 2.3., requires the user's supervision.

c) PART C- Cluster analysis: localization of the points that define different clusters in the space and calculation of the outcrop plane equations. This last part is explained in section 2.4.

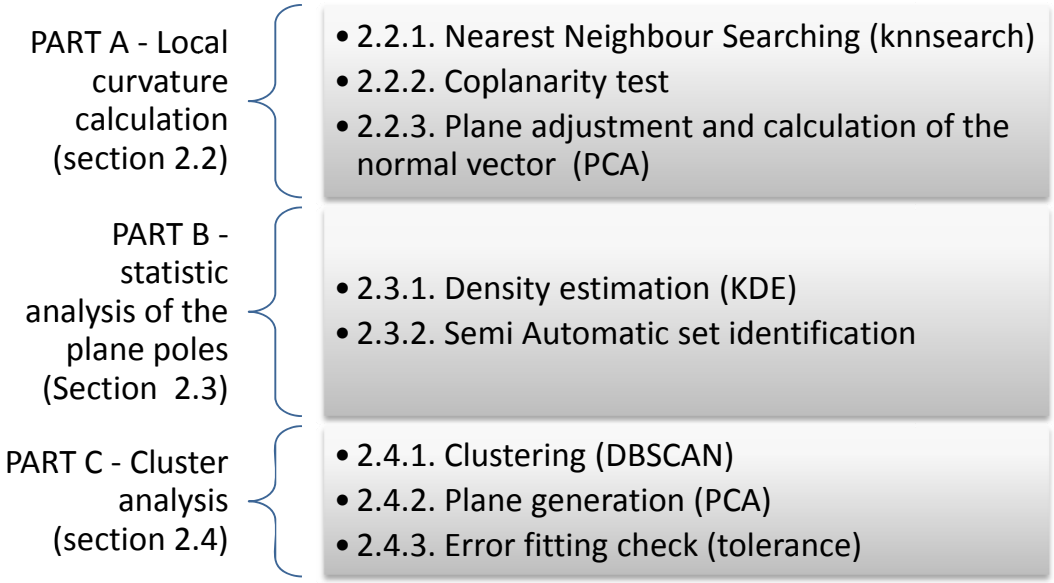


Figure 1: Flow chart of the proposed methodology.

2.1. Description of the datasets

Two different series of 3D datasets were ~~employed~~ used in our study: experimental datasets and real outcrop measurements. The first was obtained under controlled laboratory conditions and the second one is a more complex dataset corresponding to a portion of a real rock mass. We discarded using synthetic datasets due to ~~its~~ their over-simplistic characteristics.

2.1.1. Case study A

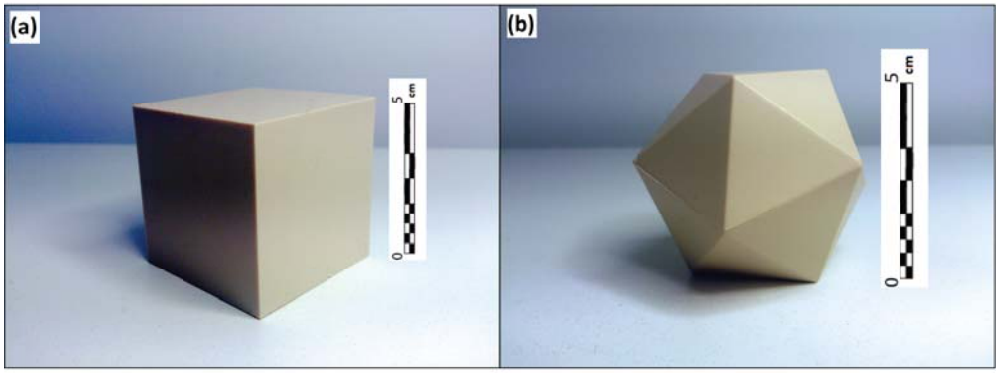
We first scanned a series of well- known geometrical solid objects using a 3D digitizer (Konica Minolta, *Vivid* 9i) from University of Lausanne, Switzerland), including a cube, dodecahedron, icosahedron, octahedron, hexagonal pyramid, hexagonal prism,

176 octagonal prism and a triangular prism. Data acquisition was performed through
177 progressive rotation of the figures around a fixed platform axis and a subsequent
178 scanning. We carried out a total of 10 ~~scanners~~ scans with a mean distance of 1406 mm
179 to the figure. The Line of Sight of the 3D digitizer was inclined about 30° to zenith. As
180 a consequence: (a) the density of points on one of the families, the horizontal planes,
181 was higher than on the other families due to the superposition of different scans, which
182 lead to an overrepresentation of these planes; (b) a higher alignment error was also
183 observed in these overlaid planes. Both effects are consistent with TLS data acquisition
184 in real case studies.

185 Then, from these figures, we selected two representative geometries: a cube and an
186 icosahedron (Figure 2a and b, respectively). The cubic geometric shape, which is
187 formed by 6 square facets grouped on three orthogonal discontinuity sets, was
188 represented by 60.488 points. The icosahedron, which is a type of polyhedron formed
189 by 20 triangular facets grouped on 10 different discontinuity sets, was represented by
190 37.226 points.

191 These simple geometries allowed the comparison of our algorithm with ~~the~~ true
192 geometries known in advance. As the data was acquired under laboratory controlled
193 conditions, it was possible to evaluate the quality of the methodology through
194 comparing our results with the real plane orientations.

195 Finally, as these figures are formed as a combination of perfectly plane surfaces, it was
196 possible to test the values of the standard deviation of the error in each single plane~~. In~~
197 ~~addition, the performed tests have allowed the identification of those parameters that~~
198 ~~proved inappropriate for some cases due to alignments.~~ The normal vectors calculated at
199 each of the 3D points allowed us to properly identify not only the normal vectors
200 corresponding to flat surfaces such as discontinuities, but also the normal vectors
201 corresponding to non-flat regions ~~surfaces~~, such as the vertex and the edges between
202 planes.



204

206 **Figure 2. Experimental datasets: scanned geometries employed in the case study A. (a) cube (50 mm side); (b)**
207 **icosahedron (50 mm height).**

207

208 ## 2.1.2. Case study B

209 The application of our method to a real case study was carried out using data from
210 publicly available LiDAR data at Rockbench repository (Lato *et al.*, 2013), providing
217 the possibility to compare our results performance to other researchers. In addition, this
218 case study has been used by other authors in several published papers for extraction of
219 rock mass characterization information (Kemeny *et al.*, 2006b). This case study consists
220 in a real rock cut located in Ouray, Colorado, USA (Figure 3 and Table 1). As will be
221 described later, in this case study, our method detected four different discontinuity sets.

216 **Table 1. Properties of the Case study B datasets**

Physical Setting	Roadcut
Location (close mjr. City)	Ouray, Colorado
Lithology	Quartzite
Scanner	Optech
Laser Type	Time of flight
Year scanned	2004
# of scan locations	4
Point spacing	< 2 cm
Number of points	1,515,722
Collected by	John Kemeny
Dataset	10a

217

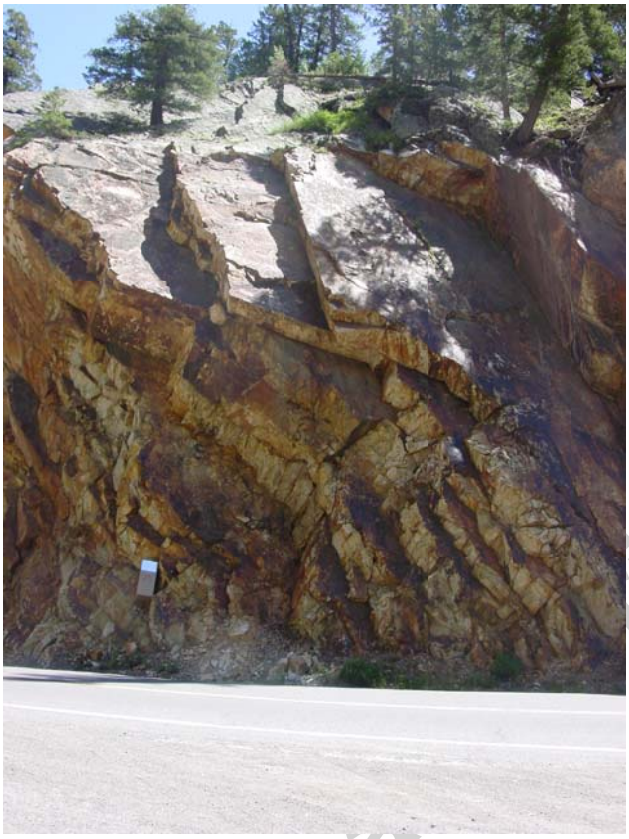


Figure 3 – Real road cut slope used in case study B. Image from Rockbench repository (Dataset 10a).

2.2. Part A –Local curvature calculation

The method requires as input the raw data points P, where (P_i) is a point member of P. Given a subset of neighbour points Q_i (where (P_i) is a member of Q_i and the size of Q_i is n_n points), it is possible to calculate ~~the~~ its best-fit adjustment plane α ~~for the subset Q_i~~ (Figure 4).

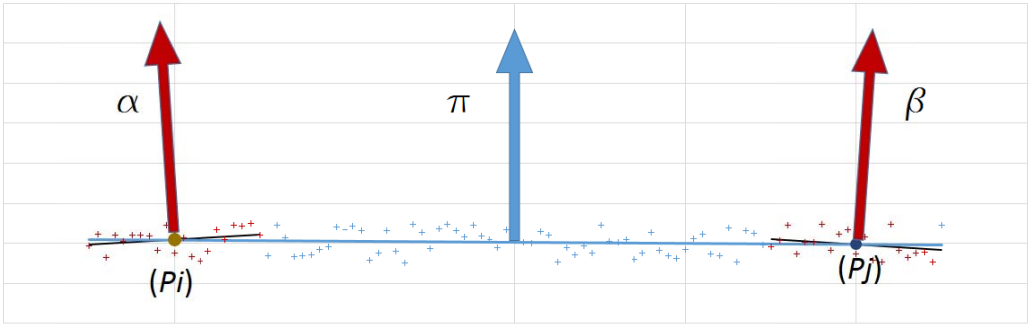


Figure 4. Subsets and normal vector orientations. The sets Q_i (left) and Q_j (right) are defined by the points (P_i) and (P_j) and their respective neighbourhoods. α, and ρ are the orientations of their respective sets *Qi, Qj* and *P* .

