Journal of Geophysical Research: Solid Earth

RESEARCH ARTICLE
10.1029/2018JB016319

Key Points:
- A nonparametric method to assess precision of best fit planes from 3-D georeferenced data is presented
- Bootstrap statistics were tested in Monte Carlo-simulated point clouds to assess the reliability of best fitting planes
- Oblateness of the point cloud underlying distribution is a straightforward parameter to assess the precision of the best fit plane

Supporting Information:
- Supporting Information S1
- Data Set S1
- Data Set S2
- Data Set S3
- Data Set S4
- Data Set S5

Correspondence to:
L. C. Gallo,
len.gallo@gmail.com;
lgallo@gl.fcen.uba.ar

Citation:
Gallo, L. C., Cristallini, E. O., & Svarc, M. (2018). A nonparametric approach for assessing precision in georeferenced point clouds best fit planes: Toward more reliable thresholds. *Journal of Geophysical Research: Solid Earth*, *123*, 10,297–10,308. https://doi.org/10.1029/2018JB016319

Received 26 DEC 2017
Accepted 31 OCT 2018
Accepted article online 5 NOV 2018
Published online 29 NOV 2018

©2018. American Geophysical Union. All Rights Reserved.

A Nonparametric Approach for Assessing Precision in Georeferenced Point Clouds Best Fit Planes: Toward More Reliable Thresholds

Leandro C. Gallo[1,2], Ernesto O. Cristallini[2,3,4], and Marcela Svarc[2,5]

[1]Instituto de Geociencias Básicas, Aplicadas y Ambientales de Buenos Aires, Departamento de Ciencias Geológicas, Facultad de Ciencias Exactas y Naturales, Ciudad Universitaria, Buenos Aires, Argentina, [2]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina, [3]Laboratorio de Modelado Geológico, Instituto de Estudios Andinos, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina, [4]Laboratorio de Termocronología (La.Te Andes), Consejo Nacional de Investigaciones Científicas y Técnicas, Salta, Argentina, [5]Departamento de Matemática y Ciencias, Universidad de San Andrés, Victoria, Argentina

Abstract The fitting of a plane to data points is essential to the geosciences. However, it is recognized that the reliability of these best fit planes depends upon the point set distribution and geometry, evaluated in terms of the eigen-based parameters derived from the moment of inertia analysis. Despite its significance, few studies have addressed the uncertainties of the analysis, which can adversely affect the reproduction of results one of the cornerstones of scientific endeavor. Aiming to contribute toward the neglected issue of the moment of inertia precision, we have developed a bootstrap resampling scheme to empirically discover the distribution of uncertainties in the orientation of best fit planes. Dispersion of the bootstrapped normal vectors to the best fit plane is regarded as a measure of precision, evaluated with the maximum angular distance from the optimal solution. This rationale was tested using Monte Carlo-generated samples covering a comprehensive range of shape parameters to assess the dependence between eigen parameters and their inherent bias. Our results show that the oblateness of the point cloud is a robust parameter to assess the reliability of the best fit plane. Given this, the method was then applied to a publicly available lidar data set. We argue that georeferenced point clouds with an oblateness parameter greater than 3 and 1.5 may be placed at 95% confidence levels of 5° and 10°, respectively. We propose using these values as thresholds to obtain robust best fit planes, guaranteeing reproducible results for scientific research.

1. Introduction

The reproduction of research is one of the cornerstones of the scientific endeavor. However, a growing lack of reproducibility is causing concern over common statistical methods (e.g., Baker, 2016; Benjamin et al., 2018). Our aim is to understand the population involved in a study and whether or not our results can be reproduced. Alternatives to common statistical methods to assess the precision and variability are therefore needed.

Motivated by a plethora of modern key digital technologies, we are now on the threshold of significant improvements in three-dimensional spatial resolution in general geoscience (McCaffrey et al., 2005). The new techniques include the digital outcrop and terrestrial altimetry acquisition of three-dimensional point clouds with improved precision and accuracy (e.g., terrestrial lidar, terrestrial digital photogrammetry or unmanned aerial vehicle-based digital photogrammetry; see Telling et al., 2017, and references therein).

A major application of these methods includes the characterization of structural heterogeneities, such as faults, joints, or fractures. The characterization of those features makes the description of fracture networks possible. Also, it can be critical for assessing rock mass stability (e.g., Riquelme et al., 2015), fluid flow (e.g., Wilson et al., 2011), or naturally fractured reservoirs (e.g., Biber et al., 2018).

Several manual, semiautomated, and unsupervised procedures have been developed for structural data reduction, allowing three-dimensional representations of structural traces (Lato et al., 2009; Li et al., 2016; Seers & Hodgetts, 2016a; Vasuki et al., 2014), as well as points of near-planar geological surfaces (Chen et al., 2016; García-Sellés et al., 2011; Gomes et al., 2016; Lato et al., 2009; Riquelme et al., 2014; Seers &

Hodgetts, 2014; Slob et al., 2005; Thiele et al., 2017; Wang et al., 2013). These representations of facets permit best fit plane appraisal to be made in order to estimate the corresponding orientation (Dueholm & Olsen, 1993; Pringle et al., 2006).

The approach to obtain an orientation from a set of points consists of estimating the moment of inertia of a set of nodes and using the axis of maximum moment of inertia as the pole to the best fit plane (i.e., the least squares plane: Fernández, 2005; Woodcock, 1977). The eigen decomposition of vertex covariance matrices is widely used, but the precision of the analysis is a neglected issue.

By using the spatial location of earthquakes, this method is also able to reconstruct the structure of the active part of a fault network (e.g., Ouillon et al., 2008; Wang et al., 2013). The extraction of information on bedding planes for the assessment of its influence on landslide abundance (e.g., Santangelo et al., 2015) is evaluated through this analysis, which is also commonly used in archeological studies (McPherron, 2017) and even in biomedical research (Palit et al., 2017).

Going beyond surfaces, the analysis of the moment of inertia has a much wider application in the geosciences. Adopted as a useful way of summarizing three-dimensional orientation data, it is useful for quantifying fabric shape in sedimentary data (Benn, 1994), grain fabric or crystallographic preferred orientation in computed tomography images (e.g., Chatzaras et al., 2016; Ketcham, 2005), and it is even used to classify ice crystal shapes (Gough et al., 2012). Furthermore, this procedure is commonly used to estimate lines and planes of best fit along demagnetization paths from paleomagnetic data (Kirschvink, 1980) and to calculate the intersection of great circles (e.g., Gallo et al., 2017).

The robustness of the best fit plane is highly dependent upon the underlying distribution of the point cloud. The precision decreases as the distribution of the data set moves away from the ideal planar distribution toward a collinear configuration (Fernández, 2005; Seers & Hodgetts, 2016b). Thus, the impact of collinear configurations results in least squares planes being less resolved and prone to noise. The extent to which this may be the case has not been explored and relatively few studies address the uncertainties in the analysis of the moment of inertia (e.g., Seers & Hodgetts, 2016b; Jones et al., 2016).

In this paper, we introduce a novel nonparametric bootstrap procedure for evaluating the uncertainty of the best fit plane estimate, we quantify its robustness using a resampling scheme to empirically discover the distribution of uncertainties for the oriented planes, without any prior information. Our new method enables us to assess the estimation error and construct reliable confidence intervals in a fully data-driven way. When analyzing the data, the estimated uncertainty can be used to rank the measured data and thus develop more accurate models with the understanding of the nature and limitations of the estimates. The bootstrap method can be used to provide accurate estimates of the variability of the desired parameters and avoids the necessity for making strong assumptions about the distribution of the data.

