

AUTHOR QUERY FORM

 ELSEVIER	Journal: CAGEO Article Number: 104322	Please e-mail your responses and any corrections to: E-mail: corrections.esch@elsevier.tnq.co.in
------------------------------------------------------------------------------------------------------	------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list. **It is crucial that you NOT make direct edits to the PDF using the editing tools as doing so could lead us to overlook your desired changes.** Note: if you opt to annotate the file with software other than Adobe Reader then please also highlight the appropriate place in the PDF file. To ensure fast publication of your paper please return your corrections within 48 hours.

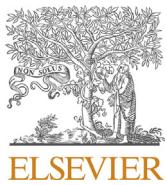
For correction or revision of any artwork, please consult <http://www.elsevier.com/artworkinstructions>.

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof.

Location in article	Query / Remark: Click on the Q link to find the query's location in text Please insert your reply or correction at the corresponding line in the proof
Q1	Please confirm that the provided email "mimichalak@us.edu.pl" is the correct address for official communication, else provide an alternate e-mail address to replace the existing one, because private e-mail addresses should not be used in articles as the address for communication.
Q2	The country name has been inserted for the second author's affiliation [c]. Please check, and correct if necessary.
Q3	The citation "Fisher, 1993" has been changed to match the author name in the reference list. Please check here and in subsequent occurrences.
Q4	Have we correctly interpreted the following funding source(s) and country names you cited in your article: Ministry of Science and Higher Education, Poland?
Q5	Please confirm that given names and surnames have been identified correctly and are presented in the desired order and please carefully verify the spelling of all authors' names.
Q6	Your article is registered as a regular item and is being processed for inclusion in a regular issue of the journal. If this is NOT correct and your article belongs to a Special Issue/Collection please contact d.perumal@elsevier.com immediately prior to returning your corrections.

Please check this box or indicate
your approval if you have no
corrections to make to the PDF file

Thank you for your assistance.



Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo1
2
3
4

Highlights

- Delaunay triangulation was used to generate local orientation measurements.
- Inertia moment analysis and the mean vector approach were used.
- Cluster analysis was applied to direction cosines for denoising.
- Robustness was compared between different clustering algorithms.

<https://doi.org/10.1016/j.cageo.2019.104322>

Available online xxxx
0098-3004/© 2019 Elsevier Ltd. All rights reserved.

5
6
7
8
9
10
11
12
13
14
15
16
17



Using delaunay triangulation and cluster analysis to determine the orientation of a sub-horizontal and noise including contact in Kraków-Silesian Homocline, Poland

Michał Michalak^{a,*}, Waldemar Bardziński^{b,2}, Lesław Teper^{a,3}, Zbigniew Małolepszy^{c,4}

^a Department of Applied Geology, Faculty of Earth Sciences, University of Silesia in Katowice, Będzińska 60, 41-205, Sosnowiec, Poland

^b Department of Fundamental Geology, Faculty of Earth Sciences, University of Silesia in Katowice, Będzińska 60, 41-205, Sosnowiec, Poland

^c Polish Geological Institute, National Research Institute Upper Silesian Branch, Królowej Jadwigi 1, 41-200, Sosnowiec, Poland

ARTICLE INFO

Keywords:

Moment of inertia

Mean vector

Orientation

K-means clustering

Hierarchical clustering

ABSTRACT

In this work, we combined a three-point problem with Delaunay triangulation to determine the average orientation of a sub-horizontal contact lying within the Kraków-Silesian Homocline, Poland. This contact was assumed to represent the regional trend due to the conformable or sub-conformable relationships between the geological units. The approach presented involved computing the local orientation of Delaunay triangles that represented the investigated surface. A C++ application was developed to compute the required figures, to which we added computer code that is open source and freely available. The pre-processing stage required the removal of collinear configurations that contributed to floating-point arithmetic inaccuracies. We then assigned dip angle and direction to the Delaunay triangles and performed a stereographic projection of the unit normal vectors. For statistical analysis, we conducted inertia moment analysis and followed the mean vector approach. As a part of exploring the orientation data—and as a way of achieving better consistency between stereonet results—we used several clustering algorithms: *k*-means, *k*-medoids and hierarchical. We indicated that the Euclidean distance could be beneficial for extracting the dominant orientation calculated for the sub-population assumed to represent the regional trend. We concluded that considering four clusters and the combination of the Euclidean distance and Ward linkage methods gave us the best extraction results for the dominant orientation. We identified limitations to the proposed approach relating to the lack of statistical information on the calculated orientations and suggested potential extensions to the research, including mixture models and investigation of spatial patterns.

1. Introduction

1.1. The general motivation

Determining the geometry of contacts between geological units plays a significant role in geological modelling (Caumon et al., 2009; Chilès et al., 2004; Lajaunie et al., 1997; Pakyuz-Charrier et al., 2018). Various methods along with the geological modelling assumptions allow the information about the geometry of contacts to be obtained. On the other

hand, geological modelling can benefit from this fundamental information in more challenging cases. They may, for example, involve considerations related to ore deposits (retrieving true thickness) or hydrogeology (direction of groundwater flow). Boundary surface orientation seems to be of even greater interest in cases of conformable or sub-conformable stratigraphic relationships (Calcagno et al., 2008; Lindsay et al., 2012) which is sometimes referred to as layer-cake stratigraphy (Thiele et al., 2016). The purpose of this paper was to make use of sub-surface geological information over a wide area in Poland to

* Corresponding author.

E-mail address: mimichalak@us.edu.pl (M. Michalak).

¹ Michał Michalak devised the project, wrote the manuscript, performed the computations and discussed the results.

² Waldemar Bardziński shared and bought selected materials (original borehole records, maps and books), revised the manuscript critically for important intellectual content and discussed the results.

³ Lesław Teper revised the manuscript critically for important intellectual content, discussed the results and supervised the project.

⁴ Zbigniew Małolepszy was responsible for the data acquisition (providing a digitised version of 820 borehole records), and revised the manuscript critically for important intellectual content.

<https://doi.org/10.1016/j.cageo.2019.104322>

Received 26 October 2018; Received in revised form 31 August 2019; Accepted 9 September 2019

Available online xxxx

0098-3004/© 2019 Elsevier Ltd. All rights reserved.

introduce a new method for determining the orientation distribution of a geological contact in the Kraków-Silesian Homocline.

1.2. Existing methods

Methods aiming to compute the geometry of contacts can incorporate the positions of points in 3D space (3D points) representing the contact surface (Jones et al., 2016; Mallet, 1989, 1992), or they may include field orientation measurements in the form of vectors in 3D space (3D vectors) (Chilès et al., 2004; Lajaunie et al., 1997; Pakyuz-Charrier et al., 2018). The first method can address the least-squares approach, aiming to find a best-fitting plane through a set of 3D points (e.g. Golub and Van Loan, 1980; Jones et al., 2016). The methods based on using 3D points as input can also refer to interpolation methods such as the Discrete Smooth Interpolation (DSI) method. The DSI method has been designed to model complex surfaces with discontinuities (Mallet, 1989, 1992), while one of its disadvantages is that the model's uncertainties cannot be determined (Mallet, 1989). Methods using 3D points along with 3D vectors (co-kriging) allow construction of a surface that passes through the 3D points of a contact, and that is compatible with orientation measurements (Chilès et al., 2004; Lajaunie et al., 1997; Lindsay et al., 2012). However, field orientation measurements can only be conducted on outcropping surfaces with substantial dip, as sub-horizontal surfaces are difficult to measure using the standard, traditional compass-clinometer approach.

Regardless of the form of the input data (3D points or 3D points with 3D vectors), the methods above also differ with respect to the results they produce. While the least-squares approach supplies the equation for the best-fitting plane passing through the given 3D points, interpolation or co-kriging methods generate a contour map representing the geometry of the investigated surface.

1.3. Combining the three-point approach with delaunay triangulation

The above methods for determining the geometry of geological contacts are not specifically designed, however, to conduct a comprehensive investigation of orientation measurements linked to specific sub-areas of investigated surfaces. Therefore, we offer a new approach allowing the quantitative aspect of the least-square approach to be combined with the more illustrative goals of the interpolation or co-kriging methods capable of tracing local changes of orientation. In our research, the resulting mean orientation is a result of averaging over the sub-population grouping similar local orientations assumed to best represent the investigated surface.

From a technical viewpoint, our approach requires the supply of 3D points as input data and is expected to produce 3D vectors as outputs. The process of obtaining local orientation in the form of 3D vectors is achieved by the combination of the three-point approach and Delaunay triangulation. The former is expected to compute the orientation of a plane, using three non-collinear points; it has many variants (Fienan, 2005; Groshong, 2006; Haneberg, 1990; Paor, 1991; Vacher, 1989) that differ in the sequences of their computational steps and in the paradigm of the computer code used. Delaunay triangulation is an algorithm that partitions the investigated surface into triangular sub-areas, allowing the three-point approach to be applied at a larger scale, while another of its inherent properties is that it minimises the roughness of the resulting surface (Rippa, 1990; see also Methods). Therefore, surfaces generated by using Delaunay triangulation can be regarded as the *smoothest* (in the sense of Rippa).

1.4. Numerical restrictions

Singularities related to floating-point arithmetic inaccuracies are of great importance (Goldberg, 1991): they arise in collinear configurations, and because they have been proven to provide unreliable results, they should be removed from further analysis—removal that can be

carried out on the basis of the introduced collinearity coefficient (Michałak, 2018). The above limitations should not obscure the general motivation for incorporating Delaunay triangulation into determining the orientation of the investigated contact, i.e., that its main advantage is that the processes of orientation sampling, and generating the *smoothest* (in the sense of Rippa) surface on the basis of the given 3D points, can be conducted at the same time.

1.5. Outline of the exploratory analysis

In our approach, the analysis of orientation data addresses only the initial, *exploratory* phase. During this phase, some fundamental questions related to the underlying orientation distribution can be raised, including over whether it is unimodal, multimodal or has no mode at all, or whether it is rotationally symmetric about a mean. The answers to these questions are crucial to the subsequent *quantitative* phase of orientation data analysis; that is, to the proper selection of the spherical distribution underlying the processed orientation data. For example, using the von Mises–Fisher distribution requires the distribution to be unimodal and rotationally symmetric about the mean. This is not the case with the Kent distribution, which is capable of dealing with bimodal and asymmetric distributions as well (Fisher et al., 1993). Using probabilistic models allows statistical information to be assigned to the results obtained, however the main difficulty with using these models is that in-depth knowledge of the orientation distribution does not always exist. Moreover, probabilistic models require the parameters of the model to be specified and, if conducted without reasonable premises, they can introduce bias into the analysis. Therefore, we have suggested using cluster analysis as part of the *exploratory* phase of orientation data analysis, to provide a preliminary and non-statistical insight into the investigated data set.

1.6. Uncertainties related to the geological knowledge

In our study, we investigated a sub-conformable contact that separates younger ore-bearing clay deposits from older sands and sandstones within the Kraków-Silesian Homocline (Fig. 1 see also Section 2 Geological Setting). To the NW, this geological unit borders on a greater geological unit called the Fore-Sudetic Homocline; while, because the western range of the Kraków-Silesian Homocline has not been established, only rough boundaries for this geological unit can be suggested (for details, see Żelaźniewicz et al., 2011). From a geometric viewpoint, there is assumed to be a general parallelism of the sediments, dipping gently to the N (Gedl and Kaim, 2012) or NE (Górecka, 1993; Wilk, 1989). More detailed studies supply an estimation of the dip angle as:

- not $>2^\circ$ (Marynowski et al., 2007; Znosko, 1960),
- not $>3^\circ$ (Szubert, 2012),
- lying in the interval of $2\text{--}5^\circ$ (Bardziński et al., 1986), or $3\text{--}5^\circ$ (Gedl and Kaim, 2012).

In the majority of these studies, the orientation measurement methodology was not explicitly revealed (Bardziński et al., 1986; Górecka, 1993; Marynowski et al., 2007; Wilk, 1989; Znosko, 1960). Because Rutkowski (1989) stated that the orientation of these deposits was generally not measurable in outcrops, due to their low dip, it could be supposed that these estimations were based on either the direction of younger deposits, on geological maps (Szubert, 2012), or on observations of a speculative nature carried out on outcrops (Gedl and Kaim, 2012). Uncertainty related to the geometry of these sediments is intensified by the presence of abrupt changes of orientation (Fig. 2), as is traditionally associated with NE–SW normal faults (Bardziński et al., 1986; Szubert, 2012).