The proposed method calculates a normal vector for each 3D point. ~~The most representative orientations are orientation is considered as the π plane orientation.~~ Thus, the above-described discontinuity planes identification is performed through three main phases:

a. For each i point of the raw data (P_i), the *K*-nearest neighbours (*knn*) have to be found in order to create the set Q_i (subsection 2.2.1).
b. For each set Q_i the coplanarity condition has to be checked (subsection 2.2.2).
c. For each set Q_i a plane adjustment has to be performed in order to calculate its normal vector (subsection 2.2.3.).

These phases are described in detail in the next subsections.

2.2.1. Nearest Neighbour Searching

The search ~~for P_i~~ of neighbours is usually carried out using two different approaches: fixed distance ~~definition – the distance from P_i to q is less or equal to a search radius, r,which is user-defined –~~; or fixed number of neighbours ~~definition – in which the point q is one of the knn nearest points to P_i –~~. Some errors may arise when using the first approach due to the heterogeneity of the density of points (Lato *et al.*, 2010). ~~Reversely~~ Thus, a fixed number of neighbours approach was preferred in our study.

The MATLAB function *knnsearch* uses an algorithm that provides a quick and efficient way to find the *knn* nearest neighbours by a selected norm (Friedman *et al.*, 1977). In the proposed approach, the *knn* neighbours are calculated by using *knn* search function and the euclidean distance. Thus, after this step, for each i point of the raw data P_i, a subset of *knn* neighbour points is defined as Q_i.

Summarizing, in this section we have identified the k nearest neighbours for each point of the 3D point cloud. The next step is to check if that set of k+1 points are coplanar or not.

2.2.2. Coplanarity test

Due to the fact that the method considers every point and its neighbours as a plane subset candidate, it is advisable to test if the Q_i sub-set of points (defined in previous steps) is coplanar (or not). This validation test must be carried out prior to α orientation

calculation (Q_i associated). If the sub-set of points Q_i is coplanar, the rest of the process will continue; otherwise the sub-set Q_i will be rejected ~~for further analysis~~.

The coplanarity test is based on the Principal Component Analysis (PCA). Given a 3D set of points, the *princomp* MATLAB function, which allows the implementation of PCA, determines its eigenvalues ($\lambda_1,\lambda_2,\lambda_3$) and eigenvectors ($V_1,V_2,V_3$). The proportion of variance accounted by the first k components H_k is determined by eq.(1)~~, while the unexplained variance is determined by eq. (2)~~ (Rencher and Christensen, 2012):

$$H_k = \frac{\sum_1^k \lambda_i}{\sum_1^3 \lambda_i} \quad (1)$$

Assuming that a portion of our measurements are arranged in a plane (π) in a $\mathbb{R}^3$ space, there will be two dimensions able to explain the majority of the data. ~~Thus, with k=2, the proportion of variance explained by the first two dimensions will be close to 1.~~ The third ~~eigenvector~~ dimension will explain the error present in the data. If the surface is not flat or the instrumental error is relevant enough, the third dimension will acquire relative importance to the first two.

In order to know if a set of points is coplanar or not, the deviation parameter (η) is defined by eq (2):

$$= \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3} \quad (2)$$

The parameter tolerance (η_{max}) is defined as the maximum allowable deviation in a subset of points, such that the subset plane is reasonably considered a plane. The η_{max} value is established through a sensitivity analysis with real data under certain test conditions. It is commonly accepted that if a set of principal components have 80% or more of the variance, these would represent the data properly ~~As a rule of thumb an 80% of percentage of the total of variance indicates that the data is enough represented~~(Rencher and Christensen, 2012). Hence, a η_{max} value of 20% is proposed. In those cases in which $\eta > \eta_{max}$, the sub-set is rejected ~~for further analysis~~.

Summarizing, in this section we have identified those points which are coplanar with their nearest neighbours. The next step is to calculate the orientation of the coplanar set of points.

2.2.3. Plane adjustment and calculation of the normal vector

Note that, subsequently, a parameters calibration will be performed in this paper. Once all those subsets of coplanar points have been found, the next step is the calculation of the best-fit adjustment plane. The algebraic expression is shown in eq (3), where A, B, C are the three components of the unit normal vector to the plane and D gives the perpendicular distance from the origin to the plane.

$$Ax + By + Cz + D = 0 \quad [A, B, C, D] \in \mathbb{R} \tag{3}$$

Some authors such as Gigli and Casagli (2011) calculate the plane equation by the singular value decomposition (SVD). In our case, since the PCA has been calculated in a previous step ~~of our analysis~~, the plane is defined in a more efficient way through the eigenvector $\vec{V}_3$ **(4).**

$$\vec{\boldsymbol{V}}_3 \equiv (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}) \tag{4}$$

Summarizing, at this section, we have computed the orientation of the previously identified sets. The next step is to calculate the most representative orientations of the 3D points and their k nearest neighbours.

2.3. Part B: statistical analysis of the planes

The subsequent methodology ~~This part~~ is based on the expected parallelism of the normal vectors associated to the points. Let's consider a set of points Q_i associated to a point (P_i) that belongs to a discontinuity defined by an unknown plane π. If α is the best-fit plane of Q_i, the orientations of the planes α and π ~~must~~ are expected to be close. Similarly, let's consider a different subset Q_j with the same size than Q_i associated to a point (P_j) member of P. This set is-also part of the π plane and it is possible to calculate

its best-fit plane β *and* the three planes (π, α and β) having a close orientation (Figure 4).

The statistical analysis ~~of the density of the poles was~~ is performed by means of the stereographic projection of the planes poles. In order to define the main discontinuity sets: (a) we calculated a normal vector for each plane and converted it to stereographic projection (Lisle, 2000); and (b) we calculated the density of the poles for each region of the stereographic projection; and (c) we calculated the local maxima.

2.3.1. Density estimation

Kernel density estimation (KDE) is a non-parametric way to estimate the probability density function of a random variable. In order to estimate the multivariable nonparametric density function, Silverman *(1986)* demonstrated a higher performance using KDE than using classic histograms. Thus, the method implementation uses the Matlab kde function *kde2d* (Botev *et al.*, 2010) by a Gaussian kernel. This script allows: (a) the automatic calculation of the width of the kernels (e.g. bandwidth); and (b) the computation of their density.

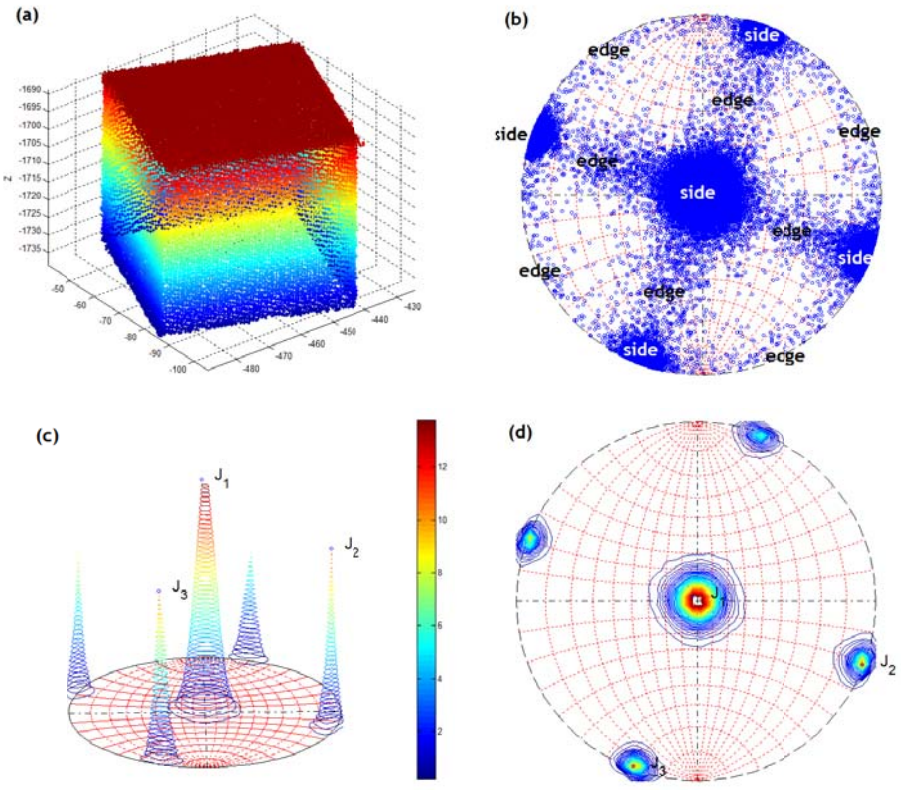


Figure 5. (a) Raw Data 3D view of a 5 cm side cube, 60.488 points. Plotted data were scanned at laboratory using a microlidar; (b) Normal vector poles stereographic projection, knn=15. Side and edge poles zones are labelled.(c) and (d) Density estimation via kernels, isolines each 2%. ~~Notice~~ Note that the identification of the main discontinuity sets is able to filter out the normal vectors calculated at the edges between planes.

Figure 5b shows the stereographic projections of the poles of the normal vector of an experimental dataset consisting in a five centimetres side cube scanned at laboratory using a 3D digitizer (Vivid 3D, Konica Minolta). Note that for the cube shown in Figure 5b the poles show three main orthogonal discontinuity sets (J1: 223.87°/4.07°; J2: 021.03°/89.47°, J3:290.91°/89.62°). Figure 5c also shows the calculated density function using the kde method. In this figure, the normal vector poles are clearly clustered into three orthogonal discontinuity sets as it was expected~~, so it is needed to calculate the stereographic projection coordinates of the most representative poles.~~

At this point, the density of the poles is known. Therefore, we can identify the peaks which reasonably represent the orientations of the 3D point cloud and its neighbours.

2.3.2. Semi-automatic set identification

In this step, the method assigns a principal orientation to every single point in the point cloud. If the method detects that the point is not represented by any principal orientation, there will be no assignment.

~~Once the normal vector is calculated and the principal orientations have been defined, the next logical step relies on labelling each point with its corresponding main family~~. ~~Those points whose estimated planes do not belong to any discontinuity set are not assigned.~~ The scheme of this step is summarized in Figure 6. Usually, the density function analysis shows many local maximums, but only a few are principal poles, which is due to the fact that the existence of reading errors and singular points of curved surfaces imply the dispersion of the poles. Hence, two requirements, which can be user-supervised, allow us to define a local maximum as a principal pole:

a) Condition num. 1 (Cone filter): the user defines a certain value (γ_1~~). which is smaller than or equal to~~ The angle formed by two principal vectors must be higher than this value ~~form an inter-normal angle~~.

b) Condition num. 2 (max. poles filter): The user indicates the maximum number of discontinuity sets (n_p) that can be established. The system filters and accepts the n_p principal planes with higher density.