To test the validity of the bootstrap statistics we apply the resampling scheme to stochastically generated (i.e., Monte Carlo simulated) point cloud and surface samples with a wide spectrum of spatial underlying distributions and also to a publicly available lidar data set.

2. Moment of Inertia Analysis for Best Fit Plane Estimates and Its Inherent Bias

Several methods exist for extracting the best fit plane directly from a point cloud, including multiple planar regressions (see equations in Press et al., 1986) and random sample consensus algorithms (Fischler & Bolles, 1981), based on the iterative segmentation of the point cloud to refine equation coefficients by least squares estimation after outlier removal (Chen et al., 2016; Ferrero et al., 2009).

In this study, we consider the eigenvalue decomposition of the covariance matrix to estimate the best fit plane. This approach follows the process used in standard structural analysis to define girdle distributions (Woodcock, 1977). Since the algorithm works directly on the point cloud, no data structuring (triangulation or approximation) is required. The moment of inertia of a set of nodes is estimated from the vectors, linking each point to the point cloud center of mass, $[\overline{x},\overline{y},\overline{z}]$ (Figure 1). Denote $\mathbf{T}$ to the orientation matrix (Scheidegger, 1965):

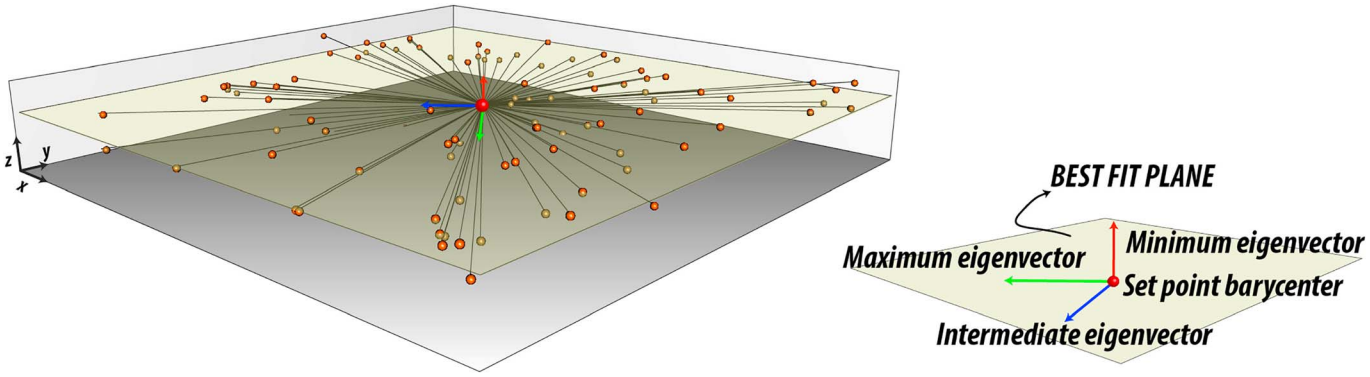

Figure 1. The moment of inertia of a set of nodes is estimated from the vectors, linking each point to the point cloud center of mass (red dot), $[\overline{x},\overline{y},\overline{z}]$. The best fit plane is the hyperplane given by the translation of subspace spanned by the eigenvectors v_1 and v_2 that contains the point cloud center of mass $[\overline{x},\overline{y},\overline{z}]$.

$$\mathbf{T} = \begin{bmatrix} \sum_{i=1}^{n}(x_i-\overline{x})^2 & \sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y}) & \sum_{i=1}^{n}(x_i-\overline{x})(z_i-\overline{z}) \\ \sum_{i=1}^{n}(y_i-\overline{y})(x_i-\overline{x}) & \sum_{i=1}^{n}(y_i-\overline{y})^2 & \sum_{i=1}^{n}(y_i-\overline{y})(z_i-\overline{z}) \\ \sum_{i=1}^{n}(z_i-\overline{z})(x_i-\overline{x}) & \sum_{i=1}^{n}(z_i-\overline{z})(y_i-\overline{y}) & \sum_{i=1}^{n}(z_i-\overline{z})^2 \end{bmatrix}$$

since $\mathbf{T}/(n-1)$ is the point cloud empirical covariance matrix, $\mathbf{T}$ is a positive definite symmetric matrix and is therefore amenable to eigen decomposition. Let λ_1, λ_2, and λ_3 be $\mathbf{T}$'s eigenvalues, which are positive, assume that $\lambda_1 \geq \lambda_2 \geq \lambda_3$, and denote v_1,v_2, and v_3 to the corresponding base of orthogonal eigenvectors. The axis of maximum moment of inertia (v_3) is the pole to the best fit plane (Figure 1), a hyperplane spanned by the eigenvectors v_1 and v_2, that contains the point cloud center of mass $[\overline{x},\overline{y},\overline{z}]$.

Eigenvalues of $\mathbf{T}$ have been widely used as indicators of fabric shape assuming that the mean of the unit vectors (i.e., observations regarded as points of unit mass) coincides with the origin [0, 0, 0] (Woodcock, 1977). Hereafter, we assume that the origin has been shifted to the center of mass, and we use the logarithm of the ratios between the eigenvalues to characterize the point cloud's shape and orientation, as stated by Woodcock (1977). The oblateness (O) is defined as $\ln(\lambda_2/\lambda_3)$, while the prolateness (P) is given by $\ln(\lambda_1/\lambda_2)$. Following the convention stated by Flinn (1962), the prolateness versus the oblateness is plotted in Figure 2. Uniaxial symmetric clusters are plotted along the line $\ln(\lambda_2/\lambda_3) = 0$, while axially symmetric oblate distributions are plotted along the line $\ln(\lambda_1/\lambda_2) = 0$. The graph shows configurations that have both oblate and prolate tendencies, which are quantified by vertex collinearity [$K = \ln(\lambda_1/\lambda_2)/\ln(\lambda_2/\lambda_3)$]. The parameter M [$M = \ln(\lambda_1/\lambda_3)$] is an estimate of the vertex coplanarity and a measure of the strength of the preferred orientation. These two parameters aim to describe the goodness of fit and the reliability of the fitted model. Figure 2 shows six stochastically generated distributions covering a wide range of distribution shapes.

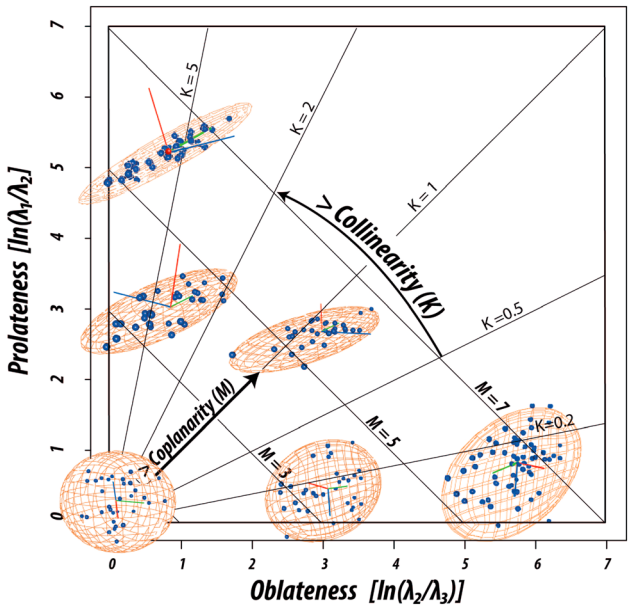

Figure 2. Spatial distribution of point data according to the different ratios between eigenvalues. $M = \ln(\lambda_1/\lambda_3)$ and $K = \ln(\lambda_1/\lambda_2)/(\lambda_2/\lambda_3)$. Data points within the clouds are independent of P and O. Modified from Woodcock (1977).