1.7. Exploratory analysis in our case

Although the above geological information sounds imprecise, the

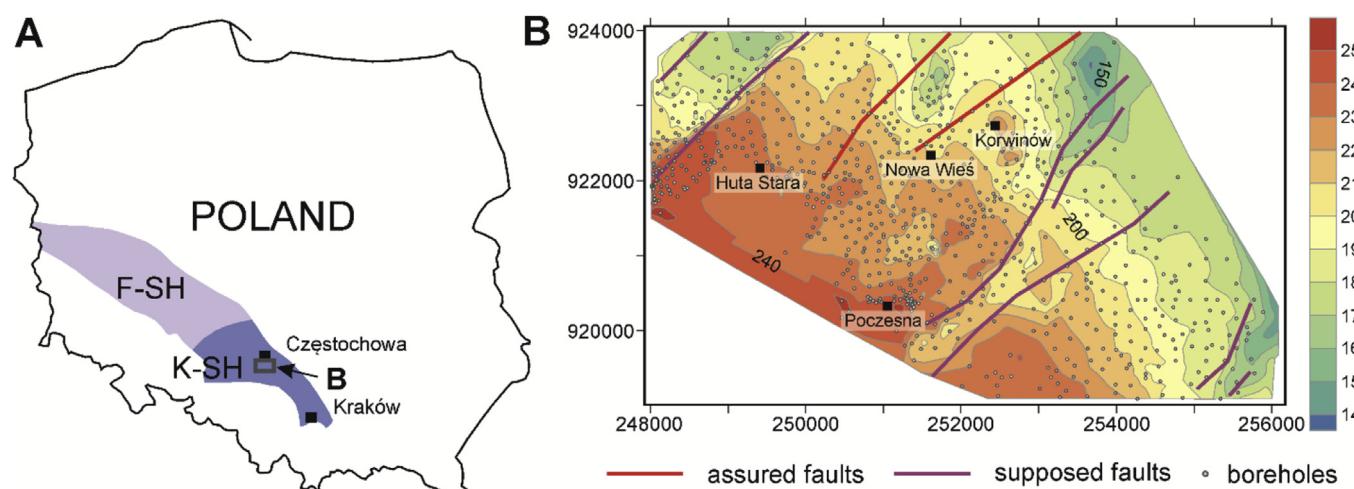


Fig. 1. Location of the investigated Middle-Jurassic ore-bearing clay deposits with 820 boreholes in the vicinity of Częstochowa: (A) the outline of the location with two major geological units: F-SH (Fore-Sudetic Homocline) and K-SH (Kraków–Silesian Homocline); (B) the contour map of the investigated surface with additional (spatial and geological) information.

outline of the geological knowledge presented above does answer one of the fundamental question related to the *exploratory* analysis; namely it states that the orientation distribution is, *in general*, unimodal. To confirm the legitimacy of this assumption and to provide a rough estimation of the mode, *inertia moment* analysis can be applied. From a technical viewpoint, orientation measurements (dip angle, dip direction) must be converted into 3D unit vectors. The Cartesian coordinates of these vectors were regarded as direction cosines, and were used in the *orientation matrix*, while the matrix eigenvectors were computed to determine the direction towards which the “moment of inertia” of the distribution was minimised. The eigenvector associated with the largest eigenvalue was then assumed to be the mean orientation (Fara and Scheidegger, 1961; Scheidegger, 1965a, 1965b; Watson 1965, 1966). To appraise the “strength” of the preferred orientation and the shape of the stereonet plot, Woodcock (1977) suggested two measures based on the eigenvalues of the *orientation matrix* described above.

1.8. Cluster analysis and spherical data

As indicated in subsection 1.5, a priori knowledge about the modality of the underlying spherical distribution does not always exist, and in fact, multimodal distribution can invalidate the results of the *inertia moment analysis* (Fisher et al., 1993), if conducted for all observations. Moreover, even if the distribution was assumed to be unimodal *in general*, the presence of outliers makes it difficult to treat all observations by using the same computational routines. Therefore, it has often been assumed that it was most appropriate to split the investigated observations into clusters, to allow more thorough investigation.

In geological considerations, cluster analysis is used to explain the variability of planar observation related to discontinuities (Assali et al., 2014; Chen et al., 2016; Duela Viana et al., 2016; Hammah and Curran, 1999; Jimenez-Rodriguez and Sitar, 2006; Marcotte and Henry, 2002; Tokhmechi et al., 2011; Zhan et al., 2017a; Zhou and Maerz, 2002), and the role of the assumptions underlying the clustering procedure has sometimes been explicitly revealed. For example, Hammah and Curran (1999), and Zhan et al. (2017a), assumed that sub-vertical planes having a dip direction difference of 180° should be regarded as similar and, consequently, clustered into one set—an assumption that has major consequences for the proper selection of a distance function. Hammah and Curran (1999) also noticed that high Euclidean distance values between sub-vertical planes with opposite dip directions were disproportionate to their relatively small angular distance (Fig. 3); therefore, in the problem of clustering discontinuities, distances based on angular

similarity have been applied (Assali et al., 2014; Chen et al., 2016; Hammah and Curran, 1999; Jimenez-Rodriguez and Sitar, 2006; Zhan et al., 2017a).

As such, we believe that our approach can also be beneficial for a treatment of outliers. Traditionally, this has been done as part of the quantitative phase, in which spherical distributions (Carmichael and Ailleres, 2016) or statistical tests were used (Fisher et al., 1993), however, we suggest that *cluster analysis* can be incorporated into the initial, *exploratory* phase of orientation data analysis.

1.9. Summary of the proposed combination

The assumed neglect of dip direction, as presented in subsection 1.8, was not legitimate in our case, however, and in fact we have assumed that planes dipping in the opposite (SW) direction should be regarded as outliers, despite their *geomorphological* similarity to those compatible with the preferred orientation (e.g. Znosko, 1960; Marynowski et al., 2007). Therefore, we have suggested using distances related to the Euclidean distance, since they can be more efficient in cases with similar underlying assumptions (unimodal orientation distribution) (Fig. 3).

The motivation and assumptions concerning the combination of three-point problem, Delaunay triangulation and cluster analysis can be summarised as follows:

1. Delaunay triangulation allows orientation data sampling from a *minimal roughness* surface (Rippa, 1990).
2. The geological units were assumed to conform to a general parallelism (Zakrzewski, 1976). This allowed the average orientation of the investigated contact to be regarded as a legitimate representative of the regional trend.
3. There was only one preferred orientation of a contact. Therefore, planes lying within fault zones, or dipping in the opposite direction, should be regarded as outliers. Distances related to Euclidean distance can be helpful in following this assumption (e.g. Hammah and Curran, 1999; Zhan et al., 2017a).
4. We used partitional algorithms (*k*-means, *k*-medoids) and hierarchical clustering with different linkage methods (Hastie et al., 2009; James et al., 2013), and compared their efficiency for retrieving information related to the average orientation.

2. Geological setting

The Jurassic deposits that were investigated are located within the

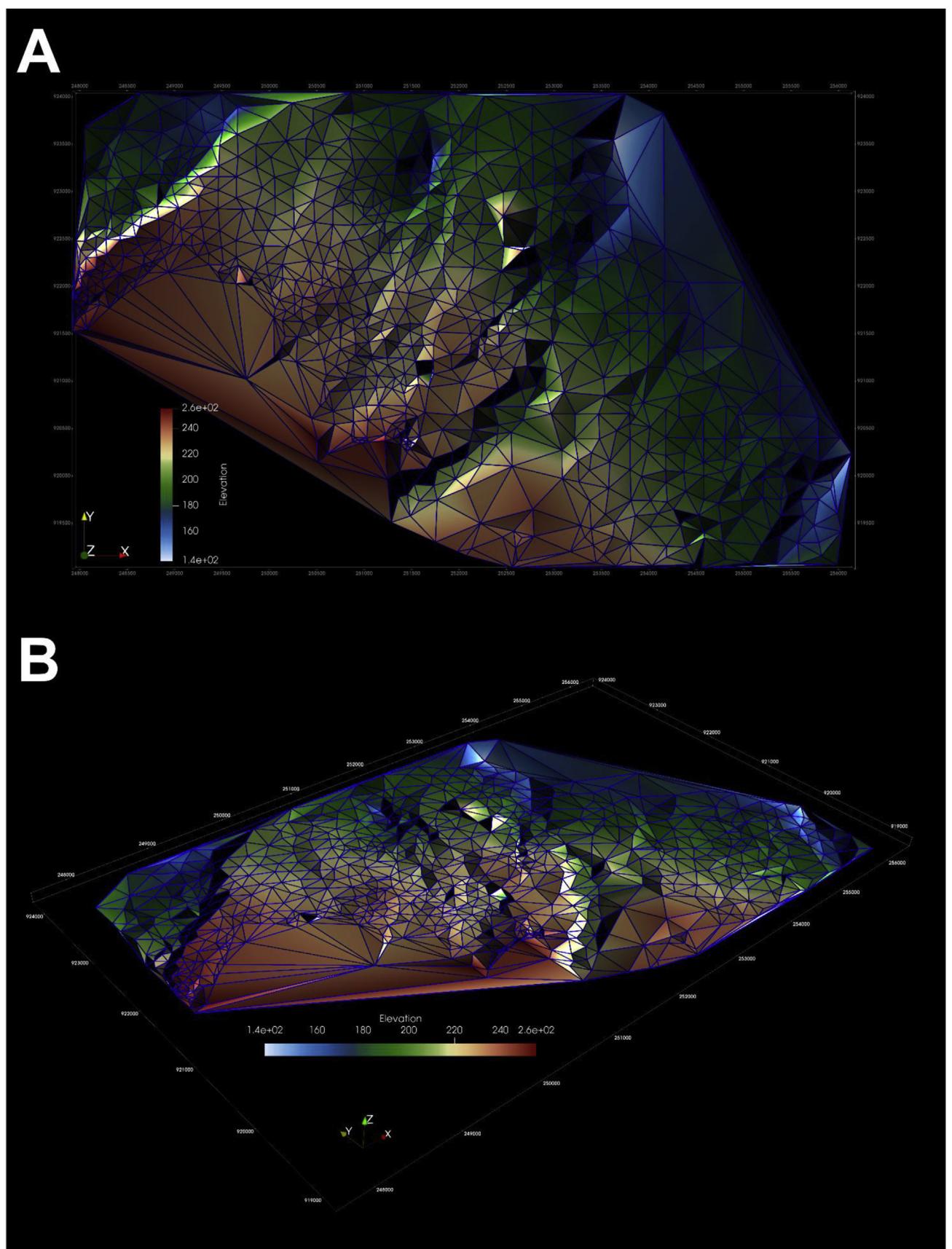


Fig. 2. A 2.5D visualization of the Delaunay triangulation. To exhibit different orientation patterns in the sub-horizontal conditions, two different positions of cameras and lighting were applied: (A) A top view with light coming from the NW; (B) Camera located to the SW of the investigated area with light coming from the SE.

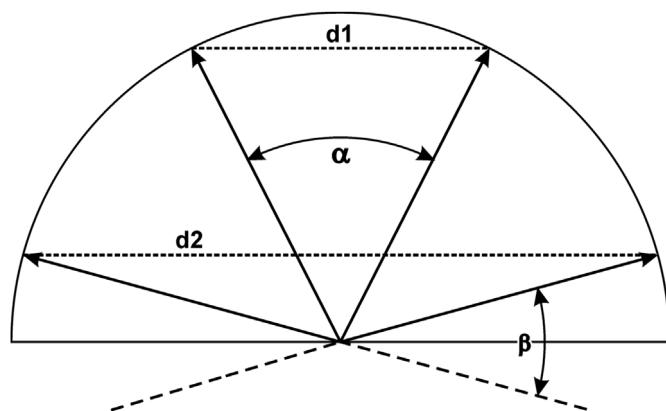


Fig. 3. Euclidean and angular distances on the sphere. Two pairs of vectors are presented having a similar angular distance but a significant difference in Euclidean distance. The latter can be therefore regarded as a better distance function, identifying planes dipping in opposite directions more adequately.

Kraków–Silesian Homocline, which is a geological unit sometimes referred to as the Permo–Mesozoic cover. It lies discordantly on folded Paleozoic rocks, with a gentle dip towards the NE of the Mesozoic sediments, that has been present since the Late Cimmerian Phase (Górecka, 1993), although it was the Alpine orogeny (Laramide phase) in the Late Cretaceous that ultimately determined their orientation (Górecka, 1993; Rutkowski, 1989). In our study, we computed the orientation of a surface that separates Late Bajocian–Late Bathonian, ore-bearing clays from the Aalenian–Early Bajocian sand and sandstones. The ore-bearing clays are often referred to as the Ore-Bearing Częstochowa Clay Formation (e.g. Kopik, 1998; Matyja and Wierzbowski, 2000), and they have been exposed in several clay pits. These deposits are composed of dark grey mudstone with horizons of siderite and calcareous concretions, and siderite bands (Leonowicz, 2015). The lower parts of the ore-bearing clays have been exploited as a source of iron from the Middle Ages up until the early 1980s (Wikowska, 2012).