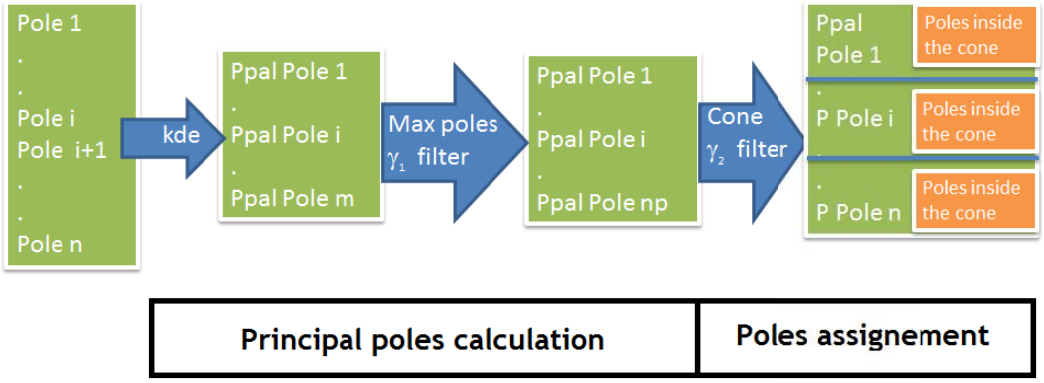


Figure 6. Scheme of the discontinuity set calculation.

Figure 7 shows the poles density function of ~~an~~ a cube in which peaks are numbered from the highest to the lowest value of the density function (*knn*=10 and η_{max}=20%). In Figure 7a the density function obtained applying no filters shows many local maxima (labelled from 1 to ~~20~~ 9). By accepting a minimum γ_1 value of 20º, a cleaner plot of the principal planes is obtained ~~(Figure 7b)~~. Specifically, the relevance of discontinuity sets 1, 2 and 3 on the other is obvious, thus the maximum number of discontinuity sets is set to 3 (Figure 7~~c~~ b).

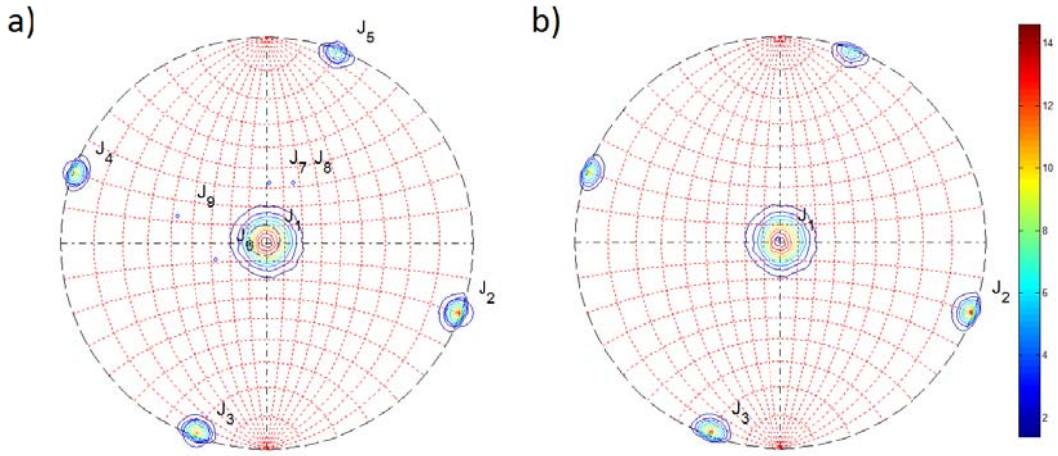


Figure 7. Poles density of the cube shown in figure 2, (a) non-filtered~~(b) option num 1, cone filter using~~

~~γ_1=20º(c)~~(b)cone filter (γ_1=20º) and max. poles filter(n_p=3). ~~Notice~~ Note that the labels (J_1 to J_9) indicate the location of the calculated relative maximums. Isolines are plotted each 2%.

The next step consists ~~in~~ of the segmentation of the point cloud: we assign a label to each point of the point cloud according to the closest principal families. For every single point, we look for the discontinuity set that provides the minimum angle (γ) between the associated normal vector and the assigned principal plane normal vector. A threshold is then defined in order to limit the maximum allowed value (γ_2). As an example, Figure 8 shows the application of these criteria for the recognition of the discontinuity sets of the cube. Points are classified according to their closest poles in Fig 8a; a threshold is then defined in Fig 8b (γ_2 = 30º), classified points are automatically filtered out.

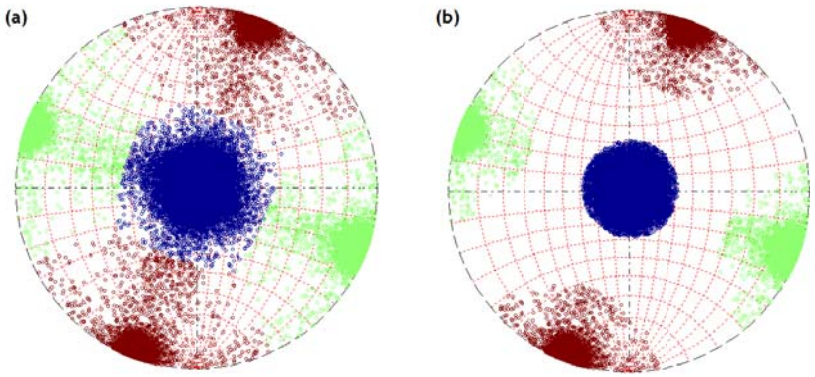


Figure 8. Stereographic representation of the principal poles assignment of the cube dataset (a) Non-filtered poles, 59.705 poles; (b) γ=30º cone filtered, 57.134 poles. ~~Notice~~ Note that this step is able to filter out the normal vectors calculated at the edges between planes.

At this point, we have identified the principal orientations of the 3D point cloud and its neighbours. The corresponding orientation has been assigned to each point depending on the point and its neighbours' orientation. We then discarded those points having an orientation considerably different to any principal pole (i.e. higher than γ). Since we have extracted the points belonging to a discontinuity set, the next step is to identify the 3D point clusters of each DS.

2.4. Part C: Cluster analysis

Given a discontinuity set i it is necessary to find its data subset R_i whose points are grouped according to planar clusters (Figure 9).The obtained clusters are members of the discontinuity set *I* which are defined in the space through ~~hits~~ its plane equation (3).

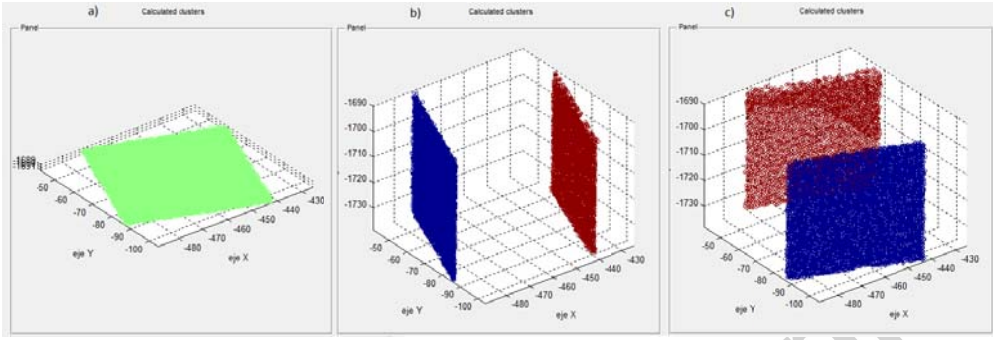


Figure 9**.** Identification of the different clusters for the three automatically recognized sets of planes of the cube shown in figure 2. ~~Notice~~ Note (left figure) that only the upper face of the hexaedron is recognized by the software because no points are available from the lower face, which is a shadow area for LiDAR.

2.4.1. Clustering

For the clustering of the 3D datasets, we employed the "Density-Based Scan Algorithm with Noise" (DBSCAN) (Ester *et al.*, 1996). This clustering algorithm for class identification in spatial databases has been proven in a previous successful application in the processing of LiDAR point clouds (Tonini and Abellan, 2014). This algorithm requires as input the following parameters: (a) ε, which is the maximum distance between two points to consider them as neighbours; (b) *Min-pts,* which is the minimum number of neighbours of point *q* to consider *q* as a core point. It is known that large differences in densities might affect the application of automatic methods ~~applications~~ (Ester *et al.*, 1996; Lato *et al.*, 2010). Although in order to develop this algorithm it is assumed that density of the point cloud is homogeneous, the effects of heterogeneous density of measurements on the results will be analysed later.

Ester et al. (1996) recommend to set the parameter *min-pts* equal to 4. In addition, ε should be determined considering the distance of the 4^{th} neighbour for each discontinuity set. The method calculates all the 4^{th} neighbour distances for all the accepted points members of a ~~discontinuity set~~ principal plane. ε can be considered as the mean, maximum or other statistically representative values. Maximum value should be inappropriate because of the existence of noisy extreme values. As the method works

with raw data, the number of distances is high enough to consider a normal distribution, so the proposed ε value is the 4^{th} neighbour distances mean plus two standard deviations.

A real case cluster analysis may find a high number of small clusters. It is possible that the user is only interested in big clusters so the method offers the option of discarding small clusters stating a selection threshold named parameter points per cluster, hereinafter *ppc*. Thus, only clusters sized by a number of points equal or higher than ppc will be in the output.

At this point, we have calculated the spatial clusters of each discontinuity set. The next step is to calculate the plane equations of these discontinuities to mathematically define the planes.

2.4.2. Plane generation

Plane generation is carried out as follows: given (a) a set of points which belong to a discontinuity set i —hereinafter R_i— and (b) a set of points members of a cluster *j* which constitute a subset of R_i —hereinafter R_{ij}—, then (c) we will find the best-fit plane of R_{ij}, which plane equation can be defined by the algebraic expression (5):

$$A_{ij} \times x + B_{ij} \times y + C_{ij} \times z + D_{ij} = 0 \quad \textbf{(5)}$$

We apply the PCA to the *k* points P_{ij}^{k} which have coordinates (x_{ij}^{k}, y_{ij}^{k}, z_{ij}^{k}),and are members of the no empty set R_{ij}, obtaining the principal vectors $\bar{V}_3$ .The parameters A_{ij}, B_{ij} and C_{ij} are calculated using eq.(4). It is also possible to calculate these parameters using the normal vector of the discontinuity set principal pole, so all the clusters will exactly have the same orientation. The independent term D_{ij} of the plane equation (3) is computed by the least square method which is mathematically defined by (6)

$$D_{ij} = -\frac{A_{ij}}{n}\sum_{k=1}^{n} x_{ij}^{\ k} - \frac{B_{ij}}{n}\sum_{k=1}^{n} y_{ij}^{\ k} - \frac{C_{ij}}{n}\sum_{k=1}^{n} z_{ij}^{\ k} \quad \textbf{(6)}$$

Where n is the cluster size of R_{ij}. The relations between the indexes are shown in Figure 10.