Using the axis of maximum moment of inertia as the pole to the best fit plane gives an estimate of the desired orientation but gives no information about its inherent precision. Schmidt (1985) first found that for converging great circle methods (i.e., best fit plane of normals to converging great circles) the precision of the intersection relied upon the distribution of the underlying data and argued that the departure from collinearity of the normal vectors to the great circles kept the inherent bias of the method to a minimum. Fernández (2005) indicates that for a set of highly colinear nodes many best fit planes can be defined with a similar degree of

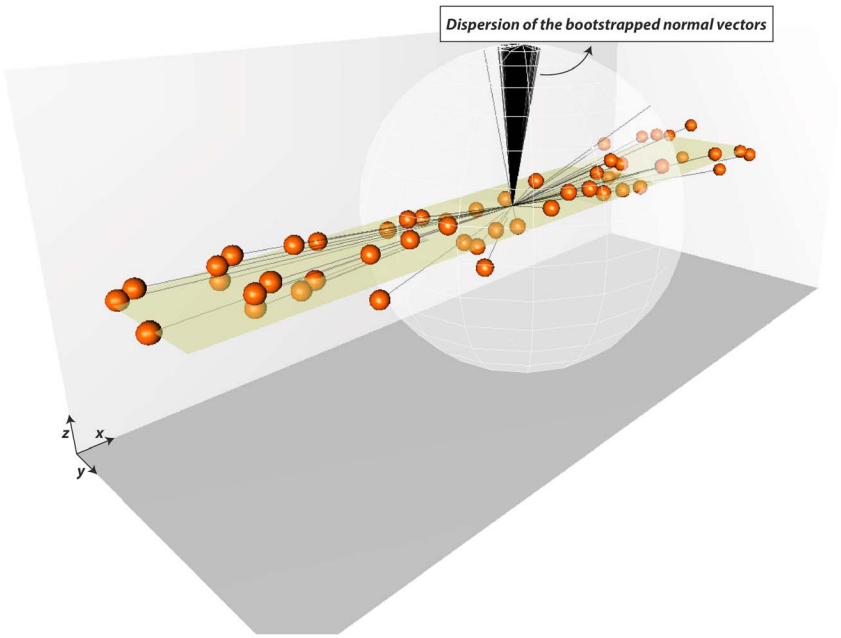

Figure 3. Procedure proposed in this article to evaluate precision of the fit applied to a synthetic data set. Dispersion of the bootstrapped normal vectors to the best fit plane is regarded as a measure of precision, represented by the quantile 95% of angular distance from the optimal solution.

fit and for a perfectly linear set of points, infinite planes with different orientations can be defined. Recent numerical Monte Carlo experiments by Seers and Hodgetts (2016b) confirmed a systematic decay in the estimated fit plane precision commensurate with increasing lineament vertex collinearity. Consequently, the reliability of the best fit plane is contingent upon the distribution of the point cloud.

We apply a bootstrap resampling technique to evaluate the reliability of the fit and find the dispersion of the bootstrapped vectors normal to the best fit plane, which is used as a measure of the precision. We then test the validity of the approach on Monte Carlo-simulated samples of point clouds covering a wide range of distributions.

3. Evaluating the Precision of Solutions: A Bootstrap Approach

We tackle the problem of assessing the goodness of fit of the best fit plane procedure. Recently, Seers and Hodgetts (2016b) faced a similar problem in the context of structural lineament best fit plane: They assessed the unit vector distributions on the circle and the sphere, using both Bingham and Fisher statistics assuming Bingham and von Mises distributions, respectively. The parametric assumptions they have made about directional statistics are thoroughly described in Fisher et al. (1987) and references therein. Adopting parametric assumptions about directional statistics offers considerable advantages for reducing the goodness of fit problem to the estimation of a few parameters. Nonetheless, in many situations the commonly used parametric distribution functions fail to model the data adequately, hence developing alternative, more flexible procedures, is necessary. To overcome this problem, we propose a nonparametric method that has the capacity to deal with the probability distributions that reflect complex shapes, since those are the problems that arise in this field. Our proposal has only a mild assumption, which is the existence of the first two moments. Once the moment of inertia analysis determines the optimal solution from a data set, the precision of the solution can be evaluated following a resampling technique. Taking advantage of computer-intensive statistical methods, it is not necessary to assume a prior underlying distribution, the uncertainties of the estimated surface can be assessed empirically by a bootstrap scheme, which is a flexible and powerful statistical tool that can be used to quantify the uncertainty associated with a given estimator. Such techniques (Efron, 1979) are used in situations where it is not feasible to use an analytical method to provide estimates of variability; repeated calculations explore possible outcomes numerically, which are used in place of complicated or intractable theoretical development (e.g., Constable & Tauxe, 1990).

The method proceeds as follows: let $X = (X_1, \ldots, X_n)$ be the point cloud that contains n observations. A pseudo-sample $X^* = (X_1^*, \ldots, X_n^*)$, of n observations, is obtained by randomly drawing data (with replacement) from X. This procedure is repeated N_b times, where $N_b >> n$. On each replicate the maximum moment of inertia is computed; hence, we obtain $[v_3^{(1)}, \ldots, v_3^{(N_b)}]$ (Figure 3).

3.1. Parameters That Indicate the Dispersion of the Best Fit Plane

Mapping out the variability of the bootstrapped normal vectors to the best fit plane gives us a robust estimate of the precision of the fit. The less scattered the vectors are, the more accurate the fit is. Namely, the dispersion of $[v_3^{(1)}, \ldots, v_3^{(N_b)}]$ on the unit hypersphere provides a measure of precision. The uncertainty of the optimal solution was evaluated through the following procedure:

1. Consider the bootstrapped minimum eigenvectors of X, $[v_3^{(1)}, \ldots, v_3^{(N_b)}]$, centered at $[\overline{x}, \overline{y}, \overline{z}]$ on a unit hypersphere. Compute its covariance matrix $\mathbf{T_B}$.
2. Denote μ to the eigenvector corresponding to the maximum eigenvalue (i.e., principal direction) of $\mathbf{T_B}$; μ gives a robust estimate of the normal vector to the best fit plane. The uncertainty of the optimal pole vector to the best fit plane was denoted by the spread of the unit vectors around μ.
3. Great circle distances of the unit vectors from μ were computed as the dot product between μ and each $v_3^{(i)}$, for $i = 1, \ldots, N_b$.

$$\Theta^{(i)} = \cos^{-1}\left(\mu \cdot v_3^{(i)}\right)$$

4. Sort the N_b vectors, $[v_3^{(1)}, \ldots, v_3^{(N_b)}]$, in ascending order by their Θ values.
5. The maximum great circle distance (Θ_{MAX}) of the bootstrapped vectors from μ at a 95% confidence level (i.e., the quantile 95% of the $\Theta^{(i)}$) was represented excluding 5% of the vectors that were the most distant (Yamaji et al., 2010). That is, let n be the maximum integer satisfying $n \leq 0.95N_b$, then $\Theta_{(n)} = \Theta_{MAX}$.

The dispersion has a maximum at $\Theta_{MAX} = 180^\circ$ when the distribution is uniform over the sphere, meaning that the orientation of the plane is not constrained at all. Precisely determined best fit planes have small Θ_{MAX} values.