There is a hiatus between the Late Bajocian–Late Bathonian ore-bearing clays and the Aalenian–Early Bajocian sand and sandstones. It covers the earliest Late Bajocian and is evidenced by the lack of a *Strenoceras subfurcatum* Ammonite Zone (Dadlez and Kopik, 1972; Garbowska et al., 1978; Kopik, 1998; Zakrzewski, 1976). According to Zakrzewski (1976), terrestrial conditions with intense chemical weathering were present in the investigated region during the hiatus. Because of the hiatus, one might expect an unconformity between these geological units; however, observations carried out in mines (e.g. Szczekaczka⁵ near Częstochowa) have not supported this hypothesis and proved the parallelism between the Aalenian–Early Bajocian sand/sandstones and the older Late Bajocian–Late Bathonian clays (Zakrzewski, 1976).

Aalenian–Early Bajocian sediments are referred to as Kościeliska⁶ beds (Razowska, 2001). These deposits are sands with admixtures of quartz gravels and siderite concretions (Gedl and Kaim, 2012), and are regarded as transgressive shelf deposits—and their age is evidenced by the presence of index ammonites (Feldman-Olszewska, 1997; Mossoczy, 1947). From a hydrogeological viewpoint, the outcrops of the Kościeliska sediments can be regarded as unconfined recharge areas, and, along with the dip direction, the presence of the younger ore-bearing clays serves to confine the aquifer system (Hermański, 1971). Because mining above the confined aquifer meant that there was always potential for mine flooding (Wang and Park, 2003), the groundwater circulation within the

⁵ The mine in Szczekaczka was transformed into a museum in 1980, although it was then closed indefinitely, in 1984.

⁶ Kościeliska is a village in the administrative district of Gmina Radłów, within Olesno County, Opole Voivodeship, in SW Poland. It lies approximately 14 km NE of Olesno town.

underlying Kościeliska sediments was of great interest to the miners. Hermański (1971) summarised the hydrogeological conditions before the rapid exploitation of the ore-bearing clays in the second half of the 20th century, and it had been noticed that the sub-horizontal dip of the Kraków–Silesian Homocline sediments in the NE direction was a key factor determining the original groundwater flow, and the orientation of the groundwater table within the Kościeliska aquifer system (Hermański, 1971; Pich and Pokora, 1982).

3. Materials and methods

3.1. Materials

We used 820 borehole records (“The borehole database - deposits of Middle-Jurassic ore-bearing clays near Częstochowa”) that were part of a greater database that was handed over to the University of Silesia in Katowice, by the “Geological Company of Częstochowa” (Częstochowskie Przedsiębiorstwo Geologiczne). The majority of these records were completed in the early 1950s, and were subsequently digitised and re-interpreted so that every borehole record contained Cartesian coordinates for three surfaces: the terrain, the top of the ore-bearing clays deposits, and their base. The precision of the processed Cartesian coordinates was 1 cm, although information was not provided regarding the uncertainty of the borehole paths. The coordinate reference system used in this study was Pulkovo 1942(58)/Poland zone V (EPSG: 2175).

3.2. Delaunay triangulation

Delaunay triangulation can be regarded as either an algorithm producing a geometrical structure, or as a structure generated by this algorithm. As for the first meaning, we used a modified, *two-dimensional* version of this algorithm. The generation of the *three-dimensional* resulting surface was accomplished in two steps: firstly, the investigated surface was partitioned into 2D triangles, on the basis of the XY coordinates associated with given 3D points (XYZ). To generate the resulting surface, it was sufficient to assign the elevation (Z) to the vertices of the 2D triangles produced. From a mathematical viewpoint, the structure produced is a graph closely related to a Voronoi diagram. The latter consists of polygons (Voronoi regions) with representatives having the property that the distance between every point, p_i , lying inside a polygon, P , and the polygon's representative, say \bar{p} , is smaller than the distance between p_i and any other polygon's representative. The vertices of the Delaunay triangulation are the given 3D points and the edges are straight-line segments that connect every pair of sites having Voronoi regions sharing a common edge (Guibas and Stolfi, 1985). Below, we list four properties of Delaunay triangulation that seemed to be crucial to our approach (De Berg et al., 2000; Rippa, 1990). The relevance of these theorems is due to their capacity of explaining selected characteristics of the resulting distribution, such as the number and quality of triangles, uniqueness and smoothness of the orientation distribution.

3.2.1. Property 1. (Valid for any triangulation)

Let P be a set of n points in the plane, not all collinear, and let k denote the number of points in P that lie on the boundary of the convex hull of P . Then any triangulation of P has $2n - 2 - k$ triangles and $3n - 3 - k$ edges.

3.2.2. Property 2

Delaunay triangulation is uniquely determined if, and only if, P contains no four points on a circle.

3.2.3. Property 3

Any Delaunay triangulation of P maximises the minimum angle over all triangulations of P .

3.2.4. Property 4 (Rippa, 1990).

The Delaunay triangulation of P , is a *minimal roughness* triangulation, for any set of P .

The first property is actually valid for any triangulation, and it provided the approximate number of triangles produced, on the basis of the number of input boreholes. The second property addresses the problem of uniqueness of computing the Delaunay triangulation and, as a consequence, the problem of uniqueness of the 3D orientation vector distributions. The risk of obtaining a non-unique distribution of the triangles' attributes naturally rises if the points lie in more regular grids, but this was not the case in our study (Fig. 1). The next property reveals the advantage of using the Delaunay triangulation in terms of the numerical limitations related to collinear configurations in the three-point problem. The assumption of considering surfaces produced by the Delaunay triangulation as *smooth* surfaces follows from the last property given. It should be noticed that an adequate explanation of this property would require the introduction of advanced mathematical concepts (e.g. Sobolev semi-norm, harmonic functions), which lies beyond the scope of our research.

3.3. Computing orientation

To compute the orientation, we used a standard linear algebra algorithm that assumed removal of collinear or almost-collinear configurations—a restriction based on the relationship between the longest edge of a 3D triangle and the sum of its remaining edges. The lengths of a triangle's edges were calculated and sorted into ascending order. A vector of lengths $[l_1, l_2, l_3]$ was obtained, such that. The coefficient of collinearity was obtained using a simple division $\frac{l_1}{l_1 + l_2}$ and was assumed to lie in the interval $[0.5, 1]$. For triangles of equilateral shape the values of this coefficient were near 0.5, with more collinear configurations, on the other hand, approaching 1 (Michalak, 2018).

In the referenced example, a uniformly oriented surface, whose dip angle was approximately symmetrically distributed about the mean, was investigated. Setting the collinearity threshold to 0.975 allowed the *numerical outliers* to be identified and removed.

The procedure for setting the threshold was experimental however, therefore, having no systematic approach, there was always a certain degree of arbitrariness in setting the upper collinearity limit. In this study, we decided to use a more restrictive approach, by setting the threshold to 0.95: this was not to say that this number can be assumed to be proven as better by experience, however, we had no evidence, at the time, that triangles having the collinearity substantially smaller than as suggested by Michalak (2018) could also produce outliers. On the other hand, choosing a greater collinearity restriction (to have more equilateral triangles) may have led to the loss of some valuable information.

3.4. Mean vector approach and inertia moment analysis

Having obtained the population of orientation measurements, we conducted statistical analysis using the mean vector approach and inertia moment analysis. Both methods require computing the Cartesian coordinates of the unit normal vectors, and the mean vector approach assumes that the mean orientation is the vector arithmetic mean of the unit normal vectors.

Inertia moment analysis is often expected to compute the maximum density of vectors (García-Sellés et al., 2011) or, equally, determine their dominant orientation (Benito-Calvo et al., 2009). It requires the eigenvectors and eigenvalues of an *orientation matrix* to be computed, as shown in (1).

$$A = \frac{1}{n} \begin{bmatrix} \sum l_i^2 & \sum l_i m_i & \sum l_i n_i \\ \sum m_i l_i & \sum m_i^2 & \sum m_i n_i \\ \sum n_i l_i & \sum n_i m_i & \sum n_i^2 \end{bmatrix}, \quad (1)$$

In (1), l_i , m_i , n_i correspond to the Cartesian coordinates X Y Z of the normalised normal vectors, respectively, and n denotes the number of observations.

The eigenvectors, v_1, v_2, v_3 , and corresponding eigenvalues, $\lambda_1, \lambda_2, \lambda_3$, were computed. The eigenvalues can be normalised using $S_i = \lambda_i/n$, such that $S_1 + S_2 + S_3 = 1$. The ratio $C = \ln(S_1/S_3)$ assesses the strength of the preferred orientation. The second coefficient, $K = \ln(S_1/S_2)/\ln(S_2/S_3)$, appraises the shape of the observations' distribution, which can be in either the form of a cluster, which indicates low variability of orientation ($K > 1$), or a girdle ($0 < K < 1$), which can arise when more heterogeneous surfaces are examined (Woodcock, 1977).

3.5. Clustering algorithms

As a part of the exploratory analysis, we performed cluster analysis on the space of the 3D unit normal vectors. Cluster analysis algorithms always require the user to arbitrarily determine certain initial conditions, which can include the number of clusters (*k-means*, *k-medoids*), the dissimilarity measure used (*k-medoids*, hierarchical clustering), the method for merging clusters (hierarchical clustering), or the range of a cluster (density-based spatial clustering of applications with noise, better known as DBSCAN). As for the selection of an adequate number of clusters, many methods have been introduced to determine it automatically, such as the *Silhouette* method (Rousseeuw, 1987), used by Chen et al. (2016), the Davies Bouldin index (Davies and Bouldin, 1979), applied by Carmichael and Ailleres (2016), or the Gap statistic (Hastie et al., 2009). Because the automatic methods provide different suggestions on the optimum number of clusters, there arises a problem of the proper justification of using them. We believed it was too early to commit to the idea that a particular automatic method of determining the reasonable number of clusters should be followed in our case. In this study, we wanted to incorporate an approach based on *geological knowledge*, as recommended by Jimenez-Rodriguez and Sitar (2006), and by Marcotte and Henry (2002).

3.6. We considered two approaches

1. A minimalistic approach. This approach is based on the rough assumption that there is only one preferred orientation, which can be disturbed by only small-scale changes (such as faults or erosive forms). From this, two clusters should be expected—one associated with the preferred orientation and another related to the noise.

2. A multi-noise approach. This approach assumes that the noise can be partitioned into more than one cluster, and follows from the simple observation that NE–SW faults can have opposite (NW and SE) dip directions. Along with two additional clusters related to faults, we expected that there could also be a fourth source, of an unknown nature (related to measurement errors, for example).

These two assumptions regarding the number of clusters and the preference for Euclidean distance suggested that partitional algorithms (*k-means*, *k-medoids*), as well as hierarchical algorithms with Euclidean distance, should be considered in the first place.

In the *k-means* algorithm, the very first centres to which observations are assigned, are selected randomly. From this it follows that the results obtained by using the *k-means* algorithm can differ at every turn. To solve this problem, some test configurations have been suggested, from which the best—with respect to the goal of minimising within-cluster variability—was selected (Hastie et al., 2009). In the next iterations, the vector arithmetic mean was used as a centre to which other observations were assigned, and the squared Euclidean distance was used as a distance function. The property of the arithmetic mean—that it minimised the squared Euclidean distances to other observations—provides that the within-cluster variance is reducing in each iteration. Minimising the within-cluster variability in the *k-medoids* algorithm requires a distance function to be selected. In the *k-medoids* algorithm, the observation

that minimises the sum of distances (according to the selected distance function) is assumed to be the centre to which other observations are assigned in each iteration. The k -medoids algorithm must check all pairwise distances and is, therefore, more computationally expensive.