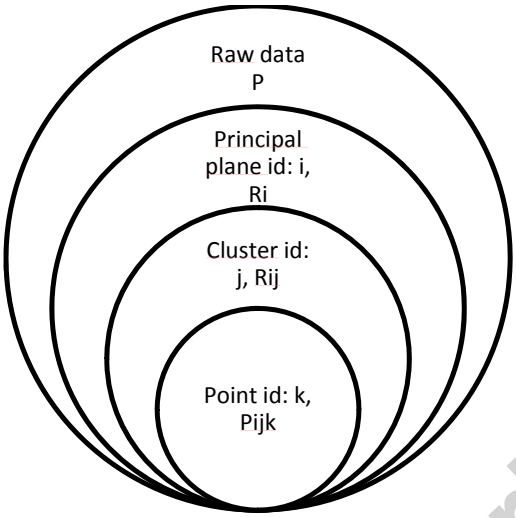


Figure 10: Relations between indexes.

2.4.3. Error fitting checking

Once the R_{ij} plane equation has been calculated, it is convenient to check the quality of the data fitting. Given that $\vec{V}_3 = 1$ ~~as in eq.(4)~~, the fitting error er_{ij} is defined as the point plane distance (7).

$$er_{ij}^{\ k} = Ax_{ijk}^{\ k} + By_{ijk}^{\ k} + Cz_{ijk}^{\ k} + D \quad (7)$$

Therefore the errors er_{ij} associated to the cluster R_{ij} can be defined by eq.(7). The set er_{ij} must satisfy two characteristics: the value of its module $\left|\vec{er}_{ij}\right|$ must be minimum (this will be satisfied as the equation is calculated by the least mean square method) and the value of its standard deviation $\sigma(er_{ij})$ must be reasonably low ~~small enough~~.

3. Results for case study A: Sensitivity analysis and calibration ~~(case study A)~~

A sensitivity test and calibration of the proposed methodology, consisting in finding the proper parameter values in order to obtain satisfactory results, is discussed in detail in next subsections. The main steps followed for the calibration are: (a) definition of an

experimental test and scanning of regular geometrical figures; (b) analysis of the surface for planar sides and for non-planar surfaces such as edges and vertexes; (c) test of the method with all the figures separately using the chosen values; (d) analysis and discussion of the results.

3.1. Coplanarity test calibration: influence on the number of neighbours

The number of neighbours *knn* and the maximum deviation η (2) are the first parameters used for the coplanarity test calibration. Given a planar surface and its scan data P_i, the deviation η for each subset Q_i member of P_i can be calculated. This process is carried out changing the value of *knn*, thus η_{max} can be analysed. It is very important to set a value of η_{max} that discards only real noise data in order to avoid the loss of valuable information. A small value of η_{max} may consider normal points as noise due to instrumental error, surface roughness or surface curvature.

The performed sensibility test (Figure 11) uses the PCA to each Q_i, which is a subset with j neighbours. For each point i and its neighbours j, the deviation η_{ij} (eq.(2)) is calculated. In order to avoid outliers the 1% lower and upper tails are removed. The final step is calculating the mean $E(\eta_{ij})$ and the standard deviation $\sigma(\eta_{ij})$.

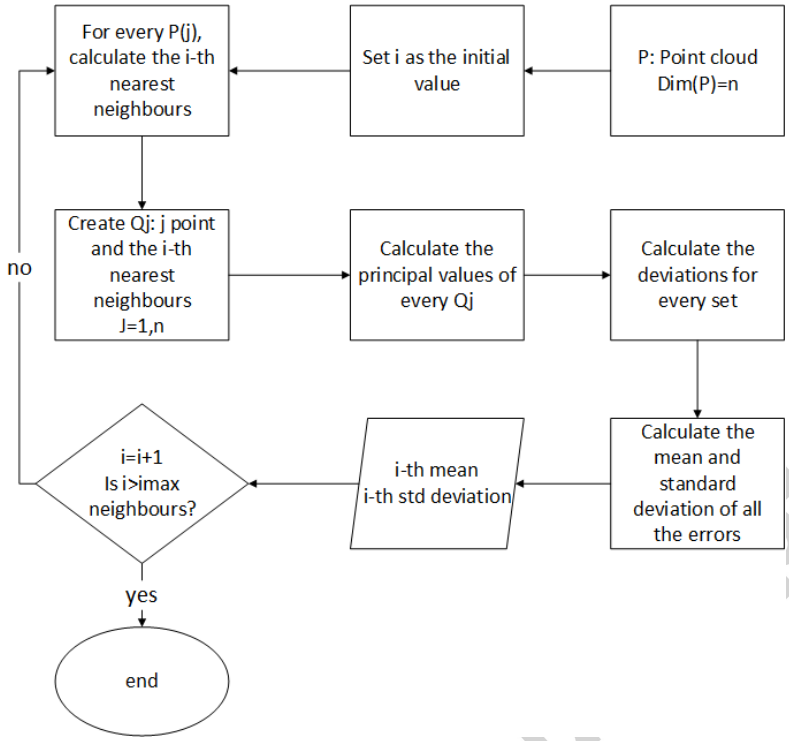


Figure 11. Model calibration scheme.

3.1.1. Plane surfaces analysis

This analysis consists of the method execution for sets extracted from known plane surfaces. The main data characteristics of the sets are:

- All the surfaces have the same area but different density of points.
- Planes 1, 2, 3 and 4 are vertical planes.
- Plane 5 is a horizontal plane with a higher standard deviation (σ) than the vertical planes due to a plane over representation, as stated in data acquisition section.
- Due to the large number of data, the representative value of the deviation (η) is the mean $E(\eta_{ij})$ plus three sigmas, $\sigma(\eta_{ij})$ (Figure 12a).

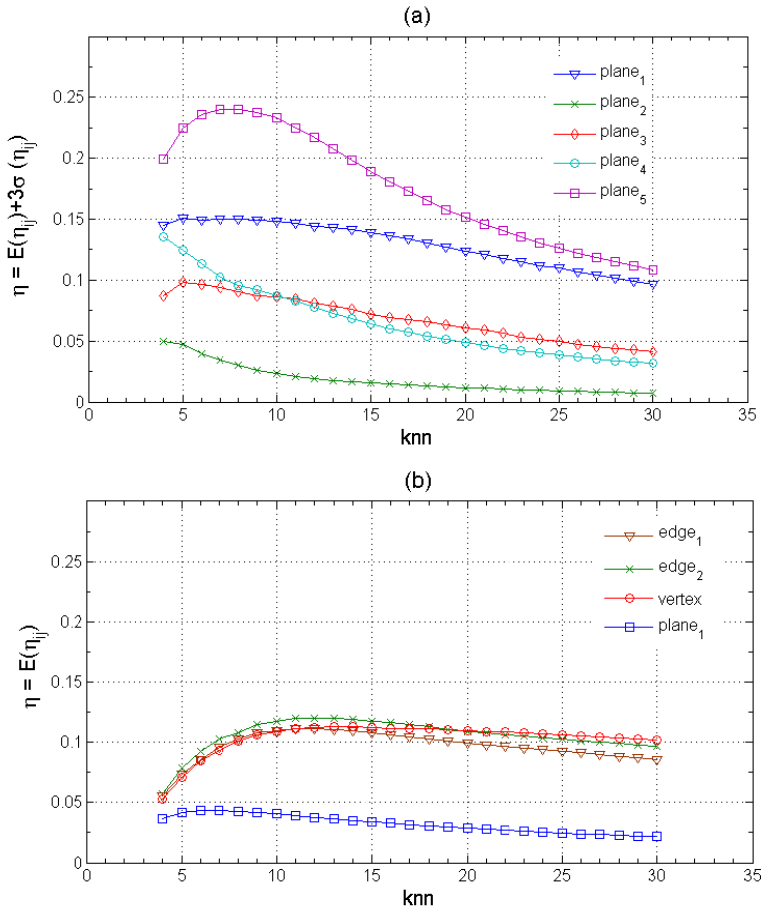


Figure 12. (a) calibration of the parameters η_{max} and *knn* for plane surfaces, (b) calibration of the parameters η and *knn* for no planar sets

This calibration study shows that generally:

- With low values of *knn* the values of η are usually high. This correlation tends to decrease when *knn* grows.
- The data bias becomes less important as the *knn* grows.

3.1.2. Non-coplanar points analysis: edges and vertexes

In order to identify the deviation (η) for the non-coplanar sets —i.e. the edges and vertex—, three sets containing two edges and one vertex have been selected and analysed (Figure 12b). The results of this calibration indicate that the mean error is ~~lower~~ higher on vertexes and edges than on planes~~, and that the convergence error reduction is also much lower~~.

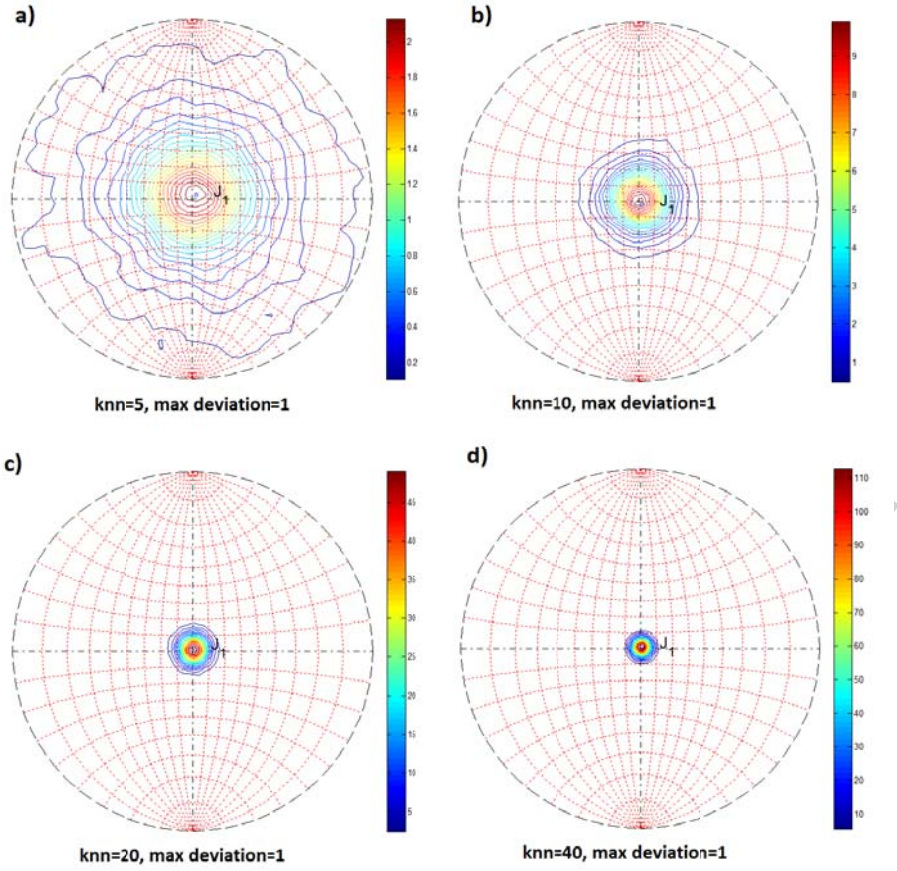
3.2. Number of neighbours and deviation tolerance determination