4. Testing the Power of the Approach on Monte Carlo-Simulated Samples

Correlating the dispersion measure, Θ_{MAX}, with the point cloud underlying distribution would enable the assessment of the incidence of the shape parameters on the precision of the best fit plane. The purpose is to study the relationship between the ratios of the eigenvalues of **T** that characterize the underlying point cloud shape (oblateness, prolateness, collinearity, and coplanarity) with the dispersion measure, Θ_{MAX}. Given this, aim two Monte Carlo simulation studies were conducted. The procedure involves the generation of synthetic point clouds. For each replicate, we generate a random sample (i.e., point cloud/surface) with its corresponding shape metric. The outcome covers a wide range of distributions of eigen characteristics of the points clouds. For each replicate, the generated point cloud random sample was then assessed by the bootstrap procedure described on section 3.

In order to eliminate a possible dependence between the sample size and the dispersion, for each replicate the sample size n was uniformly randomly chosen between 50 and 10,000 (which are regular samples sizes for these types of problems). The number of bootstrap resamples were chosen at random between 1,000 and 10,000, since no rules of thumb regarding the number of resamples have been established. Taking advantage of parallel programming, the simulation was run in eight parallel processes. Thirty-one thousand nine hundred fifty-nine replicates were generated with its inherent bootstrap approach on a typical quad-core desktop computer in a week. The subroutines written in Visual Basic.NET are included as supporting information).

In this section we describe in this section the construction of three-dimensional point clouds based on two sample generator routines. Our first aim is to thoroughly cover all the possible eigenvalues of **T** since those parameters give a measure of goodness of fit of the best fit plane. We generated uniform random samples in random boxes; even though this approach lacks realism it fulfills our needs. In the second part of the

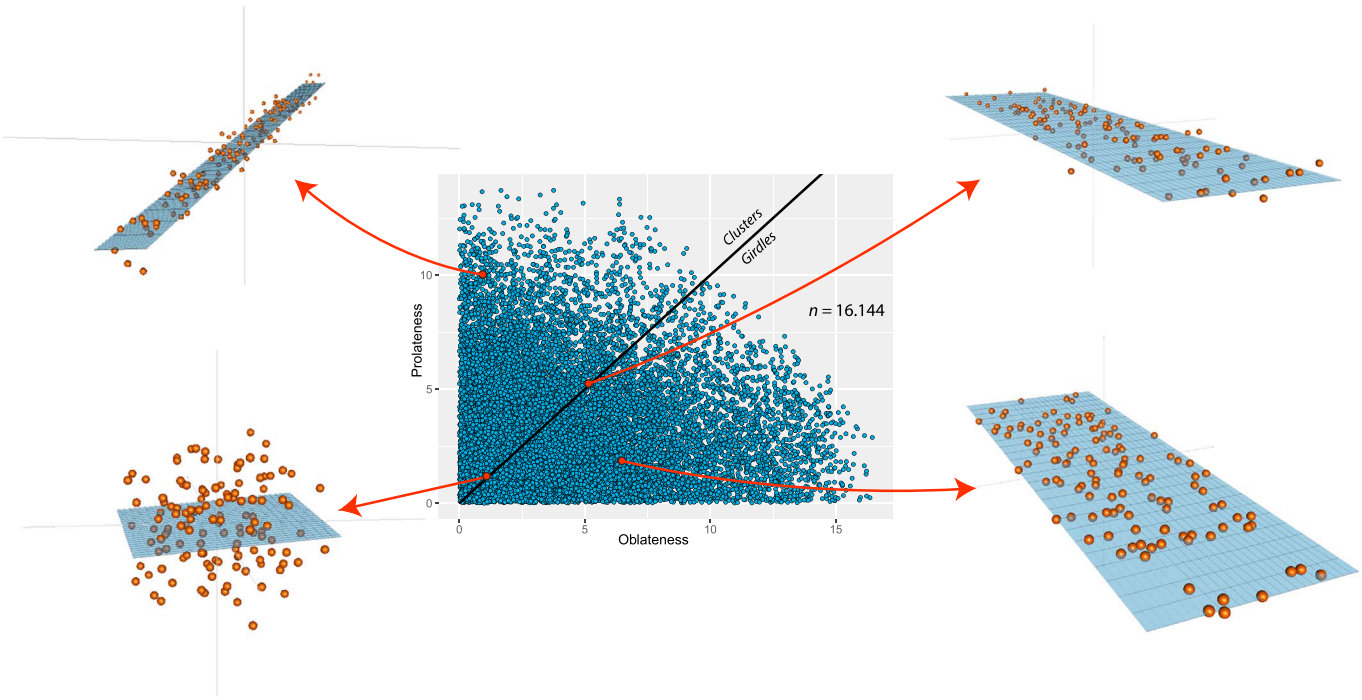

Figure 4. Eigenvalue ratio plots (Woodcock, 1977) showing the orientation tensors of Monte Carlo-simulated samples (see text for details). Data are displayed using ggplot2 (Wickham, 2016) and rgl libraries of R language.

simulation study we focused on the analysis of more realistic surfaces, and for this, we generated fractional Brownian surfaces (fBS), which is an established approach (see Fournier et al., 1982; Seers & Hodgetts, 2016b).

4.1. Uniform Random Distribution in Boxes

For each replicate, n points $X_i = (x_1, x_2, x_3)_i$ were generated, following a coordinate-wise independent uniform distribution in the interval $U(-\tau_j, \tau_j)$, where $\tau_j (0 < \tau_j \leq 1)$ is a scaling parameter that constrains the eigen characteristics of each sample, for $j = 1, 2, 3$. For each sample, components parameters τ_j are independent; over the successive iterations, multiple combinations of scaling factors are considered. Figure 4 shows the distribution of the eigen characteristic of the simulated samples.

4.2. The fBS

To simulate 2-D spatial distributions we apply the random midpoint displacement fractal algorithm (Fournier et al., 1982) also known as the diamond-square algorithm, for generation of fBS, which can be related directly to geological structures (e.g., Méheust & Schmttbuhl, 2001; Seers & Hodgetts, 2016b). The roughness of the topography created by the algorithm is controlled by the Hurst parameter H that varies from 0 to 1, and it determines the fractal dimension of the surface. In this study we cover a wide range of Hurst exponents. To generate the fBS we employed the iterative diamond-square algorithm. With the aim of enriching the study, going through a greater range of eigen parameters, in each iteration we studied the adjustment in subsets of the generated surfaces. Figure 5 shows the distribution of the eigen characteristic of the simulated samples.

4.3. Linking Shape and Uncertainty With the Maximal Information Coefficient

On each replicate of the corresponding simulation study, the orientation matrix **T** is computed from its vertex list with the best fit plane normal (v_3) and the eigen parameters (M, K, O, and P) to characterize the point cloud shape and orientation. Then, as explained in section 3, the bootstrap procedure is conducted in order to calculate the dispersion parameter Θ_{MAX}.

Finally, we proceeded to compare the dispersion value Θ_{MAX} with the eigen parameters obtained from **T**. There are several ways to measure the dependence between variables. The Pearson correlation coefficient (R) is by far the most widely used measure of the dependence between two stochastic variables.