We included hierarchical clustering in the analysis as well. These methods are different in nature, as they require the dissimilarity distance function, and the linkage method that determined the manner in which the clusters were merged in each iteration, to be specified (Hastie et al., 2009; James et al., 2013). In this study, we used Euclidean distance with complete linkage and Ward linkage methods (Ward, 1963). Complete linkage produces compact clusters, whereas the Ward method aims to minimise within-cluster variability and is, therefore, more similar to the k -means and k -medoids algorithms. We selected the agglomerative (bottom-up) version of these algorithms, which meant that initially the observations were regarded as individual clusters. In the first step of an agglomerative algorithm, the two least dissimilar (according to the selected distance function) clusters are merged into one cluster at a particular height. This merging height can also be interpreted as an intergroup dissimilarity between these two merged clusters. This procedure is applied at each of the remaining steps, producing one less cluster at the succeeding higher level (Hastie et al., 2009). This procedure can be summarised using a binary tree referred to as a *dendrogram*, and a particular number of clusters can be obtained by cutting the dendrogram horizontally at a particular height. Hierarchical clustering does not provide the elegant possibility of the centres of the clusters to be obtained, as in k -means or k -medoids algorithms, so we therefore applied the mean vector approach and inertia moment analysis again, when using hierarchical clustering (Table 2, Table 3).

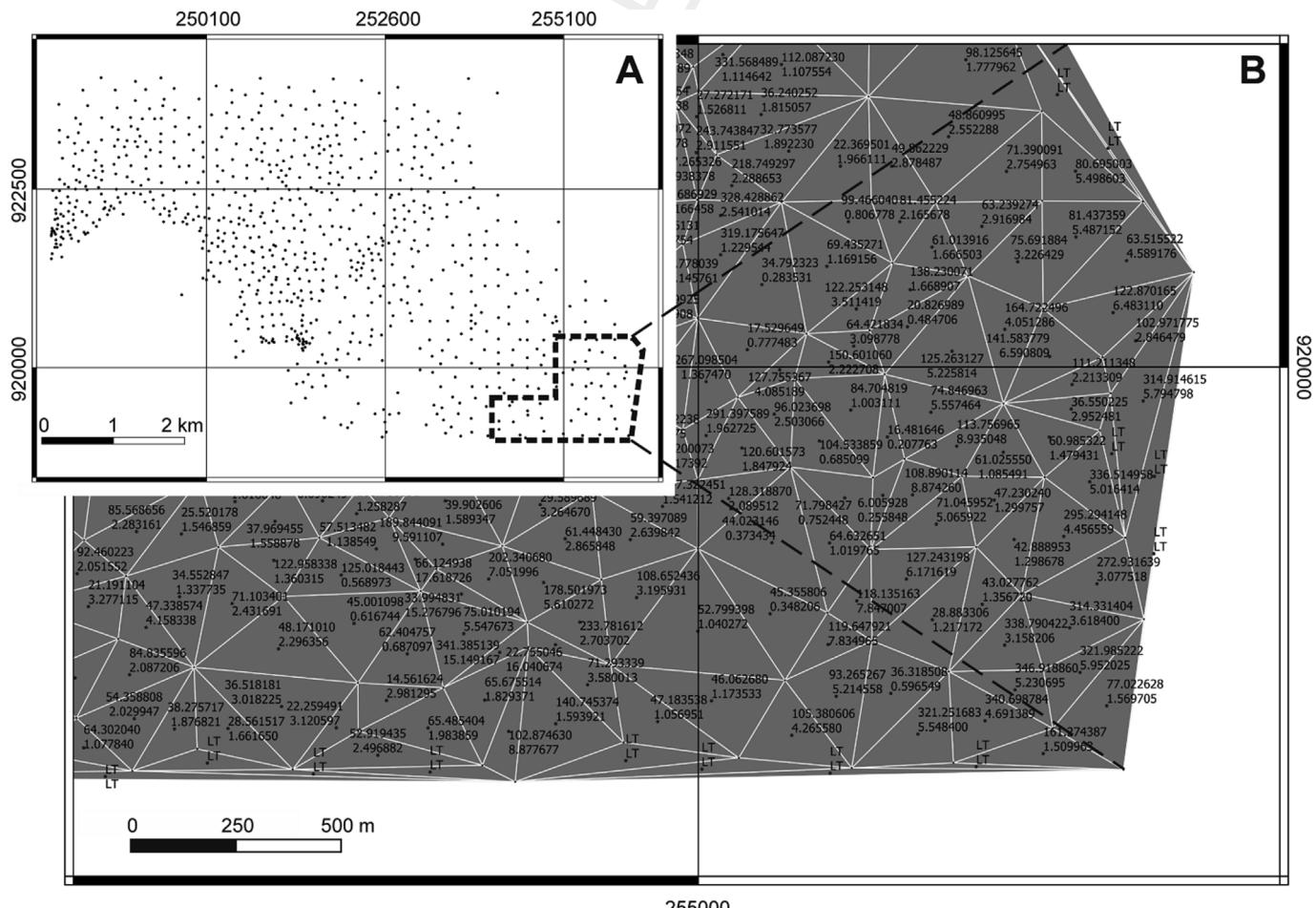


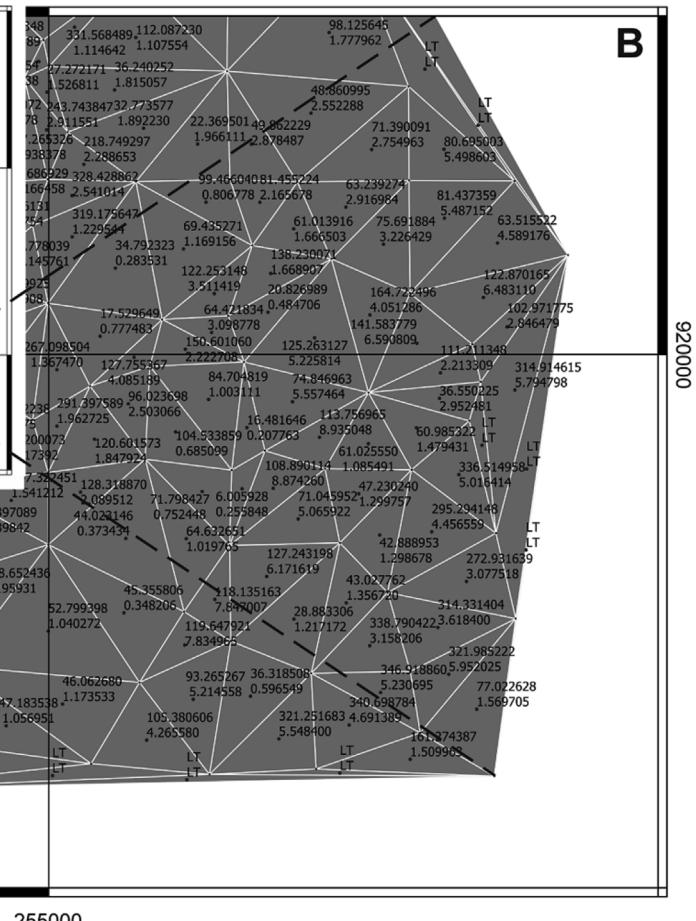
Fig. 4. Delaunay triangulation for the bottom of the Middle–Jurassic ore-bearing clay deposits. A total of 1618 triangles were generated. Labels indicate the dip direction and angle for every triangle. Long and thin triangles (DOC exceeding 0.95) are denoted by LT/LT.

3.7. Software background

From a technical point of view, to compute the Delaunay triangles, we used the CGAL library in Microsoft Visual Studio. CGAL is a software project that provides access to efficient and reliable geometric algorithms, in the form of a C++ library. The efficiency of CGAL follows from the speed of the C++ language, while the reliability of the geometric results is due to the *Exact Geometric Computation* principle, which is expected to guarantee the correctness of a produced geometric structure (Hert and Seel, 2018). We supplied 820, 3D borehole points, representing the investigated surface, as the input data. The output file contained fifteen columns that were used to remove collinear configurations and to perform stereographic projection in the Dips (Rocscience, 2017) software: X1, Y1, Z1, X2, Y2, Z2, X3, Y3, Z3, X_C, Y_C, X_N, Y_N, Z_N, Dip_ang, Dip_dir, DOC, Area. The first nine columns are the coordinates of the Delaunay triangles (X1, Y1, Z1, X2, Y2, Z2, X3, Y3, Z3). To allow visualization of the measurements, we also computed the 2D centres of these triangles (X_C, Y_C). The next three columns are the direction cosines of the computed unit observations (X_N, Y_N, Z_N), which can also be expressed as the dip angle and dip direction located in the following two columns, respectively (Dip_ang, Dip_dir). The next two columns refer to the reliability of the results—the coefficient of collinearity and the sizes of the Delaunay triangles (DOC, Area).

4. Results

We used the output data to plot the triangulated network and assigned orientation parameters to the geometric centres of the triangles



(Fig. 4). The long and thin triangles that contributed to undesirable noise among the data were denoted as LT/LT.

Fig. 4 suggests that a significant number of collinear triangles can lie on the boundary of the convex hull of the investigated points. As these collinear configurations did not provide reliable results, we deleted them prior to further processing. We then performed stereographic projections of the unit normal vectors to the Delaunay triangles, using Dips software (Fig. 5). The greatest concentration of dip vectors was observed for the dip direction lying in the 45–60° interval.

Table 1 shows the main results of computing the orientation and indicates that the computed orientation did not coincide with the maximum density of vectors, as presented in Fig. 5. Therefore, cluster analysis was conducted with respect to the motivations behind the *minimalistic* and *multi-noise* approaches described above. For simplicity, in both approaches the cluster containing the observations related to the preferred orientation will be referred to as the *expected cluster*. Similarly, the cluster (*minimalistic approach*), or clusters (*multi-noise approach*), groupings observed to be substantially dissimilar to the preferred orientation will be referred to as a *noise cluster* (*minimalistic approach*), or *noise clusters* (*multi-noise approach*).

We used the following clustering algorithms: *k*-means (with its inherent squared Euclidean distance), *k*-medoids (with Euclidean distance), and hierarchical clustering (with combinations of Euclidean distance along with complete and Ward linkage methods). To better explain the differences between the different clustering results, we plotted stereonet in two modes—pole vector (Figs. 6 and 8) and dip vector (Figs. 7 and 9). The observations associated with the *expected* clusters were always green coloured, while those related to the *noise* were either red (*minimalistic approach*) or red, blue, and yellow (*multi-noise approach*).

When considering two clusters, the main difference between the clusters produced by the *k*-means and *k*-medoids algorithms could be attributed to the range of the *noise cluster* (Fig. 6A, B, 7A, 7B). The *k*-means algorithm (Figs. 6A and 7A) assigned planes to this cluster within the range of dip directions between of 255° and 2°. The orientation associated with the *expected cluster* was 65.15/1.18, which still did not

coincide with the maximum density of vectors, as shown on the stereonet (Fig. 5). The range of the *noise cluster* when using the *k*-medoids algorithm with the Euclidean distance was greater, between 240° and 22°. The 57.51/1.14 observation was selected as being representative of the *expected cluster*, and this was closer to the results presented on the stereonet.

The results obtained using hierarchical clustering were more varied. The Euclidean distance and Ward linkage method produced a relatively large *noise cluster* (Figs. 6D and 7D), with 420 observations (Table 2). The boundary between the two considered clusters was located near the 200° azimuth, where few observations occurred (Fig. 7D). The expected orientations for inertia moment analysis and mean vector approach were 81.42/1.88 and 81.48/1.91, respectively. The *noise cluster* when Euclidean distance with complete linkage was used, turned out to be more compact, with expected orientations of 55.83/1.01 and 56.50/1.02.

When considering that four clusters, and the algorithms employed, with the exception of the *k*-means algorithm (70.02/1.05), it was noted that they supplied very similar dip direction orientations, varying between 53.47 and 55.06. Among the partitional algorithms, the *k*-medoids algorithm seemed to have delivered better extraction of the expected orientation than the *k*-means algorithm, in which even the observation of dip direction (190°), and dip angle greater than 30°, were assigned to the *expected cluster*. This was not the case in the *k*-medoids output, in which the *expected cluster* formed a narrower girdle, with a decreasing number of moderately dipping planes indicating deviations from the typical dip direction.

Moreover, the Ward linkage provided better extraction of the typical dip direction on the stereonets than the complete linkage, from among the hierarchical algorithms. Compared to the *k*-medoids algorithm, the Ward linkage supplied even better visualization of the expected orientation than *k*-medoids, as very few observations of the dip direction lying in the interval [180, 270] were observed in the Ward's clustering output.

Analysis of a dendrogram (Fig. 10) can be also beneficial for explaining the essential differences between the clustering approaches.