For a straightforward planar feature detection, non-planar features (such as edges and vertex) should be labelled and discarded according to a combination of two different parameters: η_{max} and *knn* . The calibration test (Case study A) pointed out about the convenience of choosing a high value of *knn* and a low value of η_{max} in order to optimize this segmentation. Thus, a sensitivity analysis on the combination of both parameters was carried out. ~~The calibration pointed that for plane 5, the values knn=30 η_{max}=10% covered the 99,9% of the data. Otherwise, the mean from all edges would cover approximately50% of the data, thus 50% of the non-coplanar points would be discarded. Therefore, this perspective indicates that it is convenient to choose a high value of knn and a low value of η_{max}.~~

The Micro LiDAR case study shows that low numbers of knn (e.g. k<15) retained significant noise in pole calculation. By contrast, a larger number of neighbours (e.g. *knn*>30) significantly smoothed local curvature. Regarding η_{max} parameter, values below 15% produced the discard of good candidates to coplanar points, whereas values above 25% generated the admittance of edge points and coplanar points. Accordingly, *knn* values ranging from 15 to 30 were selected as an optimal compromise between accuracy and resolution. Similarly, the optimal η_{max} value was defined in around 20%.

3.3. Influence of *knn* in the dispersion of the pole planes

The number of *knn* neighbours significantly affects to the pole dispersion in the stereoplot, as can be ~~noticed~~ seen in Figure 13. Considering *knn*=(5, 10, 20, 30) and η_{max}=1, all the poles of the top side of the cube are computed and the density function is plotted. This surface was intentionally chosen because it is horizontal, ~~As this surface was intentionally defined as horizontal,~~ so the ~~normal vector~~ principal pole must be in the centre of the stereoplot —Dip=0° for all the dip direction values—.



534

537 **Figure 13.Influence of the number of neighbours in the scattering of the poles for a single plane (cube datasets,**
538 **horizontal plane). Poles density of the plane 5, i.e. the upper side of the cube,was calculated with different**
539 **values of knn: (a) Knn=5; (b) knn=, 10; (c) knn=20; (d) knn= 30. Isolines are plotted each 5%**

538

541 The result of this calibration (Figure 13) indicates that, as the *knn* parameter grows, the
542 calculated normal vectors tend to be less scattered around the mean value, meaning a
543 greater precisions obtained when increasing the number of neighbour points.

544 This effect happens because all the points belong to the same flat surface. When the
545 points belong to a non-planar surface, such as a very irregular curved surface, the
546 principal pole would not have to converge to a point as *knn* grows.

545 ## 3.4. Pole dispersion effects in the cluster analysis

548 Cluster analysis requires constant density of points in order to obtain homogeneous
549 results (Ester *et al.*, 1996). This analysis shows that, in some particular cases, the poles
550 dispersion in the stereoplot can negatively affect the cluster analysis.

Let's consider a particular case to show this effect. If two adjacent surfaces have very close orientations and at least one surface is highly affected by bias, cluster analysis may provide poor results.

Sparse points define planes with orientations not close enough to the principal plane, so it could be closer to the adjacent surface orientation. Thus, sparse poles could be assigned to ~~another discontinuity set~~ other principal pole and the cluster analysis will extend the cluster to other surface areas or will consider them as noise.

If the icosahedron is considered, adjacent sides form an angle of 42º approximately. If the surface points' bias is high, the discontinuity set assignment could be wrong. The analysis pointed that that if *knn*=15 some points that belong to a surface were assigned to the adjacent one by the cluster analysis. When *knn* was set to 30, the planes concentrated around the principal pole with less deviation. Therefore, the poles were assigned to points correctly and the cluster analysis offered a good result.

3.5. Proposal of the optimal parameters

The previous performed analyses from well-known regular figures using 3D digitizer data allowed us to conclude that the optimal processing parameters for the different processing stages are: (a) For the step A —planes detection—, we set *knn*=30 and η_{max}=20% as optimal parameters; (b) For the Part B —Statistical analysis and poles assignment to discontinuity sets—, we determined γ_1=20º n_p =20; — and γ_2=30º; Finally, (c) for the part C—Cluster analysis— we defined the optimal value of *ppc* =50.

Figure 14 shows the case ~~of~~ study of an icosahedron processed using the above listed parameters. As ~~it~~ can be seen, the analysis ~~has successfully allowed to obtain~~ successfully obtained the different clusters of this Platonic solid.

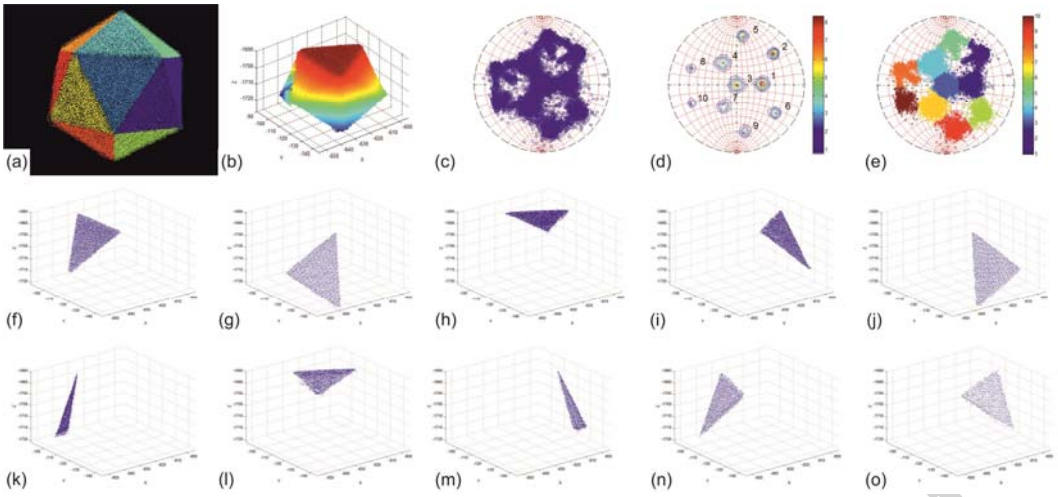


Figure 14.Icosahedron scan results. a) 3D points and discontinuity sets coloured plot, b) 3D data plot, c) calculated poles stereoplot, d) density function plot and discontinuity sets identification, e) poles assigned to discontinuity sets plot, f) – o) calculated clusters. Note that only ten clusters (faces) have been recognized because the additional ten clusters (faces) were in a shaded area of the scan and as a consequence no data were obtained from them.

With the proposed parameters, all the geometrical solid objects were successfully processed. Reasonably good results were obtained: (a) the orientation of each face matches with the compass lectures; furthermore, (b) the visual analysis inspection showed that outliers and non-planar parts of the objects (edges and vertex) were properly segmented and discarded. In the following section, these parameters will be used in a real case (Case study B).

4. Results for case study B: Application to a real roadcut ~~(case study B)~~

Once the methodology has been applied and calibrated using regular geometrical figures, the next step consists in the application of the methodology to a real case of study. The datasets —which are publically available at Rockbench.org_(Lato *et al.*, 2013)—consist of a 3D point cloud on a quartzitic roadcut in Ouray (Colorado).

Since the point cloud was acquired from a single station, no alignment artefacts were detected. Thus the *knn* value was set in 15 and η_{max} to 20% based on the above discussion of calibration.

During the analysis, the method detected five principal discontinuity sets (Figure 15b and Table 2). This figure shows that despite ~~the fact that~~ J_1 is the ~~vastest~~ most visible outcrop, there are other discontinuity sets less represented (i.e. J_2, J_3, J_4 and J_5). As the

method ~~searches~~ searched local maximum values, the J_2, J_3, J_4 and J_5 sets ~~where~~ were identified. ~~In addition, a fifth set was considered, but due to its small density value and the stereoplot closeness to J_1, it was discarded.~~ A visual analysis of the results pointed out that the J_1 orientation (249,04/36,66; Figure 15b) seems to successfully represent the stratification plane.

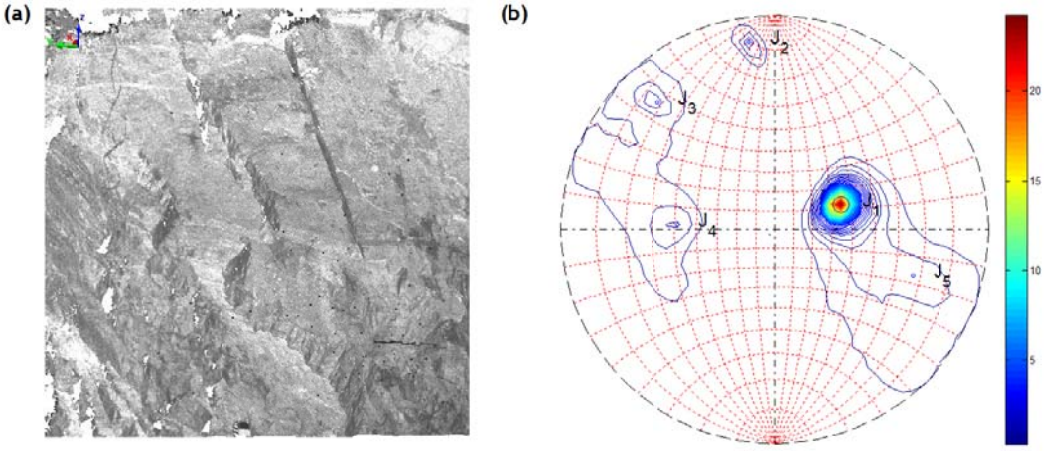
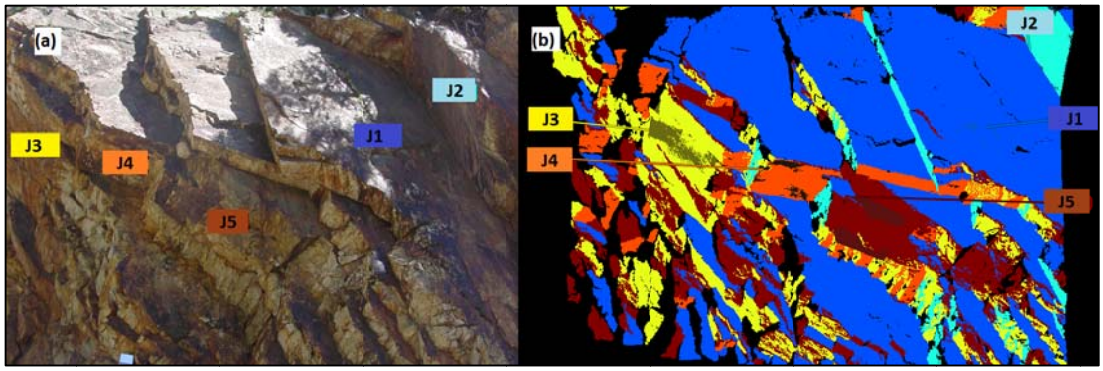


Figure 15.(a) Scattered point cloud . (b) Normal vector density plot of the different planes. Five principal poles were found: J_1 (249/37), J_2 (172/83), J_3 (137/78), J_4 (093/49) and J_5 (288/68). Isolines are plotted each 1,25%.