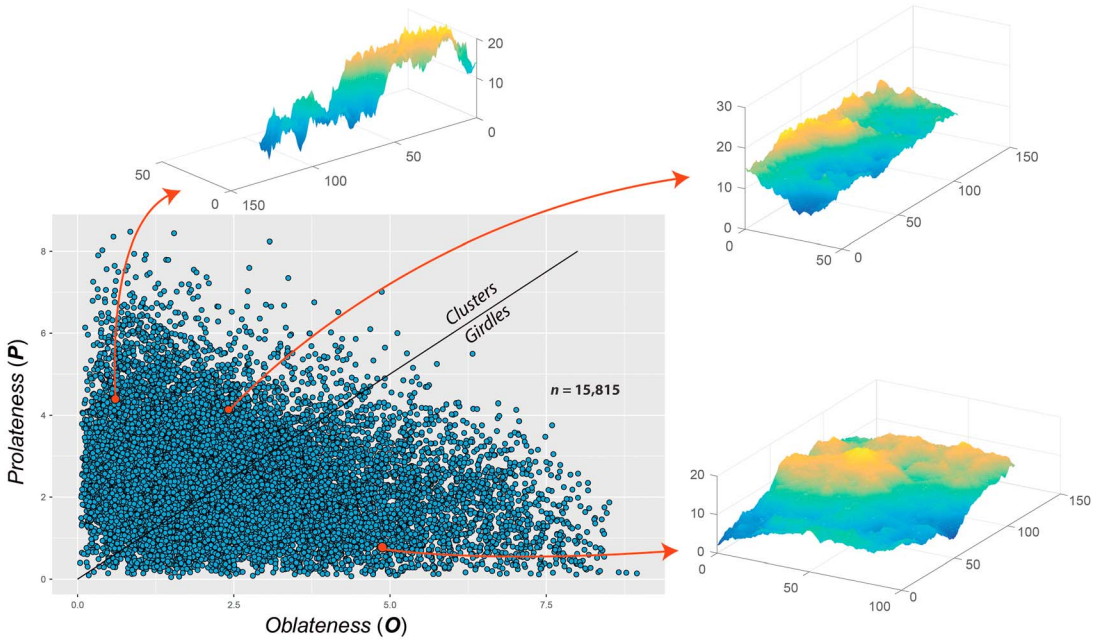

Figure 5. Eigenvalue ratio plots (Woodcock, 1977) showing the orientation tensors of the Monte Carlo-simulated subsets of fractional Brownian surfaces with Hurst exponents indicative of persistent behavior, $0.8 < H < 1$ (see text for details). Data are displayed using ggplot2 (Wickham, 2016) and rgl libraries of R language.

Nonetheless, it is a known fact that it can only correctly assess dependence when the relationship between the variables being studied is linear. At first glance, if the variables are dependent, one may attempt to transform the variables to obtain linear relationships. Then, the Pearson correlation coefficient will still be an appropriate measure of dependence. However, if the variables are dependent but no clear transformation can be done to get a linear dependence, then some other measure of dependence between two variables should be used. This matter has been studied in depth over the last few decades and the novel maximal information coefficient (MIC) introduced by Reshef et al. (2011). This nonparametric coefficient measures the association between two variables, even when they are correlated in nonlinear fashions. MIC has been largely proven to be the most adequate and flexible dependence measure (Speed, 2011). The main idea is to locally analyze the mutual information between two variables. The MIC coefficient is symmetrical and normalized into a range [0, 1]. An MIC value approaching 1 suggests a dependency between the investigated variables, whereas $MIC = 0$ describes the relationship between two independent variables.

5. Results

The results of the simulation were stored in a multivariate matrix (a full breakdown is included in the supporting information). The fundamental aim of this study was to assess the quality of the best fit solution by the eigen characteristics of the underlying point cloud. Given this purpose, we explore the relationship between these characteristics and the dispersion of the best fit plane, regarded as a measure of precision. Scatterplots for each of the shape parameters versus the dispersion of the best fit summarize the results (Figure 6). No correlation with dispersion was found in the number of bootstrap resampling ($MIC = 0.11$) nor in the sample size ($MIC = 0.12$), ruling out possible flaws in the procedure.

Overall, comparisons of values of Θ_{MAX} suggest a rise in the precision of the best fit plane with increasing vertex coplanarity ($MIC = 0.51$ and 0.53) and oblateness ($MIC = 0.88$ and 0.76), while the prolateness did not show significant dependence ($MIC = 0.12$ and 0.17), as shown in Figure 6. Surprisingly, collinearity (K), regarded in previous studies as a measure of precision, is found to be a flawed parameter to assess reliability since the correlation is not straightforward ($MIC = 0.42$ and 0.53).

In summary, the straightforward correlation found between oblateness of the point cloud underlying distribution and dispersion allows it to be proposed as a good predictor of reliability. Figure 7 summarizes the

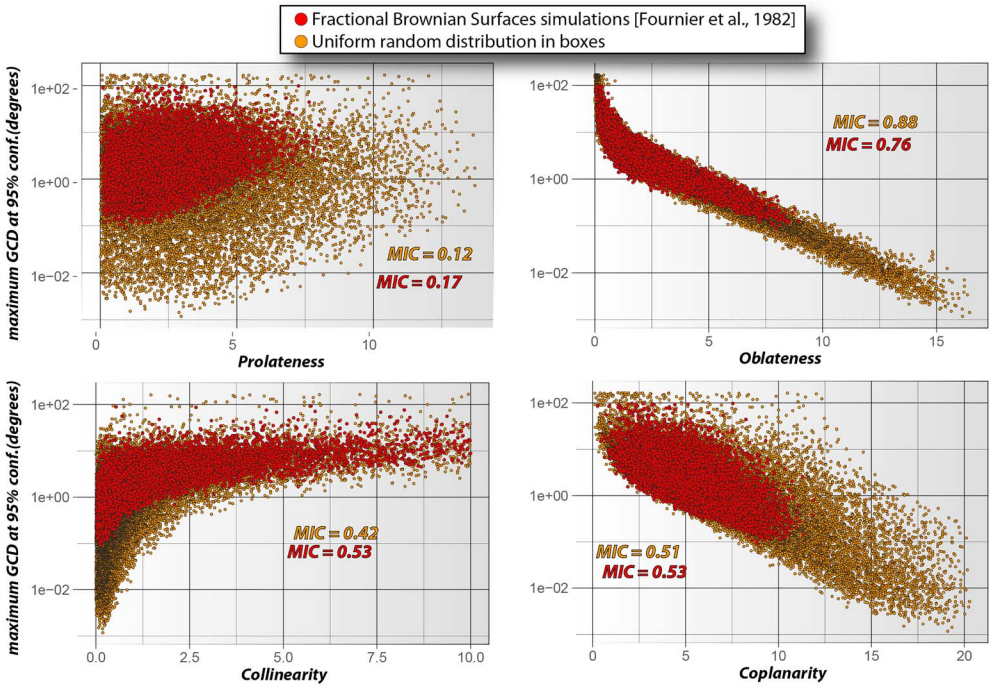

Figure 6. Scatter plot showing the relationship between the eigen characteristics lineation, foliation, collinearity, and coplanarity with dispersion of the best fit plane, represented by the quantile 95% of the $\Theta^{(b)}$ on a logarithmic scale (see text for details) for the simulations described in section 4.1 (brown dots) and section 4.2 (red dots). Note that the straightforward correlation, assessed by the maximal information coefficient (MIC; Reshef et al., 2011) between oblateness and dispersion allows it to be proposed as a potent predictor of reliability. Data are displayed using ggplot2 (Wickham, 2016) library of R language.