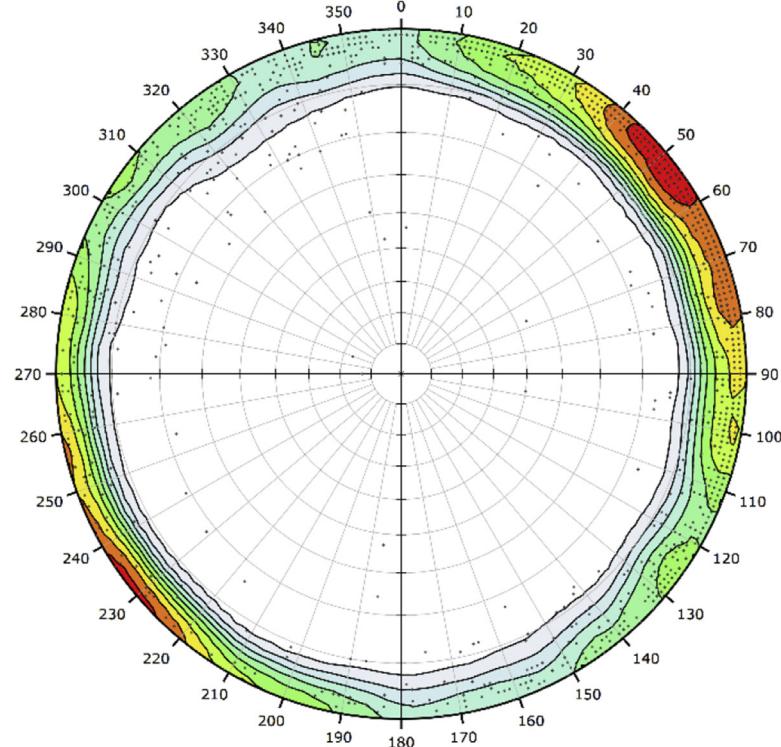


Fig. 5. Density contour plot related to the measurements of the ore-bearing clay deposits bases: dip vector mode. The greatest density of vectors is to be observed for the dip direction lying in the interval 45–60°. The greater density of vectors lying on the opposite side of the stereonet remains unexplained.

Symbol	Feature
Dip Vectors	
Color	Density Concentrations
	0.00 - 1.20
	1.20 - 2.40
	2.40 - 3.60
	3.60 - 4.80
	4.80 - 6.00
	6.00 - 7.20
	7.20 - 8.40
	8.40 - 9.60
	9.60 - 10.80
	10.80 - 12.00
Contour Data	Dip Vectors
Maximum Density	11.70%
Contour Distribution	Fisher
Counting Circle Size	1.0%
Plot Mode	Dip Vectors
Vector Count	1564 (1564 Entries)
Hemisphere	Lower
Projection	Equal Angle

Table 1

Results from using the mean vector and inertia moment analysis approach.

<i>d</i> (collinearity)	No restriction	<i>d</i> < 0.95		
n	1617		1564	
Delaunay triangles	Area [m ²]		Area [m ²]	
	Min	Max	Min	Max
	59.82	606712	59.82	606712
Orientation matrix	Dip direction	Dip angle	Eigenvalue	Mean
	S ₁	34.22	0.98617	19022.86
	S ₂	143.29	0.00816	0.99009
	S ₃	233.29	0.00567	0.00628
Woodcock C	5.16		5.61	0.00363
Woodcock K	13.14		9.23	
Fisher mean	Dip direction	Dip angle	Dip direction	Dip angle
	32.36	1.06	29.22	1.12

For instance, the complete linkage produced more asymmetric dendograms (Fig. 10) in which many observations and smaller clusters associated with the expected cluster were merged at a very low height. The dendograms also exhibited a possible advantage from considering two clusters only, as the expected and noise clusters could be compared more easily in terms of their within-dissimilarity. When using two clusters, the height at which the noise cluster was generated was greater than that at which the expected cluster was produced. As the dendrogram showed that the dissimilarity within observations grouped in one cluster did not exceed the height at which this cluster was produced, the variability within the noise cluster was greater than that within the expected cluster.

After performing clustering, Woodcock's *C* (strength of the preferred orientation) and *K* (homogeneity of the orientation) values increased (Tables 1 and 2), although, even before starting the clustering algorithms, these values exceeded the typical referenced values (Bertran et al., 1997; Marich et al., 2005; Zuchiewicz, 1997). In our study, *C* and *K* minima were 5.16 and 9.23, and the maxima were 8.92 and 16.11, respectively. In the referenced examples, the *C* and *K* minima were 0.61 and 0.02, while the maxima were 3.93 and 6.23, respectively (See Discussion).

5. Discussion

In this paper, we have presented a new approach for determining the orientation of a sub-horizontal contact that separates two geological units, employing the three-point approach and Delaunay triangulation. Compared to the best-fitting plane approaches (Jones et al., 2016), or those based on interpolation (Mallet, 1989, 1992) or co-kriging (Lajaunie et al., 1997; Lindsay et al., 2012), the proposed combination has been shown to be capable of linking orientation measurements with points lying within the interiors of the produced Delaunay triangles.

From the viewpoint of topology and graph theory, the partition obtained using Delaunay triangulation establishes a *many-to-one* function, say *f*, between an infinite set of 3D points lying on the *smooth* surfaces (in the sense of Rippa), and a finite set of output 3D vectors. To expand on this, it should be first noted that 3D points that do not lie in the *interiors* of the produced Delaunay triangles cannot be formally associated with any 3D vector. The set of these 3D points includes the input 3D points, or points corresponding to the edges of Delaunay triangles that are shared by more than one triangle. Therefore, this *many-to-one* relationship can be expressed in a *quasi-formal* way, as follows: for every 3D output vector, *v*, there exists an interior of exactly one Delaunay triangle, say *t_i*, such that *f(t_i) = v*. The proposed approach allowed further application of stereonets and cluster analysis.

We have indicated the limitations of our procedure related to collinear configurations which did not provide reliable results—which led us to remove triangles with collinearity coefficients exceeding 0.95 from further computation. The removal of these configuration was not sufficient, however, to obtain results consistent with the density contour plot on the stereonet (Fig. 5). Neither the mean vector approach nor the inertia moment analysis were sufficient to precisely determine the

maximum density of vectors alone. Therefore, cluster analysis was conducted, to supply a better approximation of the average orientation, and to investigate potential patterns in the data set. We indicated that, compared to the angular distance, the Euclidean distance could be a better candidate for separating planes dipping in the opposite direction. More generally, in our case, the angular distance seemed to be a better function for grouping planes similar from the *geomorphological* point of view, while the Euclidean distance was better from the *genetic* point of view. Having selected the Euclidean distance (or squared Euclidean distance in *k*-means), we followed the *minimalistic* and *multi-noise* approaches, assuming that there could be two or four clusters, respectively. The first approach was assumed to produce two clusters, one with the preferred orientation and the second that could be regarded as noise.

In the first case, the algorithms employed were incapable of totally following the assumption that all observations in the *expected cluster* were related to the expected orientation. This was because the *expected cluster* also included observations that dipped in the opposite direction, as can be seen from the Euclidean distance and Ward's linkage method results. This combination did not supply accurate definition of the expected orientation, which could be explained by the fact that the Ward linkage generally produced more balanced partitions (Fig. 10), contributing to a relatively large *noise cluster*.

Except for the Ward's case, the observations assigned to the *noise cluster* did not significantly disturb the calculations of the expected orientation, due to their low dip angle. Therefore, the *expected cluster* in the minimalistic approach should be described as one that included observations that were not significantly different from the expected orientation and that did not vitally affect our calculations. As for the *noise cluster*, a common group of observations for all partial plots can be distinguished, on Fig. 7. This group included planes dipping moderately or even steeply to the NW, perpendicular to the average direction. The exact origin of these planes remains unknown; however, one possible explanation is that they could be associated with NE-SW faults (Bardziński et al., 1986), one of which can be seen in the NW part of the investigated region (Fig. 1).

When four clusters were considered, the average orientation was even better visualized (Figs. 8 and 9). In all cases, the *expected cluster* still contained the greatest amount of observations. This concentration of observations proved that the noise could indeed be associated with small-scale deviations. As for the motivation behind the multi-modal approach, analysis of the dendrogram (Fig. 10) could serve as an a posteriori justification for considering two or four clusters, as the candidates for natural clusters were groups that merged at high values, relative to the merger values of the clusters contained within them (Hastie et al., 2009).

Regarding the Woodcock's distribution coefficients *C* and *K*, one could argue that the values were higher in our study because the investigated surface is relatively smooth and homogenous. These results could be also explained using Rippa's theorem on *minimal roughness* properties of surfaces generated by Delaunay triangulation. However, the smaller values of the Woodcock coefficients related to compass-clinometer

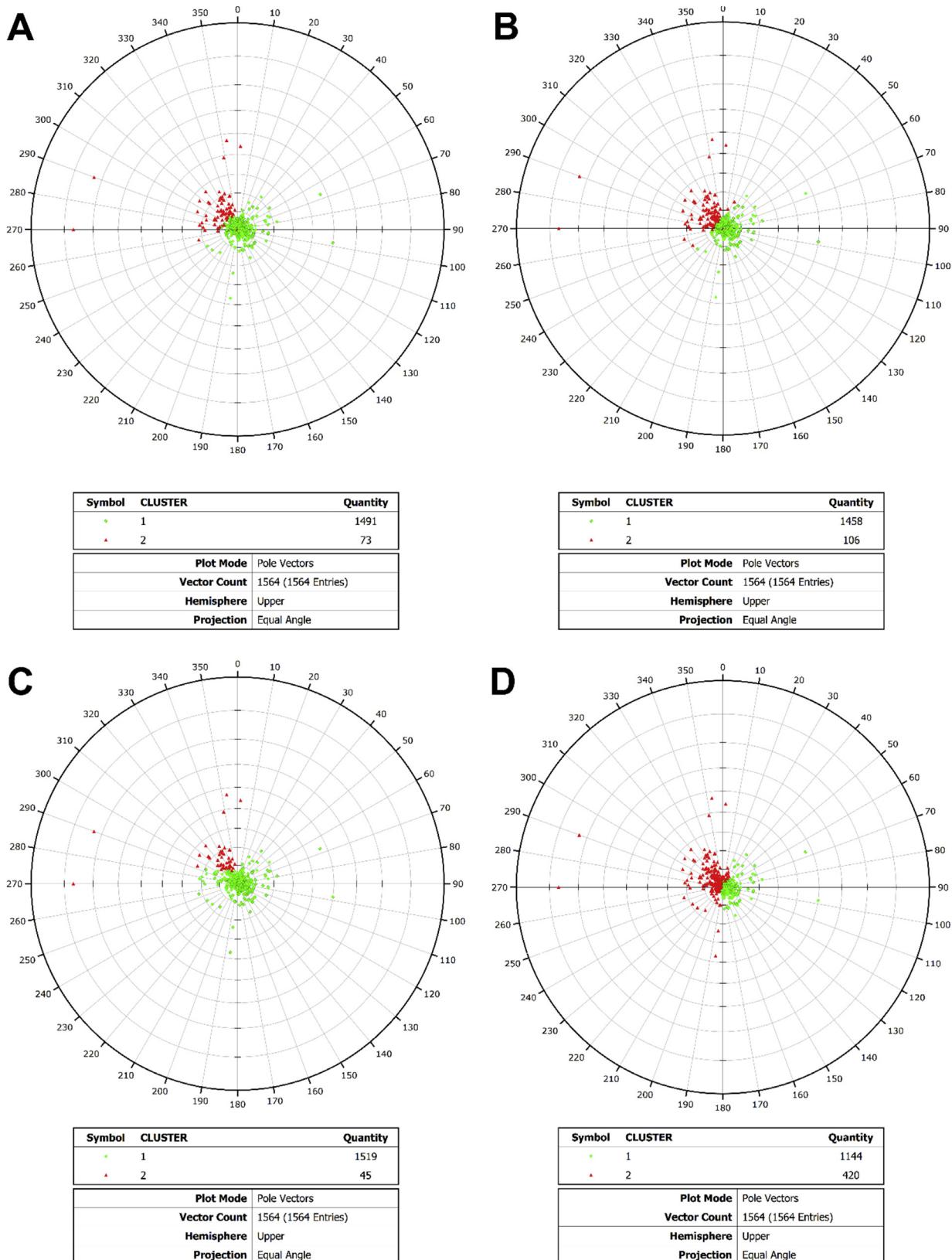


Fig. 6. Stereographic projection of the pole vectors. Two clusters were obtained using different clustering methods. The expected cluster is green and the noise cluster is red: (A) k-means algorithm; (B) k-medoids algorithm with Euclidean distance; (C) hierarchical clustering with Euclidean distance and the complete linkage method; (D) hierarchical clustering with Euclidean distance and the Ward linkage method. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