In order to obtain good cluster visualization the maximum number of points per cluster (ppc) was set to 500. Each cluster has associated an equation (eq. (5)). Table 2 provides the details of the planes adjustment analysis.

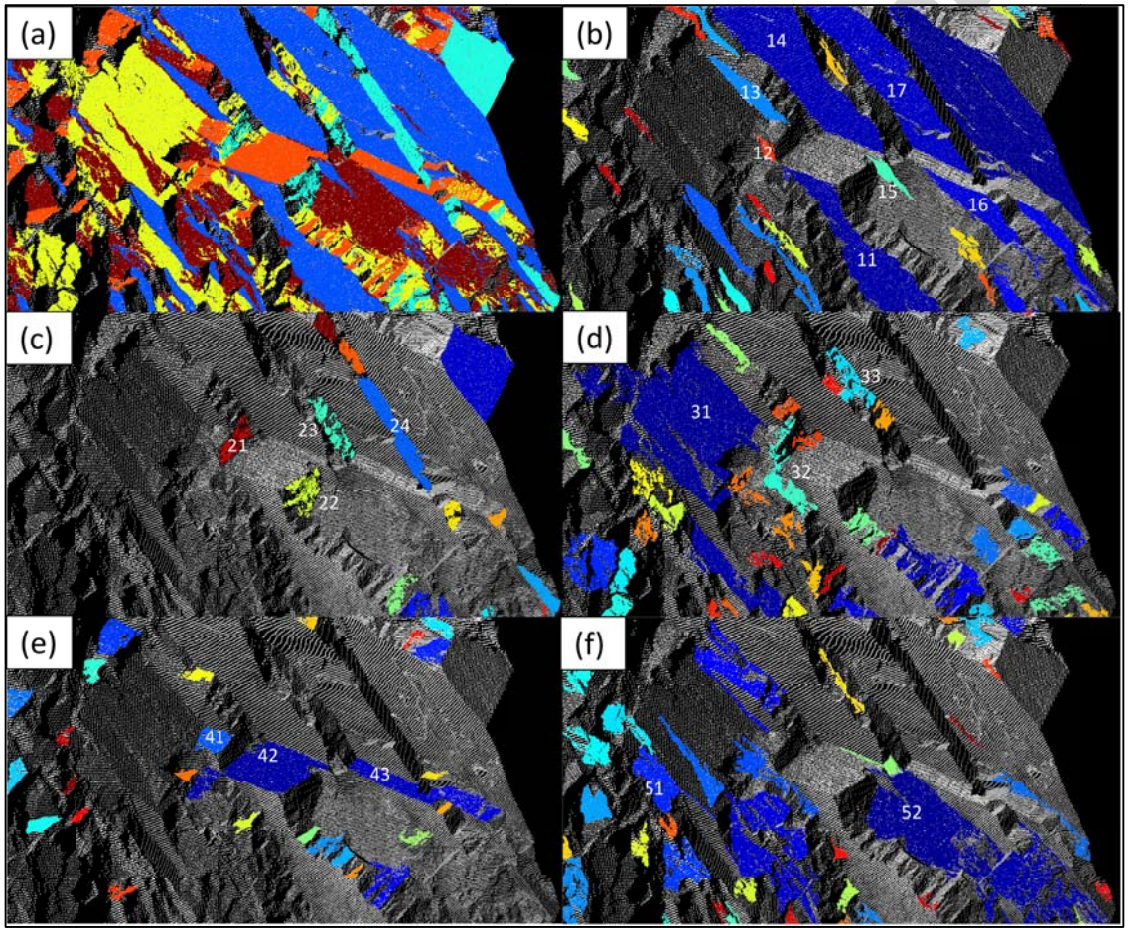
Table 2: Results: application of the proposed methodology to the studied roadcut (Case study B). Dip orientation and dip are in degrees. Error is calculated by eq. (7. See the orientation and location of the discontinuity sets in Figure 16 and Figure 17 respectively.

Discontinuity set	Azimuth º	Dip º	Number of clusters	Number of points	Mean Error, (m)	Std. Dev. Error (m)
J_1	249,04	36,66	59	558.921	$2{,}61 \times 10^{-4}$	$1{,}43\times 10^{-1}$
J_2	172,29	83,16	14	36.781	$2{,}59\times 10^{-4}$	$1{,}36\times 10^{-1}$
J_3	137,33	77,87	56	135.858	$1{,}10\times 10^{-5}$	$1{,}42\times 10^{-1}$
J_4	092,96	48,74	34	96.348	$1{,}67\times 10^{-4}$	$1{,}08\times 10^{-1}$
J_5	288,45	68,22	57	196.613	$2{,}38\times 10^{-4}$	$2{,}13\times 10^{-1}$



611

614 **Figure 16.a) Picture showing the section of the scanned area, b) segmented 3D point cloud: discontinuity set**
615 **assignment (families J1, J2, J3, J4 and J5) to each point; the points not associated to any family are not**
616 **represented**



615

617 **Figure 17. Clusters identification in a section of the Case Study B. a) One colour per discontinuity set with all**
618 **clusters labelled, b) J1, c) J2, d) J3, e) J4 and f) J5 sets representation using one colour per cluster.**

618

620 As an output data example of cluster computation, a crop of the surface shown in Figure
621 16b has been analysed. This window shows the five discontinuity sets and different

clusters. In Figure 17a, we observe the results of the point cloud segmentation: several labelled clusters, whose planes are mentioned in Table 3, are shown with different colour labelling. ~~some of the clusters of the selected area have been labelled and their respective clusters equations (plane equations) have been obtained~~. Note that sets J_1 (represented in Figure 17b by point labels 11 to 17; Table 3), J_3 (represented in Figure 17d by point labels 31 to 33; Table 3) and J_4 (represented in Figure 17e by point labels 41 to 43; Table 3) are almost perpendicular. This can also be observed in Figure 16b.

So, the proposed methodology defines the scanned surface by algebraic expressions following eq.(3).

Table 3.Cluster equations identified from the rock mass crop shown in Figure 15.

Point label (Figure 17)	Discontinuity set id	Cluster id	Parameters of the cluster equation: Ax+By+Cz+D=0 (eq. (5))			
			A	B	C	D
11	J_1	2	-0.576	-0.254	0.777	16.247
12		49	-0.770	-0.180	0.612	20.752
13		15	0.551	0.198	-0.811	-14.102
14		3	-0.554	-0.173	0.814	13.056
15		26	-0.550	-0.203	0.810	13.369
16		5	0.553	0.196	-0.810	-12.713
17		4	-0.522	-0.139	0.841	11.055
21	J_2	13	-0.348	0.930	0.118	2.999
22		8	0.230	-0.945	-0.232	-2.657
23		6	0.339	-0.941	0.003	-6.033
24		3	0.109	-0.968	0.227	-3.548
31	J_3	1	0.681	-0.720	0.129	-12.309
32		24	0.746	-0.654	0.128	-15.079
33		19	0.589	-0.808	-0.005	-12.748
41	J_4	7	0.885	-0.117	0.451	-22.315
42		1	0.768	-0.014	0.640	-19.203
43		2	0.738	-0.086	0.670	-17.861
51	J_5	9	0.810	-0.535	-0.239	-18.112
52		1	0.904	-0.255	-0.343	-20.759

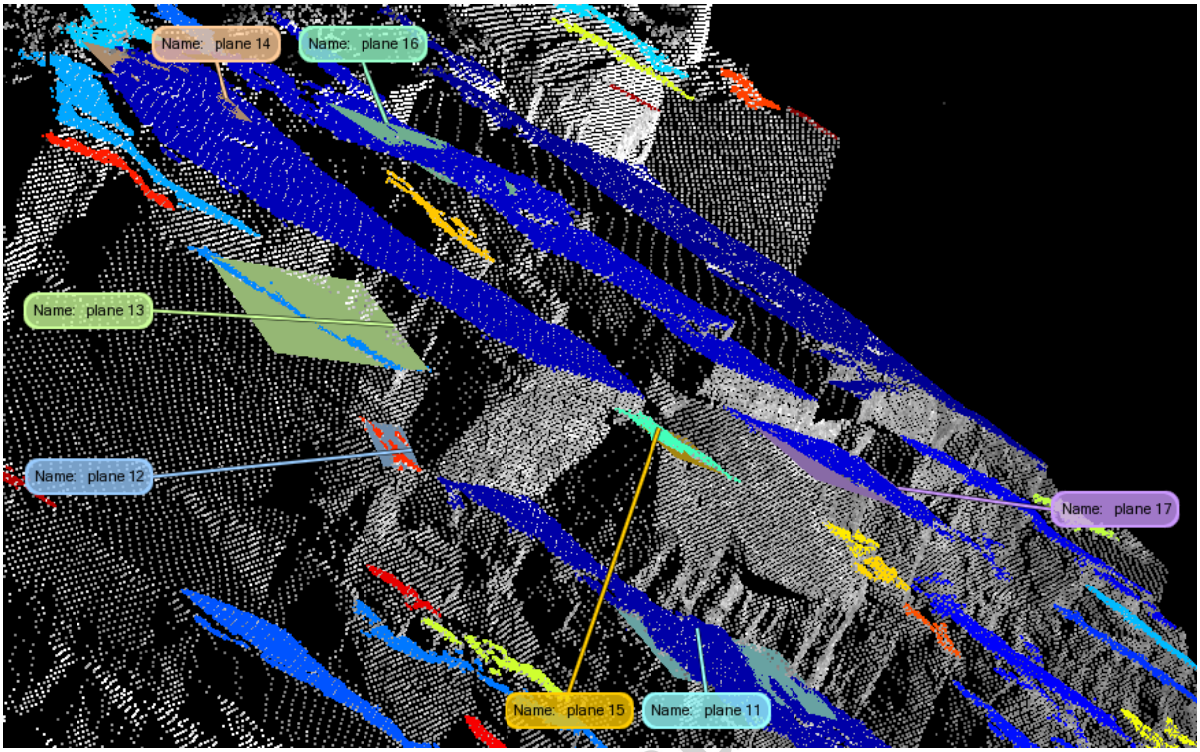


Figure 18. Best-Fit planes extraction of some selected set of points, using Polyworks