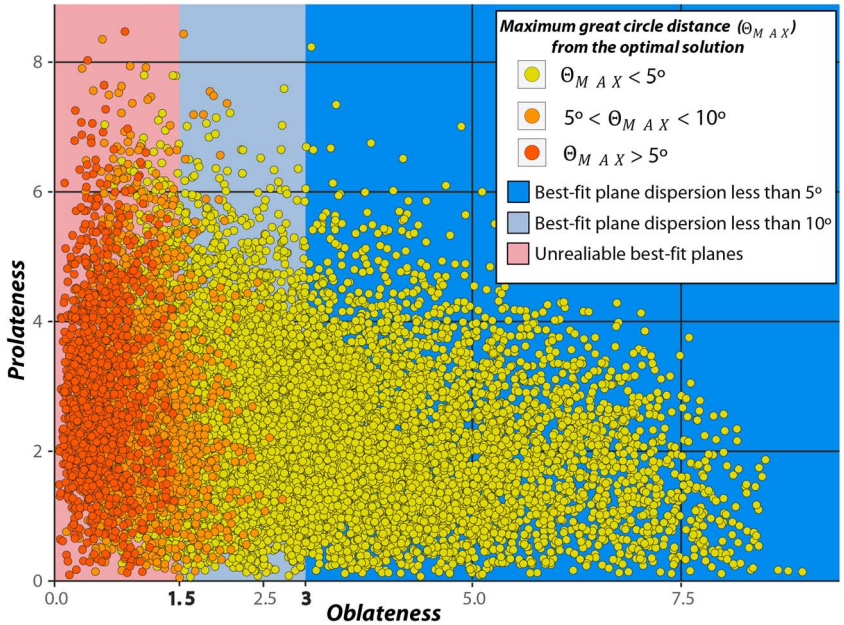

Figure 7. Eigenvalue ratio plots (Woodcock, 1977) showing the orientation tensors of Monte Carlo-simulated samples with their corresponding dispersion parameter Θ_{MAX} at 95% confidence level. Data are displayed using ggplot2 (Wickham, 2016) library of R language.

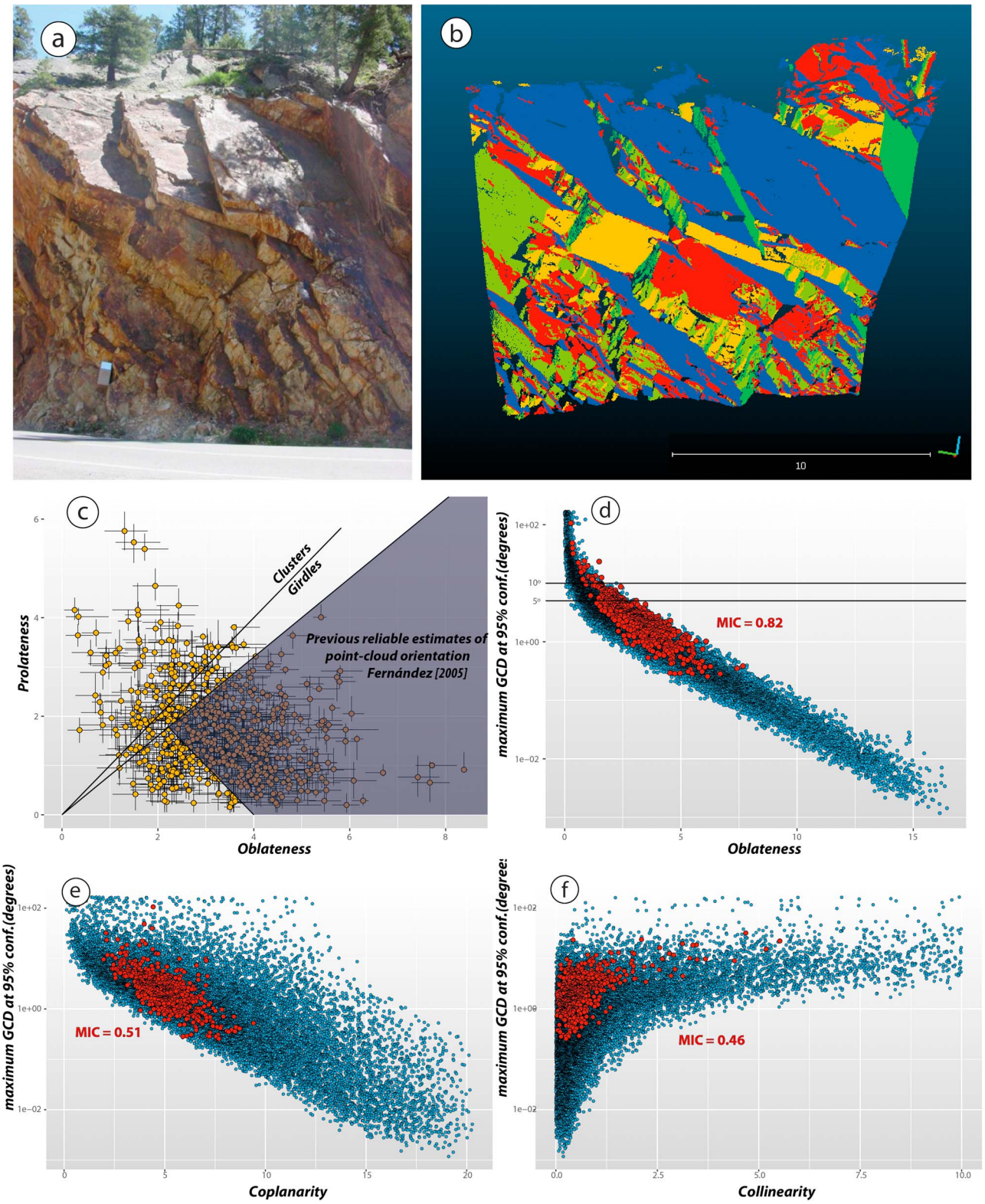

Figure 8. (a) Image of the road cut slope from the Rockbench repository. (b) Discontinuity planar facets extracted with the method proposed by Riquelme et al. (2014). The color indicates different sets of joints. (c) Point cloud oblateness [$O = \ln(\kappa_2/\kappa_3)$] versus prolateness [$O = \ln(\kappa_1/\kappa_2)$]. The variability of the bootstrapped parameters was used to estimate the 95% confidence bounds (vertical and horizontal lines), which lies between the 2.5% and 97.5% values of the empirical cumulative distribution functions (see text for details). Dependence, estimated using the maximal information coefficient (MIC), of Θ_{MAX} with respect to oblateness (d), coplanarity (e), and collinearity (f) of the extracted planar facets (red dots); blue dots represent the results of the Monte Carlo simulations.

precision of the solutions in the two-axis ratio graph of Woodcock (1977). This simulation study strongly suggests that the method proposed in this paper is valid and produces accurate results to a 95% level of confidence. Our results show that point clouds with oblateness parameter greater than 3 and 1.5 may be placed at 95% confidence levels of 5° and 10°, respectively.

6. Application to Planar Facets of a Real Outcrop

The method introduced in this work was applied to publicly available lidar data using a standalone program written in Visual Basic.NET that performs the bootstrap statistics developed in this contribution for each planar facet. This case is located in Ouray, Colorado, USA (Figure 8a). The data sets—which are publicly available at Rockbench.org (Lato et al., 2013)—consist of a 3-D point cloud on a quartzitic road cut in Ouray (Colorado). The planar facets were extracted with the method proposed by Riquelme et al. (2014; supporting information S2). The extracted discontinuity planes have very different eigen characteristics (Figure 8c, supporting information S3). The bootstrap resampling method described generates sets of bootstrap eigen estimates and the cumulative distribution function for each planar facet. A 95% bootstrap percentile confidence interval (see Chernik, 1999, for details) for these estimates is obtained by taking the 2.5th and 97.5th percentiles (between which 95% of the data lie) of the eigen estimates empirical cumulative distribution function. The bootstrap procedure is conducted in order to calculate the dispersion parameter Θ_{MAX} of each planar facet.