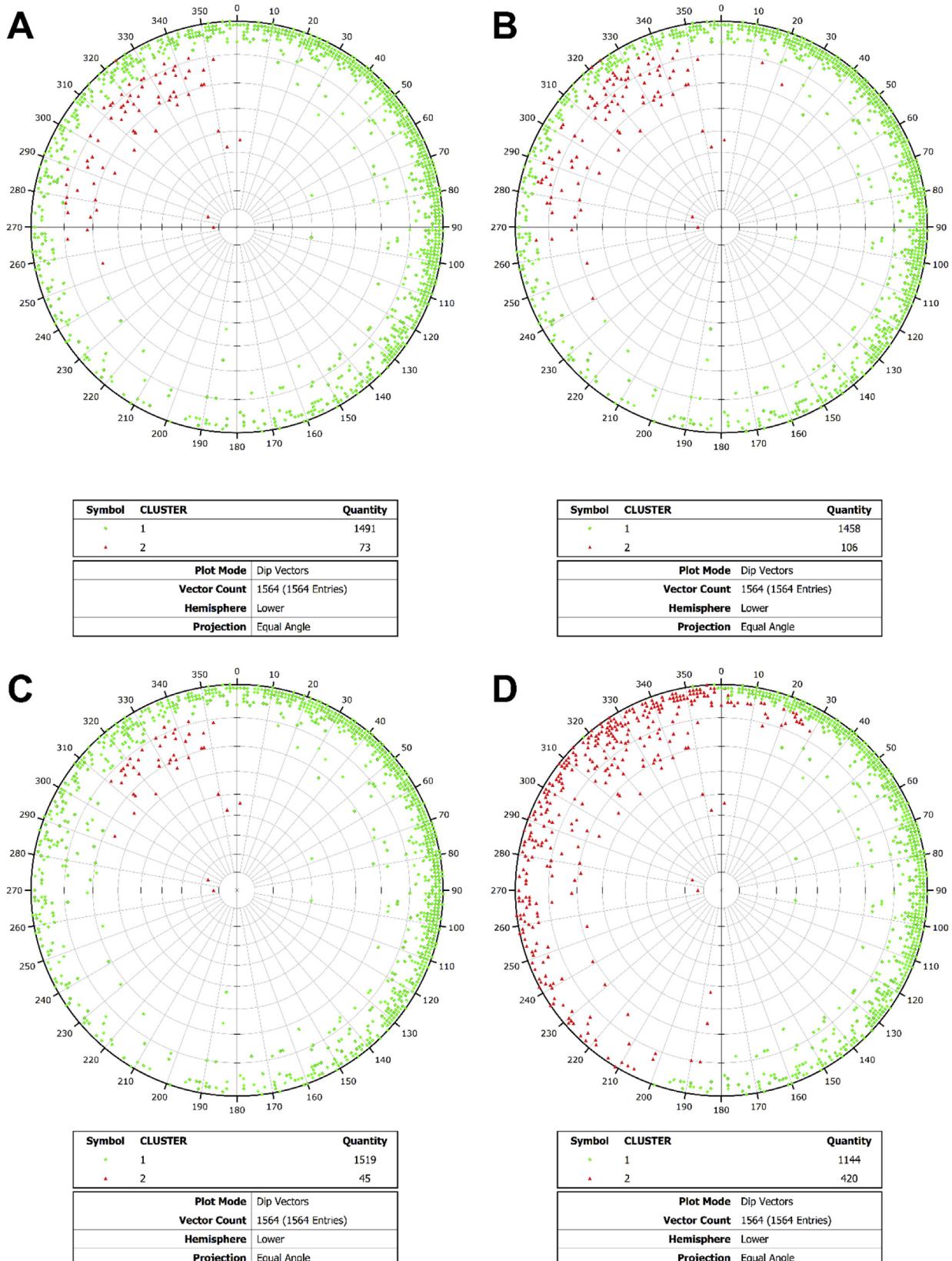


Fig. 7. Stereographic projection of the dip vectors. Two clusters were obtained using different clustering methods. The expected cluster is green and the noise cluster is red: (A) k-means algorithm; (B) k-medoids algorithm with Euclidean distance; (C) hierarchical clustering with Euclidean distance and the complete linkage method; (D) hierarchical clustering with Euclidean distance and the Ward linkage method. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

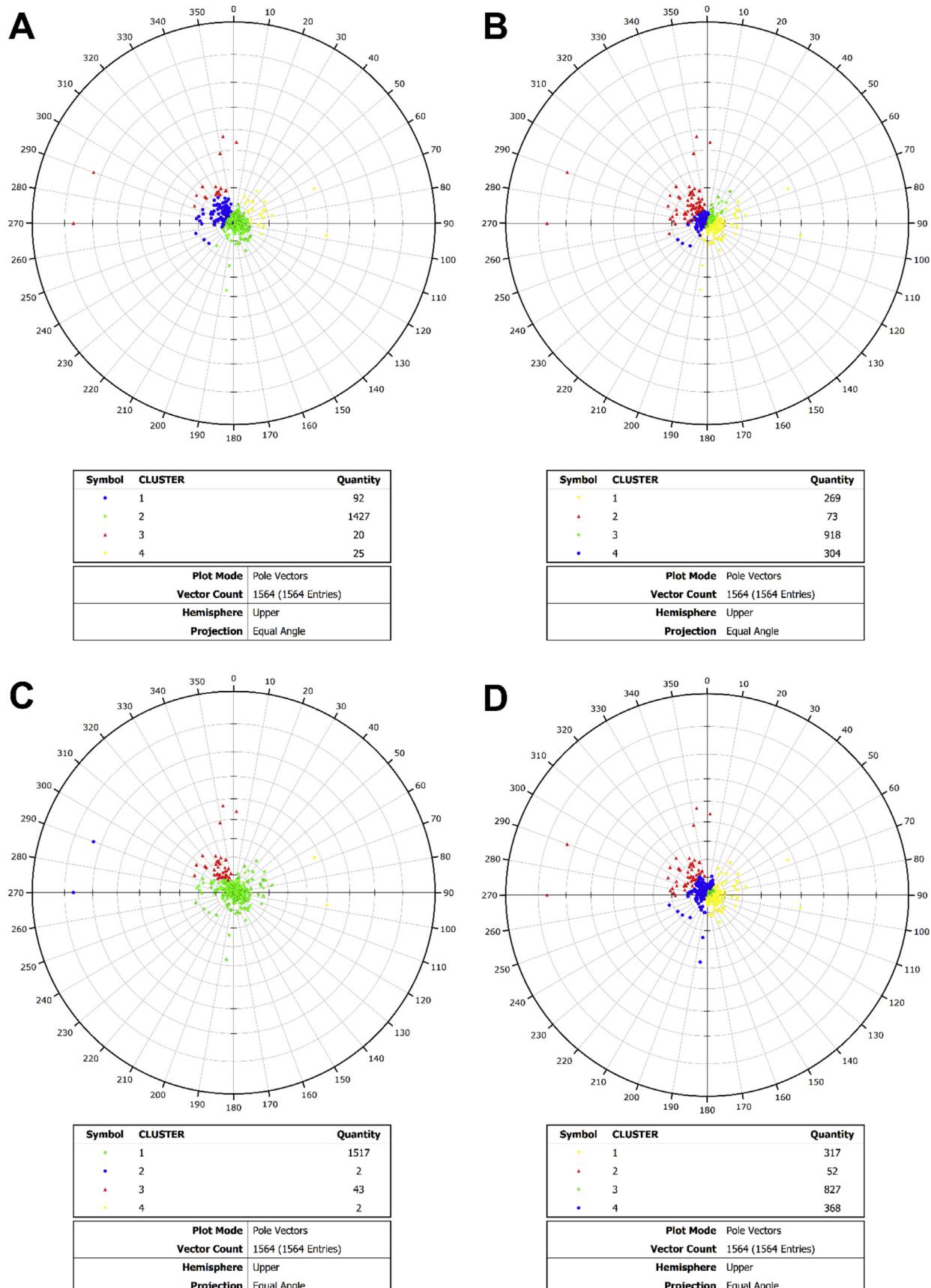


Fig. 8. Stereographic projection of the pole vectors. Four clusters were obtained using different clustering methods. The expected cluster is green and the noise clusters are red, blue and yellow: (A) k-means algorithm; (B) k-medoids algorithm with Euclidean distance; (C) hierarchical clustering with Euclidean distance and the complete linkage method; (D) hierarchical clustering with Euclidean distance and the Ward linkage method.

67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131

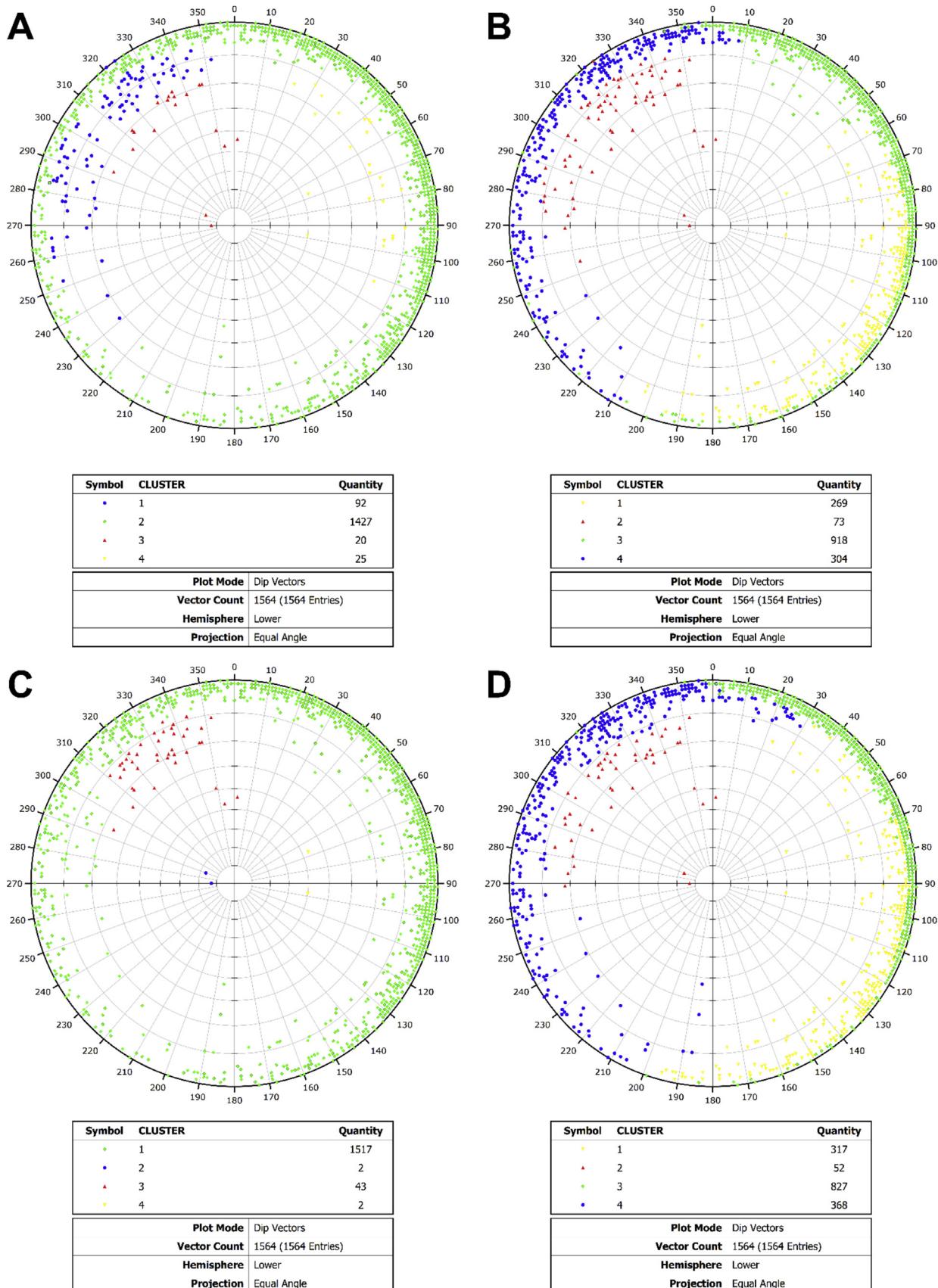


Fig. 9. Stereographic projection of the dip vectors. Four clusters were obtained using different clustering methods. The expected cluster is green and the noise clusters are red, blue and yellow: (A) k-means algorithm; (B) k-medoids algorithm with Euclidean distance; (C) hierarchical clustering with Euclidean distance and the complete linkage method; (D) hierarchical clustering with Euclidean distance and the Ward linkage method.