Discontinuity orientations measured on field were not available for this case study. Therefore, the results were validated by comparing our method with the classical approach for normal vector estimation, meaning the best fit plane to a subset of points was calculated according to Fernández (2005). Two different indicators were used (Table 2): (a) The mean error for the different discontinuity sets and their associated standard deviations were analysed and included in Table 2; (b) Furthermore, we computed several best-fit planes using Polyworks and compared them with the result (α angle). The extracted vectors (EV) and the method vectors (MV) are almost parallel (Table 4). Nevertheless, the discontinuity J_3 shows the highest angle difference, which was expected due to its surface irregularity. ~~In other words, J_3 and J_4 angle depends on the size and the location of the plane extraction~~.

We finally performed a visual checking of the recognized discontinuities, where each point was coloured according to its discontinuity set or its cluster. Despite the intrinsic waviness of the discontinuities, results agree with the observed field relationships ~~true~~ (Figure 16 and Figure 17).

Table 4. Validation of the proposed method through comparison of our method with classical best-fit plane using Polyworks. The second column shows the cluster orientation corresponding to the Table 3 label

Discontinuity set	Plane orientation using classical approach (Figure 18)	Plane orientation using the proposed method (Figure 17 and Table 3)	Angle between planes (º)
J_1	249.18/40.23 (Plane 11)	246.24/39.02 (Label 11)	2
J_1	264.23/57.02 (Plane 12)	256.86/52.30 (Label 12)	8
J_1	263.97/41.91 (Plane 13)	070.26/35.80 (Label 13)	11
J_1	252.58/36.53 (Plane 14)	252.68/35.48 (Label 14)	1
J_1	248.71/36.98 (Plane 15)	249.74/35.91 (Label 15)	1
J_1	254.77/29.86 (Plane 16)	070.47/35.92 (Label 16)	6
J_1	249.85/35.94 (Plane 17)	255.12/32.72 (Label 17)	4
J_2	338.68/82.35 (Plane 21)	339.47/83.25 (Label 21)	1
J_2	347.47/79.01 (Plane 22)	166.33/76.58 (Label 22)	3
J_2	341.04/89.50 (Plane 23)	160.20/89.86 (Label 23)	1
J_2	353.50/76.40 (Plane 24)	173.55/76.85 (Label 24)	0
J_3	314.10/77.18 (Plane 31)	136.59/82.58 (Label 31)	6
J_3	302.36/75.92 (Plane 32)	131.25/82.67 (Label 32)	11
J_3	330.19/83.01 (Plane 33)	143.91/89.70 (Label 33)	10
J_4	286.12/58.91 (Plane 41)	097.55/63.22 (Label 41)	9
J_4	274.18/51.09 (Plane 42)	091.07/50.19 (Label 42)	3
J_4	277.22/46.42 (Plane 43)	096.64/47.97 (Label 43)	2
J_5	305.04/77.62 (Plane 51)	123.42/76.15 (Label 51)	2
J_5	290.16/66.99 (Plane 52)	105.75/69.94 (Label 52)	5

5. Conclusions

In this work a new method for the semi-automatic calculation of the orientations and position of rock mass discontinuities from 3D LiDAR data is presented. The method is based on the: (a) ~~the~~ calculation of the normal vector using PCA; (b) ~~the~~ removal of anomalous points through the creation of a coplanarity test; (c) ~~the~~ semi-automatic identification of the main discontinuity sets using a KDE analysis; (d) ~~the~~ assignment of each point to a given main family set (or to a noise); and (e) ~~the~~ automatic extraction of single discontinuities using DBSCAN algorithm.

A complete sensitivity analysis of the parameters has been carried out as well, playing a key role on the method, and showing the strong influence that the number of neighbours has in the quality of the method, both for planar features, edges and vertex.

664 The method has been tested using three sources of information —synthetic data, 3D
665 digitized and Terrestrial LiDAR scans— showing a good adaptability of the method to
666 the different sources of information. The case study A allowed us to validate the method
667 and to provide a range of values for the method's parameter, which were then
668 successfully applied in case study B. Furthermore, the method has been tested with
669 more than two million points in a Intel Core i3-350M, 8GB DDR3 RAM with a total
670 processing time of 5307 seconds (Table 5). The slowest step is the coplanarity test,
671 which increases the execution time geometrically as the *knn* parameter grows. In
672 addition, it was necessary to adapt the DBSCAN algorithm to large point clouds.
673 Finally, our experience indicates that it is not recommended to analyse a huge number
674 of points in the same test since principal orientations could be masked due to an excess
675 of poles in stereoplot.

Step	CPU time (s)	%
Part A: local curvature calculation	4277	80.59%
Part B: statistical analysis	2	0.04%
Part C: cluster analysis	1028	19.37%
Total:	5307	100.00%

676 **Table 5. CPU time**

677 One of the strengths of the method consists in using the original information contained
678 in the 3D points during all the process, instead of commonly used approaches that
679 utilise 2.5D interpolated surface model. ~~Thus, our method is able to analyse~~
680 ~~multivaluated surfaced (e.g. over hanged or bended areas)~~. Although a great
681 improvement in workflow automation is obtained using the proposed methodology, a
682 solid background in structural geology and rock mechanics together with the use of
683 useful material such as field pictures and visual recognition of the results is required for
684 an optimum application of the proposed method.

685 Further research lines point to a continuous software development in order to
686 automatically obtain geomechanical parameters (e.g. spacing, persistence, etc.) from the
687 scanned rock masses; furthermore, we support the development of a more reproducible
688 research thanks to the new trend in code and data sharing under Creative-Commons
689 license (e.g. www.reproducibleresearch.net). In order to contribute to the latter, the

690 complete 3D RAW and processed datasets are ~~will be~~ publically available in our
691 website (www.3D-landslide.com/discontinuity).

692

693 ## ACKNOWLEDGEMENTS

694 The laser scanner raw data from the case study B was obtained from Rockbench open
695 repository (www.rockbench.org), the work carried out by the RockBench founders (R.
696 Harrap, M. Lato, J. Kemeny and G. Bevan) is kindly appreciated. This work was
697 partially funded by the University of Alicante (vigrob-157, uausti11–11, and gre09–40
698 projects), the Swiss National Science Foundation (FNS-138015 and FNS-144040
699 projects) and by the Generalitat Valenciana (project GV/2011/044).

700

REFERENCES

Abellán, A., Oppikofer, T., Jaboyedoff, M., Rosser, N.J., Lim, M., Lato, M.J., 2014. Terrestrial laser scanning of rock slope instabilities. Earth Surface Processes and Landforms 39, 80-97.

Abellán, A., Vilaplana, J.M., Martínez, J., 2006. Application of a long-range Terrestrial Laser Scanner to a detailed rockfall study at Vall de Núria (Eastern Pyrenees, Spain). Engineering Geology 88, 136-148.

Bieniawski, Z.T., 1989. Engineering rock mass classifications.

Botev, Z., Grotowski, J., Kroese, D., 2010. Kernel density estimation via diffusion. The Annals of Statistics 38, 2916-2957.

Brodu, N., Lague, D., 2012. 3D terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: Applications in geomorphology. ISPRS Journal of Photogrammetry and Remote Sensing 68, 121-134.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise, KDD, pp. 226-231.

Feng, Q., Sjögren, P., Stephansson, O., Jing, L., 2001. Measuring fracture orientation at exposed rock faces by using a non-reflector total station. Engineering Geology 59, 133-146.

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

Ferrero, A.M., Forlani, G., Roncella, R., Voyat, H.I., 2009. Advanced Geostructural Survey Methods Applied to Rock Mass Characterization. Rock mechanics and rock engineering 42, 631-665.

Friedman, J.H., Bentley, J.L., Finkel, R.A., 1977. An algorithm for finding best matches in logarithmic expected time. ACM Transactions on Mathematical Software (TOMS) 3, 209-226.

García-Sellés, D., Falivene, O., Arbués, P., Gratacos, O., Tavani, S., Muñoz, J.A., 2011. Supervised identification and reconstruction of near-planar geological surfaces from terrestrial laser scanning. Computers & Geosciences 37, 1584-1594.

Gigli, G., Casagli, N., 2011. Semi-automatic extraction of rock mass structural data from high resolution LIDAR point clouds. International Journal of Rock Mechanics and Mining Sciences 48, 187-198.

Hack, R., Price, D., Rengers, N., 2003. A new approach to rock slope stability–a probability classification (SSPC). Bulletin of Engineering Geology and the Environment 62, 167-184.

Harrison, J.P., Hudson, J.A., 2000. Engineering rock mechanics-an introduction to the principles. Access Online via Elsevier.

Hoek, E., Bray, J., 1981. Rock slope engineering. Taylor & Francis.

Ioannou, Y., 2012. Difference of Normals as a Multi-scale Operator in Unorganized Point Clouds, In: Taati, B., Harrap, R., Greenspan, M. (Eds.), pp. 501-508.

Jaboyedoff, M., Metzger, R., Oppikofer, T., Couture, R., Derron, M., Locat, J., Turmel, D., 2007. New insight techniques to analyze rock-slope relief using DEM and 3D-imaging cloud points: COLTOP-3D software, Rock Mechanics: Meeting Society's Challenges and Demands: Proceedings of the 1st Canada-US Rock Mechanics Symposium, Vancouver, Canada, p. 31.

Jaboyedoff, M., Oppikofer, T., Abellán, A., Derron, M.-H., Loye, A., Metzger, R., Pedrazzini, A., 2012. Use of LIDAR in landslide investigations: a review. Natural Hazards 61, 5-28.

Kemeny, J., Norton, B., Turner, K., 2006a. Rock slope stability analysis utilizing ground-based LIDAR and digital image processing. Felsbau 24, 8+10-15.