After performing the bootstrap of the maximum moment of inertia of the data, Θ_{MAX} dependence with oblateness is evaluated, with two considerations to highlight. First, it is found that 83%(96%) of the extracted facets may be placed at 95% confidence levels of 5°(10°), respectively (Figure 8d). Second, it is found a strong dependence of Θ_{MAX} with respect to oblateness (*MIC* = 0.82), as observed in the Monte Carlo-simulated samples. The results allowed the comparisons between the Monte Carlo-simulated samples and planar facets and verify the dependence of the precision upon the oblateness parameter.

7. Conclusions

The analysis of the moment of inertia has diverse useful applications in the geosciences. A great advance in three-dimensional orientation data came from the proliferation of digital outcrop acquisition techniques and the moment of inertia approach is a widely used way to define best fit planes from many of these types of referenced points. However, the reliability of a best fit plane relies upon the distribution of the underlying point cloud. The eigenvalues derived from the analysis proved to be a way to assess the reliability of the fit (Fernández, 2005; Seers & Hodgetts, 2016b), but its probabilistic constraining necessitates making strong assumptions about distribution statistics. To ensure the accuracy and precision of the orientation estimates, collinearity/coplanarity (*K*/*M*) were the chosen ratios to assess precision although no other alternative was sensibly tested. Here a bootstrap method for estimating the reliability of a best plane was presented. The variability of the bootstrapped vectors normal to the best fit plane, evaluated with the maximum distance from the optimal solution, gives us a robust estimate of the precision of the fit. A noteworthy result of this study is the contrasting correlation between oblateness of the point cloud and dispersion of the results. As discussed by Jones et al. (2016), we find experimentally that neither *M* nor *K* (functions of λ_1) is suitable for assessing the quality of a best fit. Instead it is argued that oblateness ($O = \ln(\lambda_2/\lambda_3)$) is a suitable parameter to assess precision. We argue that georeferenced point clouds with an oblateness (*O*) parameter greater than 3 and 1.5 may be placed at 95% confidence levels of 5° and 10°, respectively, and we propose using these values as thresholds to obtain robust best fit planes from georeferenced data, guaranteeing reproducible results for scientific research.

Acknowledgments
We thank the Universidad de Buenos Aires (UBA) and CONICET for the funding that enabled the conduction of these studies and L.C.G. Doctoral Grant. E. O. C. acknowledges founding from PICT 2013-1309 and BACYT 20020130100613BA. T. D. Seers provided an insightful and helpful review; we are very grateful. We also thank the Editor, Paul Tregoning; the Associate Editor, Andy Hooper; and an anonymous reviewer. The authors would like to thank Renata N. Tomezzoli for constructive discussions. The data used are available in tables, cited references, and supporting information.

References

Baker, M. (2016). Statisticians issue warning over misuse of P values. *Nature*, *531*(7593), 151. https://doi.org/10.1038/nature.2016.19503

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., et al. (2018). Redefine statistical significance. *Nature Human Behaviour*, *2*(1), 6–10. https://doi.org/10.1038/s41562-017-0189-z

Benn, D. I. (1994). Fabric shape and the interpretation of sedimentary fabric data. *Journal of Sedimentary Research*, *64A*(4), 910–915. https://doi.org/10.1306/D4267F05-2B26-11D7-8648000102C1865D

Biber, K., Khan, S. D., Seers, T. D., Sarmiento, S., & Lakshmikantha, M. R. (2018). Quantitative characterization of a naturally fractured reservoir analog using a hybrid lidar-gigapixel imaging approach. *Geosphere*, *14*(2), 710–730. https://doi.org/10.1130/GES01449.1

Chatzaras, V., Kruckenberg, S. C., Cohen, S. M., Medaris, L. G., Withers, A. C., & Bagley, B. (2016). Axial-type olivine crystallographic preferred orientations: The effect of strain geometry on mantle texture. *Journal of Geophysical Research: Solid Earth*, *121*, 4895–4922. https://doi.org/10.1002/2015JB012628

Chen, J., Zhu, H., & Li, X. (2016). Automatic extraction of discontinuity orientation from rock mass surface 3D point cloud. *Computers and Geosciences*, *95*, 18–31. https://doi.org/10.1016/j.cageo.2016.06.015

Chernik, M. (1999). *Bootstrap methods: A practitioner's guide*. New York: John Wiley.

Constable, C., & Tauxe, L. (1990). The bootstrap for magnetic susceptibility tensors. *Journal of Geophysical Research*, *95*(B6), 8383–8395. https://doi.org/10.1029/JB095iB06p08383

Dueholm, K. S., & Olsen, T. (1993). Reservoir analog studies using multimodel photogrammetry: A new tool for the petroleum industry. *American Association of Petroleum Geologists Bulletin*, *77*. https://doi.org/10.1306/BDFF8FBA-1718-11D7-8645000102C1865D

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, *7*(1), 1–26. https://doi.org/10.1214/aos/1176344552

Fernández, O. (2005). Obtaining a best fitting plane through 3D georeferenced data. *Journal of Structural Geology*, *27*(5), 855–858. https://doi.org/10.1016/j.jsg.2004.12.004

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*(4), 631–665. https://doi.org/10.1007/s00603-008-0010-4

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, *24*(6), 381–395. https://doi.org/10.1145/358669.358692

Fisher, N. I., Lewis, T., & Embleton, B. J. J. (1987). Statistical analysis of spherical data, 343. https://doi.org/10.1017/CBO9780511623059

Flinn, B. Y. D. (1962). On folding during three-dimensional progressive deformation. *Geological Society of London Quarterly Journal*, *118*(1-4), 385–428. https://doi.org/10.1144/gsjgs.118.1.0385

Fournier, A., Fussell, D., & Carpenter, L. (1982). Computer rendering of stochastic models. *Communications of the ACM*, *25*(6), 371–384. https://doi.org/10.1145/358523.358553

Gallo, L. C., Tomezzoli, R. N., & Cristallini, E. O. (2017). A pure dipole analysis of the Gondwana apparent polar wander path: Paleogeographic implications in the evolution of Pangea. *Geochemistry, Geophysics, Geosystems*, *18*, 1499–1519. https://doi.org/10.1002/2016GC006692

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 and Geosciences*, *37*(10), 1584–1594. https://doi.org/10.1016/j.cageo.2011.03.007

Gomes, R. K., De Oliveira, L. P. L., Gonzaga, L., Tognoli, F. M. W., Veronez, M. R., & De Souza, M. K. (2016). An algorithm for automatic detection and orientation estimation of planar structures in LiDAR-scanned outcrops. *Computers and Geosciences*, *90*, 170–178. https://doi.org/10.1016/j.cageo.2016.02.011

Gough, A. J., Mahoney, A. R., Langhorne, P. J., Williams, M. J. M., & Haskell, T. G. (2012). Sea ice salinity and structure: A winter time series of salinity and its distribution. *Journal of Geophysical Research*, *117*, C03008. https://doi.org/10.1029/2011JC007527

Jones, R. R., Pearce, M. A., Jacquemyn, C., & Watson, F. E. (2016). Robust best-fit planes from geospatial data. *Geosphere*, *12*(1), 196–202. https://doi.org/10.1130/GES01247.1