Table 2

Results from computing orientation using different clustering algorithms (k = number of clusters; n_e , n_n = number of observations in the expected and noise clusters, respectively).

K = 2, d < 0.95						
K-means algorithm	n_e 1491	n_n 73	Dip direction 65.15	Dip angle 1.18		
K-medoids algorithm (Euclidean distance)	n_e 1458	n_n 106	Dip direction 57.51	Dip angle 1.14		
Hierarchical clustering (Euclidean distance, complete linkage method)	n_e 1519	n_n 45	Orientation matrix S_1 55.83 S_2 273.46 S_3 183.45	Dip direction 6.14	Eigenvalue 1.01 89.20 89.38	
			Woodcock C Woodcock K Fisher mean	Dip angle 10.51 56.50	0.99418 0.00367 0.00215	
Hierarchical clustering (Euclidean distance, Ward linkage method)	n_e 1144	n_n 420	Orientation matrix S_1 81.42 S_2 263.90 S_3 173.90	Dip direction 6.53	Dip angle 1.88 88.13 89.92	Eigenvalue 0.99616 0.00239 0.00145
			Woodcock C Woodcock K Fisher mean	Dip direction 12.07 81.48	Dip angle 1.91	

Table 3

Results from computing orientation using different clustering algorithms (k = number of clusters; n_e = number of observations in the expected cluster; n_{n1} , n_{n2} , n_{n3} = numbers of observations in the clusters associated with noise).

K = 4, d < 0.95						
K-means algorithm	n_e 1427	$n_{n1} + n_{n2} + n_{n3}$ 92 + 20 + 25 = 137	Dip direction 70.02	Dip angle 1.05		
K-medoids algorithm (Euclidean distance)	n_e 918	$n_{n1} + n_{n2} + n_{n3}$ 269 + 73 + 304 = 646	Dip direction 54.67	Dip angle 1.23		
Hierarchical clustering (Euclidean distance, complete linkage method)	n_e 1517	$n_{n1} + n_{n2} + n_{n3}$ 2 + 43 + 2 = 47	Orientation matrix S_1 54.86 S_2 279.62 S_3 189.62	Dip direction 6.17	Dip angle 0.98 89.31 89.31	Eigenvalue 0.99489 0.00303 0.00208
			Woodcock C Woodcock K Fisher mean	Dip direction 15.27 55.06	Dip angle 0.97	
Hierarchical clustering (Euclidean distance, Ward linkage method)	n_e 827	$n_{n1} + n_{n2} + n_{n3}$ 317 + 52 + 368 = 737	Orientation matrix S_1 53.47 S_2 204.60 S_3 294.61	Dip direction 8.92	Dip angle 1.35 88.82 89.35	Eigenvalue 0.99964 0.00023 0.00013
			Woodcock C Woodcock K Fisher mean	Dip direction 16.11 53.47	Dip angle 1.35	

measurements could also be explained by the difficulty in picking up the trend at the scale of the outcrops (Pakyuz-Charrier et al., 2018). Regardless of which was the most adequate explanation, we suggest that

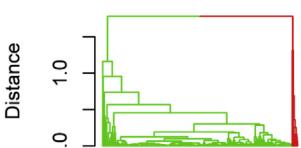
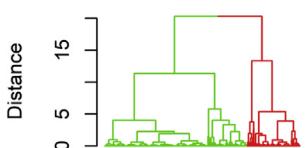
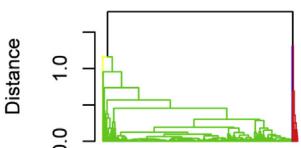
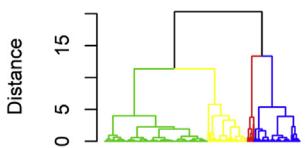
Euclidean distance, complete linkage**Euclidean distance, Ward linkage****Euclidean distance, complete linkage****Euclidean distance, Ward linkage**

Fig. 10. Dendograms for different hierarchical clustering approaches. When considering two clusters, the expected cluster is green and the noise cluster is red. If four cluster are considered, the expected cluster is green and the noise clusters are red, blue and yellow.

comparing Woodcock's C and K values for two different surfaces should only be done if the same procedure was applied for collecting the surface orientation data.

6. Conclusions

- In this study, we combined the three-point approach with Delaunay triangulation to determine the average orientation of a sub-horizontal contact within the Kraków-Silesian Homocline in Poland. The significance of this study can be referred to by the fact that dip directions for sub-horizontal surfaces are difficult to measure using standard compass-clinometer approaches. Moreover, the regional trend was a key factor in determining groundwater circulation before the rapid exploitation of ore-bearing clays in the second half of the 20th century.
- One main advantage of this approach is the possibility of using the exploratory phase of the orientation data analysis (Fisher et al., 1993). Moreover, one property of the Delaunay triangulation method is that it guarantees sampling the orientation data from the minimal roughness surface.
- The mean vector approach does not provide accurate results if the input data set contains “noisy” observations, such as faults or erosive forms. In such cases, the inertia moment analysis will not compute the maximum density of vectors very well.
- Among the partitional algorithms, the k -means algorithm should not be recommended for obtaining reproducible results, due to the inherently random step at the very beginning of the algorithm. It should be also noticed that in the k -medoids algorithm, the clusters' centres are not calculated—they are only selected from the input data set.
- Considering four clusters allowed the dip direction to be constrained within an interval smaller than 2° (53.47° – 54.67°), and the results obtained were more consistent with the initial density contour plot (Fig. 5). The dip angle was varying from 0.97° to 1.35° . The advantage of using only two clusters can be that it allows a concise comparison between the expected and noise clusters.
- The minor hiatus between the separated geological units and the presence of abrupt changes of orientation made it difficult to regard the boundary surface as a purely stratigraphic contact. In fact, it is

partly stratigraphic (observations in mines), partly unconformable (hiatus) and partly faulted (NE-SW faults). The relative concentration of the observations in the expected clusters indicated, however, that the areas that deviated from the regional pattern were relatively small.

- When considering hierarchical clustering algorithms, using Euclidean distance and the Ward linkage method, with four clusters, allowed the best visualization of the expected orientation to be generated. As hierarchical clustering did not select the clusters' centres, we computed the orientation using inertia moment analysis and the Fisher mean.
- The denoising aspect of cluster analysis in this study was also reflected in the increased Woodcock's C and K values (hierarchical clustering)—which were, in fact, relatively large, even before beginning clustering. Therefore, comparing C and K for two different surfaces would only be valid if the same procedure had been applied for collecting the orientation data for both surfaces.
- Application of cluster analysis allows “noisy” observations to be removed, provided that the approximate expected orientation is known. One disadvantage of the proposed approach is that the results of clustering cannot directly supply statistical information for each orientation cluster (compare Zhan et al., 2017b). It is suggested that mixture models, with probability density functions assigned to the clusters, are used to fill this gap.
- The proposed approach could be expanded in another direction by investigating whether the clusters produced indicated spatial patterns. This could be beneficial in delineating areas exhibiting different orientation patterns.

Declarations of interest

None.

Computer code availability

Name of code: OCT. **Developer:** Michał Michalak. **Contact address:** Department of Applied Geology, Faculty of Earth Sciences, University of Silesia in Katowice, Poland. **Telephone number:** +48 736-944-132. E-mail: mimichalak@us.edu.pl. **Year first available:** 2018. **Hardware required:** Celeron CPU or better. **Software required:** Microsoft Visual Studio (2015 or 2017), CGAL library (ver. 4.8). **Program language:** C++. **Program size:** 600 KB. **How to access the source code:** Available at: <https://github.com/michalmichalak997/OCT>.

Data availability

Original and unprocessed borehole data related to this article can be found and ordered at <http://otworywiertnicze.pgi.gov.pl/>, a borehole repository of the Polish Geological Institute – National Research Institute.

Acknowledgements

We are grateful to Ewa Kurowska for sharing the digital version of the borehole records, and to the Museum of Częstochowa (Archive of the History Department) for sharing the geological documentation. The study was funded by the Faculty of Earth Sciences, University of Silesia in Katowice from a subsidy from the Polish Ministry of Science and Higher Education for research and development works. We are also grateful to two anonymous reviewers who helped us achieve a more systematic approach to our reporting of this research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cageo.2019.104322>.

References

- Assali, P., Grussenmeyer, P., Villemin, T., Pollet, N., Viguer, F., 2014. Surveying and modeling of rock discontinuities by terrestrial laser scanning and photogrammetry: semi-automatic approaches for linear outcrop inspection. *J. Struct. Geol.* 66, 102–114. <https://doi.org/10.1016/j.jsg.2014.05.014>.
- Bardziński, W., Lewandowski, J., Więckowski, R., Zieliński, T., 1986. Objasnienia Do Szczegółowej Mapy Geologicznej Polski W Skali 1:50000, Arkusz Częstochowa. Wydawnictwa Geologiczne, Warszawa, 67pp. [in Polish].
- Benito-Calvo, A., Martínez-Moreno, J., Jordá Pardo, J.F., de la Torre, I., Torcal, R.M., 2009. Sedimentological and archaeological fabrics in palaeolithic levels of the south-eastern pyrenees: cova gran and roca dels bous sites (Ileida, Spain). *J. Archaeol. Sci.* 36, 2566–2577. <https://doi.org/10.1016/j.jas.2009.07.012>.
- Bertran, P., Hetu, B., Texier, J.-P., Steijn, H., 1997. Fabric characteristics of subaerial slope deposits. *Sedimentology* 44, 1–16. <https://doi.org/10.1111/j.1365-3091.1997.tb00421.x>.
- Calcagno, P., Chilès, J.P., Courrioux, G., Guillen, A., 2008. Geological modelling from field data and geological knowledge: Part I. Modelling method coupling 3D potential-field interpolation and geological rules. *Phys. Earth Planet. Inter.* 171 (1–4), 147–157. <https://doi.org/10.1016/j.pepi.2008.06.013>.
- Carmichael, T., Ailleres, L., 2016. Method and analysis for the upscaling of structural data. *J. Struct. Geol.* 83, 121–133. <https://doi.org/10.1016/j.jsg.2015.09.002>.
- Caumon, G., Collon-Drouillet, P.L.C.D., De Veslud, C.L.C., Viseur, S., Sausse, J., 2009. Surface-based 3D modeling of geological structures. *Math. Geosci.* 41 (8), 927–945. <https://doi.org/10.1007/s11004-009-9244-2>.
- Chen, J., Zhu, H., Li, X., 2016. Automatic extraction of discontinuity orientation from rock mass surface 3D point cloud. *Comput. Geosci.* 95, 18–31. <https://doi.org/10.1016/j.cageo.2016.06.015>.
- Chiles, J.P., Aug, C., Guillen, A., Lees, T., 2004. Modelling the geometry of geological units and its uncertainty in 3D from structural data: the potential-field method. In: *Proceedings Orebody Modelling and Strategic Mine Planning*, Perth, WA, 22–24 November 2004, pp. 313–320.
- Dadlez, R., Kopik, J., 1972. Stratygrafia i paleogeografia jury w Polsce. *Biul. Inst. Geol.* 252, 153–174 [in Polish].
- Davies, D.L., Bouldin, D.W., 1979. A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* 2, 224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>.
- De Berg, M., Cheong, O., Van Kreveld, M., Overmars, M., 2000. Computational Geometry: Algorithms and Applications. Springer-Verlag, Berlin, p. 364pp.
- Duelos Viana, C., Endlein, A., Ademar da Cruz Campanha, G., Henrique Grohmann, C., 2016. Algorithms for extraction of structural attitudes from 3D outcrop models. *Comput. Geosci.* 90, 112–122. <https://doi.org/10.1016/j.cageo.2016.02.017>.
- Fara, H., Scheidegger, A., 1961. Statistical geometry of porous media. *J. Geophys. Res.* 66, 3279–3284. <https://doi.org/10.1029/JZ066i010p03279>.
- Feldman-Olszewska, A., 1997. Depositional architecture of the polish epicontinental Middle jurassic basin. *Geol. Q.* 41, 491–508.
- Fienan, M.N., 2005. The three-point problem, vector analysis and extension to the n-point problem. *J. Geosci. Educ.* 53, 257–263. <https://doi.org/10.131402110716480>.
- Fisher, N.I., Lewis, T., Embleton, B.J., 1993. *Statistical Analysis of Spherical Data*. Cambridge University Press, p. 329pp.
- Garbowska, J., Łacka, B., Pazdro, O., 1978. Interrelation between microfauna and nature of dogger deposits of the Częstochowa Jura (Poland). *Acta Palaeontol. Pol.* 23 (1).
- 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. *Comput. Geosci.* 37, 1584–1594. <https://doi.org/10.1016/j.cageo.2011.03.007>.
- Gedl, P., Kaim, A., 2012. An introduction to the palaeoenvironmental reconstruction of the bathonian (Middle jurassic) ore-bearing clays at gnaszn, kraków-silesia homoclinal. *Acta Geol. Pol.* 62, 267–280. <https://doi.org/10.2478/v10263-012-0014-y>.
- Goldberg, D., 1991. What every computer scientist should know about floating-point arithmetic. *ACM Comput. Surv.* 23 (1), 5–48. <https://doi.org/10.1145/103162.103163>.
- Golub, G.H., Van Loan, C.F., 1980. An analysis of the total least squares problem. *SIAM J. Numer. Anal.* 17, 883–893. <https://doi.org/10.1137/0717073>.
- Górecka, E., 1993. Geological setting of the Silesian-Cracow Zn-Pb deposits. *Geol. Q.* 37, 127–146.
- Groshong, R.H., 2006. 3-D Structural Geology: A Practical Guide to Quantitative Surface and Subsurface Map Interpretation. Springer, Berlin, 400pp.
- Guibas, L., Stolfi, J., 1985. Primitives for the manipulation of general subdivisions and the computation of Voronoi. *ACM Trans. Graph.* 4 (2), 74–123. <https://doi.org/10.1145/282918.282923>.
- Hammah, R.E., Curran, J.H., 1999. On distance measures for the fuzzy K-means algorithm for joint data. *Rock Mech. Rock Eng.* 32 (1), 1–27. <https://doi.org/10.1007/s00050041>.
- Haneberg, W.C., 1990. A Lagrangian interpolation method for three-point problems. *J. Struct. Geol.* 12, 945–947. [https://doi.org/10.1016/0191-8141\(90\)90069-B](https://doi.org/10.1016/0191-8141(90)90069-B).
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning - Data Mining, Inference and Prediction*. Springer Science+Business Media, New York, p. 745pp.
- Hermaniśki, S., 1971. Wpływ prac odwadniających kopalnictwa rud żelaza na kształcenie warunków hydrogeologicznych w rejonie częstochowsko-kłobuckim. *Rudy Żelaza* 9–10, 13–16 [in Polish].
- Hert, S., Seel, M., 2018. dD convex hulls and Delaunay triangulations. In: *CGAL User and Reference Manual*, 13 edition, vol. 4. CGAL Editorial Board.