Kemeny, J., Turner, K., Norton, B., 2006b. LIDAR for rock mass characterization: hardware, software, accuracy and best-practices. Laser and photogrammetric methods for rock face characterization. ARMA Golden, Colorado.

Khoshelham, K., Altundag, D., Ngan-Tillard, D., Menenti, M., 2011. Influence of range measurement noise on roughness characterization of rock surfaces using terrestrial laser scanning. International Journal of Rock Mechanics and Mining Sciences 48, 1215-1223.

Lato, M., Diederichs, M.S., Hutchinson, D.J., Harrap, R., 2009. Optimization of LiDAR scanning and processing for automated structural evaluation of discontinuities in rockmasses. International Journal of Rock Mechanics and Mining Sciences 46, 194-199.

Lato, M., Kemeny, J., Harrap, R.M., Bevan, G., 2013. Rock bench: Establishing a common repository and standards for assessing rockmass characteristics using LiDAR and photogrammetry. Computers & Geosciences 50, 106-114.

Lato, M.J., Diederichs, M.S., Hutchinson, D.J., 2010. Bias correction for view-limited Lidar scanning of rock outcrops for structural characterization. Rock mechanics and rock engineering 43, 615-628.

Lato, M.J., Vöge, M., 2012. Automated mapping of rock discontinuities in 3D lidar and photogrammetry models. International Journal of Rock Mechanics and Mining Sciences 54, 150-158.

Lisle, R.J., 2000. The techniques of modern structural geology: Applications of continuum mechanics in structural geology. Elsevier.

Olariu, M.I., Ferguson, J.F., Aiken, C.L., Xu, X., 2008. Outcrop fracture characterization using terrestrial laser scanners: Deep-water Jackfork sandstone at Big Rock Quarry, Arkansas. Geosphere 4, 247-259.

Oppikofer, T., Jaboyedoff, M., Blikra, L., Derron, M.-H., Metzger, R., 2009. Characterization and monitoring of the Åknes rockslide using terrestrial laser scanning. Natural Hazards and Earth System Science 9, 1003-1019.

Pantelidis, L., 2009. Rock slope stability assessment through rock mass classification systems. International Journal of Rock Mechanics and Mining Sciences 46, 315-325.

Rencher, A.C., Christensen, W.F., 2012. Methods of multivariate analysis. John Wiley & Sons.

Rengers, N., 1967. Terrestrial photogrammetry: a valuable tool for engineering geological purposes. Rock Mechanics and Engineering Geology. Vol. V/2-3.

Romana, M., 1985. New adjustment ratings for application of Bieniawski classification to slopes. Proc. Int. Symp. on the Role of Rock Mechanics, 49-53.

Roncella, R., Forlani, G., 2005. Extraction of planar patches from point clouds to retrieve dip and dip direction of rock discontinuities. Proceedings of laser scanning, 162-167.

Rosser, N.J., Petley, D.N., Lim, M., Dunning, S., Allison, R.J., 2005. Terrestrial laser scanning for monitoring the process of hard rock coastal cliff erosion. Quarterly Journal of Engineering Geology and Hydrogeology 38, 363-375.

Silverman, B.W., 1986. Density estimation for statistics and data analysis. CRC press.

Slob, S., Hack, H., Feng, Q., Röshoff, K., Turner, A., 2007. Fracture mapping using 3D laser scanning techniques, Proceedings of the 11th Congress of the International Society for Rock Mechanics, Lisbon, Portugal, pp. 299-302.

Slob, S., Hack, R., 2004. 3D terrestrial laser scanning as a new field measurement and monitoring technique. Lecture Notes in Earth Sciences, 179-189.

Slob, S., van Knapen, B., Hack, R., Turner, K., Kemeny, J., 2005. Method for Automated Discontinuity Analysis of Rock Slopes with Three-Dimensional Laser Scanning. Transportation Research Record: Journal of the Transportation Research Board 1913, 187-194.

Sturzenegger, M., Stead, D., 2009a. Close-range terrestrial digital photogrammetry and terrestrial laser scanning for discontinuity characterization on rock cuts. Engineering Geology 106, 163-182.

Sturzenegger, M., Stead, D., 2009b. Quantifying discontinuity orientation and persistence on high mountain rock slopes and large landslides using terrestrial remote sensing techniques. Nat. Hazards Earth Syst. Sci. 9, 267-287.

Sturzenegger, M., Stead, D., Elmo, D., 2011. Terrestrial remote sensing-based estimation of mean trace length, trace intensity and block size/shape. Engineering Geology 119, 96-111.

Tomás, R., Delgado, J., Serón, J.B., 2007. Modification of slope mass rating (SMR) by continuous functions. International Journal of Rock Mechanics and Mining Sciences 44, 1062-1069.

Tonini, M., Abellan, A., 2014. Rockfall detection from terrestrial LiDAR point clouds: a clustering approach using R. Journal of Spatial Information Science, Accepted, typesetting (OGRS 2012 special feature).

Turner, A.K., Kemeny, J., Slob, S., Hack, R., 2006. Evaluation, and management of unstable rock slopes by 3-D laser scanning. International association for engineering geology and the environment. The Geological Society of London, 1-11.

Viero, A., Teza, G., Massironi, M., Jaboyedoff, M., Galgaro, A., 2010. Laser scanning-based recognition of rotational movements on a deep seated gravitational instability: The Cinque Torri case (North-Eastern Italian Alps). Geomorphology 122, 191-204.

Vöge, M., Lato, M.J., Diederichs, M.S., 2013. Automated rockmass discontinuity mapping from 3-dimensional surface data. Engineering Geology 164, 155-162.

809

Table 1. Properties of the Case study B datasets

Physical Setting	Roadcut
Location (close mjr. City)	Ouray, Colorado
Lithology	Quartzite
Scanner	Optech
Laser Type	Time of flight
Year scanned	2004
# of scan locations	4
Point spacing	< 2 cm
Number of points	1,515,722
Collected by	John Kemeny
Dataset	10a

810

811

812 **Table 2: Results: application of the proposed methodology to the studied roadcut (Case study B). Dip**
813 **orientation and dip are in degrees. Error is calculated by eq. (9). See the orientation and location of the**
814 **discontinuity sets in Figure 16 and Figure 17 respectively.**

Discontinuity set	Azimuth º	Dip º	Number of clusters	Number of points	Mean Error (m)	Std. Dev. Error (m)
J_1	249,04	36,66	59	558.921	$2,61 \times 10^{-4}$	$1,43\times 10^{-1}$
J_2	172,29	83,16	14	36.781	$2,59\times 10^{-4}$	$1,36\times 10^{-1}$
J_3	137,33	77,87	56	135.858	$1,10\times 10^{-5}$	$1,42\times 10^{-1}$
J_4	092,96	48,74	34	96.348	$1,67\times 10^{-4}$	$1,08\times 10^{-1}$
J_5	288,45	68,22	57	196.613	$2,38\times 10^{-4}$	$2,13\times 10^{-1}$

815

816

817 **Table 3.Cluster equations identified from the rock mass crop shown in Figure 15.**

Point label (Figure 17)	Discontinuity set id	Cluster id	Parameters of the cluster equation: Ax+By+Cz+D=0 (eq. (5))			
			A	B	C	D
11	J_1	2	-0.576	-0.254	0.777	16.247
12		49	-0.770	-0.180	0.612	20.752
13		15	0.551	0.198	-0.811	-14.102
14		3	-0.554	-0.173	0.814	13.056
15		26	-0.550	-0.203	0.810	13.369
16		5	0.553	0.196	-0.810	-12.713
17		4	-0.522	-0.139	0.841	11.055
21	J_2	13	-0.348	0.930	0.118	2.999
22		8	0.230	-0.945	-0.232	-2.657
23		6	0.339	-0.941	0.003	-6.033
24		3	0.109	-0.968	0.227	-3.548
31	J_3	1	0.681	-0.720	0.129	-12.309
32		24	0.746	-0.654	0.128	-15.079
33		19	0.589	-0.808	-0.005	-12.748
41	J_4	7	0.885	-0.117	0.451	-22.315
42		1	0.768	-0.014	0.640	-19.203
43		2	0.738	-0.086	0.670	-17.861
51	J_5	9	0.810	-0.535	-0.239	-18.112
52		1	0.904	-0.255	-0.343	-20.759

818

819

820 **Table 4. Validation of the proposed method through comparison of our method with classical best-fit**
821 **plane using Polyworks. The second column shows the cluster orientation corresponding to the Table 3**
822 **label**

Discontinuity set	Plane orientation using classical approach (Figure 18)	Plane orientation using the proposed method (Figure 17 and Table 3)	Angle between planes (º)
J_1	249.18/40.23 (Plane 11)	246.24/39.02 (Label 11)	2
J_1	264.23/57.02 (Plane 12)	256.86/52.30 (Label 12)	8
J_1	263.97/41.91 (Plane 13)	070.26/35.80 (Label 13)	11
J_1	252.58/36.53 (Plane 14)	252.68/35.48 (Label 14)	1
J_1	248.71/36.98 (Plane 15)	249.74/35.91 (Label 15)	1
J_1	254.77/29.86 (Plane 16)	070.47/35.92 (Label 16)	6
J_1	249.85/35.94 (Plane 17)	255.12/32.72 (Label 17)	4
J_2	338.68/82.35 (Plane 21)	339.47/83.25 (Label 21)	1
J_2	347.47/79.01 (Plane 22)	166.33/76.58 (Label 22)	3
J_2	341.04/89.50 (Plane 23)	160.20/89.86 (Label 23)	1
J_2	353.50/76.40 (Plane 24)	173.55/76.85 (Label 24)	0
J_3	314.10/77.18 (Plane 31)	136.59/82.58 (Label 31)	6
J_3	302.36/75.92 (Plane 32)	131.25/82.67 (Label 32)	11
J_3	330.19/83.01 (Plane 33)	143.91/89.70 (Label 33)	10
J_4	286.12/58.91 (Plane 41)	097.55/63.22 (Label 41)	9
J_4	274.18/51.09 (Plane 42)	091.07/50.19 (Label 42)	3
J_4	277.22/46.42 (Plane 43)	096.64/47.97 (Label 43)	2
J_5	305.04/77.62 (Plane 51)	123.42/76.15 (Label 51)	2
J_5	290.16/66.99 (Plane 52)	105.75/69.94 (Label 52)	5

823

824

825

Step	Computing time (s)	%
Part A: local curvature calculation	4277	80,59%
Part B: statistical analysis	2	0,04%
Part C: cluster analysis	1028	19,37%
Total:	5307	100%

826 **Table 5. CPU time**

827

828

829