Ketcham, R. A. (2005). Three-dimensional grain fabric measurements using high-resolution X-ray computed tomography. *Journal of Structural Geology*, *27*(7), 1217–1228. https://doi.org/10.1016/j.jsg.2005.02.006

Kirschvink, J. L. (1980). The least-squares line and plane and the analysis of paleomagnetic data. *Geophysical Journal of the Royal Astronomical Society*, *62*(3), 699–718. https://doi.org/10.1111/j.1365-246X.1980.tb02601.x

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*(1), 194–199. https://doi.org/10.1016/j.ijrmms.2008.04.007

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 and Geosciences*, *50*, 106–114. https://doi.org/10.1016/j.cageo.2012.06.014

Li, X., Chen, J., & Zhu, H. (2016). A new method for automated discontinuity trace mapping on rock mass 3D surface model. *Computers and Geosciences*, *89*, 118–131. https://doi.org/10.1016/j.cageo.2015.12.010

McCaffrey, K. J. W., Jones, R. R., Holdsworth, R. E., Wilson, R. W., Clegg, P., Imber, J., et al. (2005). Unlocking the spatial dimension: Digital technologies and the future of geoscience fieldwork. *Journal of the Geological Society*, *162*(6), 927–938. https://doi.org/10.1144/0016-764905-017

McPherron, S. P. (2017). Additional statistical and graphical methods for analyzing site formation processes using artifact orientations coming from total station measurements. *PLoS One*, *13*(1), 1–21. https://doi.org/10.1371/journal.pone.0190195

Méheust, Y., & Schmttbuhl, J. (2001). Geometrical heterogeneities and permeability anisotropy of rough fractures. *Journal of Geophysical Research*, *106*(B2), 2098–2102. https://doi.org/10.1029/2000JB900306

Ouillon, G., Ducorbier, C., & Sornette, D. (2008). Automatic reconstruction of fault networks from seismicity catalogs: Three-dimensional optimal anisotropic dynamic clustering. *Journal of Geophysical Research*, *113*, B01306. https://doi.org/10.1029/2007JB005032

Palit, A., Williams, M. A., Turley, G. A., Renkawitz, T., & Weber, M. (2017). Femur First navigation can reduce impingement severity compared to traditional free hand total hip arthroplasty. *Scientific Reports*, *7*(1), 7238–7239. https://doi.org/10.1038/s41598-017-07644-4

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1986). *Numerical recipes: The art of scientific computing*. Cambridge, MA: Cambridge University Press.

Pringle, J. K., Howell, J. A., Hodgetts, D., Westerman, A. R., & Hodgson, D. M. (2006). Virtual outcrop models of petroleum reservoir analogues: A review of the current state-of-the-art. *First Break*, *24*(1093), 33–42. https://doi.org/10.3997/1365-2397.2006005

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., et al. (2011). Detecting novel associations in large data sets. *Science*, *334*(6062), 1518–1524. https://doi.org/10.1126/science.1205438

Riquelme, A. J., Abellán, A., & Tomás, R. (2015). Discontinuity spacing analysis in rock masses using 3D point clouds. *Engineering Geology*, *195*, 185–195. https://doi.org/10.1016/j.enggeo.2015.06.009

Riquelme, A. J., Abellán, A., Tomás, R., & Jaboyedoff, M. (2014). A new approach for semi-automatic rock mass joints recognition from 3D point clouds. *Computers and Geosciences*, *68*, 38–52. https://doi.org/10.1016/j.cageo.2014.03.014

Santangelo, M., Marchesini, I., Cardinali, M., Fiorucci, F., Rossi, M., Bucci, F., & Guzzetti, F. (2015). A method for the assessment of the influence of bedding on landslide abundance and types. *Landslides*, *12*(2), 295–309. https://doi.org/10.1007/s10346-014-0485-x

Scheidegger, A. E. (1965). On the statistics of the orientation of bedding planes, grain axes, and similar sedimentological data. *U.S. Geological Survey Professional Paper*, *525*, 164–167.

Schmidt, P. W. (1985). Bias in converging great circle methods. *Earth and Planetary Science Letters*, *72*(4), 427–432. https://doi.org/10.1016/0012-821X(85)90063-9

Seers, T. D., & Hodgetts, D. (2014). Comparison of digital outcrop and conventional data collection approaches for the characterization of naturally fractured reservoir analogues. *Geological Society Special Publication*, *374*(1), 51–77. https://doi.org/10.1144/SP374.13

Seers, T. D., & Hodgetts, D. (2016a). Extraction of three-dimensional fracture trace maps from calibrated image sequences. *Geosphere*, *12*(4), 1323–1340. https://doi.org/10.1130/GES01276.1

Seers, T. D., & Hodgetts, D. (2016b). Probabilistic constraints on structural lineament best fit plane precision obtained through numerical analysis. *Journal of Structural Geology*, *82*, 37–47. https://doi.org/10.1016/j.jsg.2015.11.004

Slob, S., van Knapen, B., Hack, H. R. G. K., Turner, K., & Kemeny, J. (2005). Method for automated discontinuity analysis of rock slopes with three-dimensional laser scanning. *Transportation Research Record*, *1913*(1), 187–194. https://doi.org/10.3141/1913-18

Speed, T. (2011). A correlation for the 21st century. *Science*, *334*(6062), 1502–1503. https://doi.org/10.1126/science.1215894

Telling, J., Lyda, A., Hartzell, P., & Glennie, C. (2017). Review of Earth science research using terrestrial laser scanning. *Earth-Science Reviews*, *169*(November 2016), 35–68. https://doi.org/10.1016/j.earscirev.2017.04.007

Thiele, S. T., Grose, L., Samsu, A., Micklethwaite, S., Vollgger, S. A., & Cruden, A. R. (2017). Rapid, semi-automatic fracture and contact mapping for point clouds, images and geophysical data. *Solid Earth*, *8*(6), 1241–1253. https://doi.org/10.5194/se-8-1241-2017

Vasuki, Y., Holden, E. J., Kovesi, P., & Micklethwaite, S. (2014). Semi-automatic mapping of geological structures using UAV-based photogrammetric data: An image analysis approach. *Computers and Geosciences*, *69*, 22–32. https://doi.org/10.1016/j.cageo.2014.04.012

Wang, Y., Ouillon, G., Woessner, J., Sornette, D., & Husen, S. (2013). Automatic reconstruction of fault networks from seismicity catalogs including location uncertainty. *Journal of Geophysical Research: Solid Earth*, *118*, 5956–5975. https://doi.org/10.1002/2013JB010164

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer. https://doi.org/10.1007/978-3-319-24277-4

Wilson, C. E., Aydin, A., Karimi-Fard, M., Durlofsky, L. J., Sagy, A., Brodsky, E. E., et al. (2011). From outcrop to flow simulation: Constructing discrete fracture models from a LIDAR survey. *AAPG Bulletin*, *95*(11), 1883–1905. https://doi.org/10.1306/03241108148

Woodcock, N. H. (1977). Specification of fabric shapes using an eigenvalue method. *Geological Society of America Bulletin*, *88*(9), 1231–1236. https://doi.org/10.1130/0016-7606(1977)88<1231:SOFSUA>2.0.CO;2

Yamaji, A., Sato, K., & Tonai, S. (2010). Stochastic modeling for the stress inversion of vein orientations: Paleostress analysis of Pliocene epithermal veins in southwestern Kyushu, Japan. *Journal of Structural Geology*, *32*(8), 1137–1146. https://doi.org/10.1016/j.jsg.2010.07.001