- James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning with Applications in R. Springer Science+Business Media, New York, p. 426pp.
- Jimenez-Rodriguez, R., Sitar, N., 2006. A spectral method for clustering of rock discontinuity sets. *Int. J. Rock Mech. Min. Sci.* 43 (7), 1052–1061 <https://doi.org/10.1016%2Fj.ijrmms.2006.02.003>.
- Jones, R.R., Pearce, M.A., Jacquemyn, C., Watson, F.E., 2016. Robust best-fit planes from geospatial data. *Geosphere* 12, 196–202. <https://doi.org/10.1130/GES01247.1>.
- Kopik, J., 1998. Lower and Middle jurassic of the north-eastern margin of the upper silesian coal basin. *Biul. Państwowego Inst. Geol.* (1989) 378, 67–130 ([in Polish]).
- Lajaunie, C., Courrioux, G., Manuel, L., 1997. Foliation fields and 3D cartography in geology: principles of a method based on potential interpolation. *Math. Geol.* 29 (4), 571–584. <https://doi.org/10.1007/BF02775087>.
- Leonowicz, P., 2015. Tubular tempestites from Jurassic mudstones of southern Poland. *Geol. Q.* 60, 385–394. <https://doi.org/10.7306/gq.1246>.
- Lindsay, M.D., Aillères, L., Jessell, M.W., de Kemp, E.A., Betts, P.G., 2012. Locating and quantifying geological uncertainty in three-dimensional models: analysis of the Gippsland Basin, southeastern Australia. *Tectonophysics* 546, 10–27. <https://doi.org/10.1016/j.tecto.2012.04.007>.
- Mallet, J.L., 1989. Discrete smooth interpolation. *ACM Trans. Graph.* 8 (2), 121–144. <https://doi.org/10.1145/62054.62057>.
- Mallet, J.L., 1992. Discrete smooth interpolation in geometric modelling. *Comput. Aided Des.* 24 (4), 178–191. [https://doi.org/10.1016/0010-4485\(92\)90054-E](https://doi.org/10.1016/0010-4485(92)90054-E).
- Marcotte, D., Henry, E., 2002. Automatic joint set clustering using a mixture of bivariate normal distributions. *Int. J. Rock Mech. Min. Sci.* 39 (3), 323–334. [https://doi.org/10.1016/S1365-1609\(02\)00033-3](https://doi.org/10.1016/S1365-1609(02)00033-3).
- Marich, A., Batterson, M., Bell, T., 2005. The morphology and sedimentological analyses of rogen moraines, central avalon peninsula, newfoundland, current research. Geological Survey, Newfoundland and Labrador Department of Natural Resources 1–14.
- Marynowski, L., Zatoń, M., Simoneit, B., Otto, A., 2007. Compositions, sources and depositional environments of organic matter from the Middle Jurassic clays of Poland. *Appl. Geochem.* 22, 2456–2485. <https://doi.org/10.1016/j.apgeochem.2007.06.015>.
- Matyja, B.A., Wierzbowski, A., 2000. Ammonites and stratigraphy of the uppermost bajocian and lower bathonian between częstochowa and wieluń Central Poland. *Acta Geol. Pol.* 50, 191–209.
- Michałak, M., 2018. Numerical limitations of the attainment of the orientation of geological planes. *Open Geosci.* 10, 395–402. <https://doi.org/10.1515/geo-2018-0031>.
- Mossoczy, Z., 1947. Sprawozdanie z badań geologicznych na zachód od Częstochowy w r. 1946. *Biul. Państwowego Inst. Geol.* (1989) 31, 43–50 ([in Polish]).
- Pakyuz-Charrier, E., Lindsay, M., Ogarko, V., Giraud, J., Jessell, M., 2018. Monte Carlo simulation for uncertainty estimation on structural data in implicit 3-D geological modeling, a guide for disturbance distribution selection and parameterization. *Solid Earth* 9 (2), 385–402. <https://doi.org/10.5194/se-9-385-2018>.
- Paor, D.G. De, 1991. A modern solution to the classical three-point problem. *J. Geosci. Educ.* 39 (4), 322–324. <https://doi.org/10.5408/0022-1368-39.4.322>.
- Pich, J., Pokora, M., 1982. Depressional cone in area of the Częstochowa-Kłobuck iron mines. *Przegląd Geol.* 30, 132–141 ([in Polish]).
- Razowska, L., 2001. Changes of groundwater chemistry caused by the flooding of iron mines (Częstochowa region, southern Poland). *J. Hydrol.* 244, 17–32. [https://doi.org/10.1016/S0022-1694\(00\)00420-0](https://doi.org/10.1016/S0022-1694(00)00420-0).
- Rippa, S., 1990. Minimal roughness property of the Delaunay triangulation. *Comput. Aided Geomet. Des.* 7 (6), 489–497. [https://doi.org/10.1016/0167-8396\(90\)90011-F](https://doi.org/10.1016/0167-8396(90)90011-F).
- Rosscience, 2017. Dips - Graphical and Statistical Analysis of Orientation Data.
- Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- Rutkowski, J., 1989. Geological structure of the cracow region, south Poland. *Przegląd Geol.* 37, 302–307 ([in Polish]).
- Scheidegger, A.E., 1965a. The tectonic stress and tectonic motion direction in the Pacific and adjacent areas as calculated from earthquake fault plane solutions. *Bull. Seismol. Soc. Am.* 55, 147–152.
- Scheidegger, A.E., 1965b. On the Statistics of the Orientation of Bedding Planes, Grain Axes, and Similar Sedimentological Data, vol. 525. U.S. Geological Society Professional Paper, pp. 164–167.
- Szubert, M., 2012. Pleistoceńska Morfogeneza Wyżyny Woźnicko-Wieluńskiej Związana Ze Stadiątem Maksymalnym Zlodowacenia Odry W Świecie Geostatystycznej Rekonstrukcji Powierzchni Podpleistoceńskiej. Wydawnictwo Naukowe Uniwersytetu Pedagogicznego, Kraków, 178pp. [in Polish].
- The Borehole Database - Deposits of Middle-Jurassic Ore-Bearing Clays Near Częstochowa, The Archive of Department of Fundamental Geology (University of Silesia in Katowice, Faculty of Earth Sciences), 1950–1960.
- Thiele, S.T., Jessell, M.W., Lindsay, M., Ogarko, V., Wellmann, J.F., Pakyuz-Charrier, E., 2016. The topology of geology 1: topological analysis. *J. Struct. Geol.* 91, 27–38. <https://doi.org/10.1016/j.jsg.2016.08.009>.
- Tokhmechi, B., Memarian, H., Moshiri, B., Rasouli, V., Noubari, H.A., 2011. Investigating the validity of conventional joint set clustering methods. *Eng. Geol.* 118 (3–4), 75–81. <https://doi.org/10.1016/j.enggeo.2011.01.002>.
- Vacher, H.L., 1989. The three-point problem in the context of elementary vector analysis. *J. Geol. Educ.* 37 (4), 280–287. <https://doi.org/10.5408/0022-1368-37.4.280>.
- Wang, J.A., Park, H.D., 2003. Coal mining above a confined aquifer. *Int. J. Rock Mech. Min. Sci.* 40 (4), 537–551. [https://doi.org/10.1016/S1365-1609\(03\)00029-7](https://doi.org/10.1016/S1365-1609(03)00029-7).
- Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58 (301), 236–244. <https://doi.org/10.1080/01621459.1963.10500845>.
- Watson, G., 1965. Equatorial distributions on a sphere. *Biometrika* 52, 193–201. <https://doi.org/10.1093/biomet/52.2.193>.
- Watson, G.S., 1966. The statistics of orientation data. *J. Geol.* 74, 786–797. <https://doi.org/10.1086/627211>.
- Wilk, Z., 1989. Hydrogeological problems of the Cracow-Silesia Zn-Pb ore deposits. *Dev. Earth Surf. Process* 1, 513–531. <https://doi.org/10.1016/B978-0-444-98874-4.50036-7>.
- Witkowska, M., 2012. Palaeoenvironmental significance of iron carbonate concretions from the bathonian (Middle jurassic) ore-bearing clays at gnaszny, kraków-silesia homoclinal, Poland. *Acta Geol. Pol.* 62 (3), 307–324. <https://doi.org/10.2478/v10263-012-0017-8>.
- Woodcock, N.H., 1977. Specification of fabric shapes using an eigenvalue method. *Geol. Soc. Am. Bull.* 88 (9), 1231–1236 [https://doi.org/10.1130/0016-7606\(1977\)88<1231:SOFSUA>2.0.CO;2](https://doi.org/10.1130/0016-7606(1977)88<1231:SOFSUA>2.0.CO;2).
- Zakrzewski, M., 1976. Charakterystyka Geologiczna Spągowego Poziomu Rud Syderytowych W Obszarze Częstochowskim. Wydawnictwa Geologiczne, Warszawa, 60pp. [in Polish].
- Zhan, J., Chen, J., Xu, P., Zhang, W., Han, X., Zhou, X., 2017a. Automatic identification of rock fracture sets using finite mixture models. *Math. Geosci.* 49 (8), 1021–1056. <https://doi.org/10.1007/s11004-017-9702-1>.
- Zhan, J., Xu, P., Chen, J., Wang, Q., Zhang, W., Han, X., 2017b. Comprehensive characterization and clustering of orientation data: a case study from the Songta dam site, China. *Eng. Geol.* 225, 3–18. <https://doi.org/10.1016/j.enggeo.2017.01.010>.
- Zhou, W., Maerz, N.H., 2002. Implementation of multivariate clustering methods for characterizing discontinuities data from scanlines and oriented boreholes. *Comput. Geosci.* 28 (7), 827–839. [https://doi.org/10.1016/S0098-3004\(01\)00111-X](https://doi.org/10.1016/S0098-3004(01)00111-X).
- Znosko, J., 1960. Tektonika obszaru częstochowskiego. *Przegląd Geol.* 8 (8), 418–424 ([in Polish]).
- Zuchiewicz, W., 1997. Distribution of jointing within magura nappe, west carpathians, Poland, in the light of statistical analysis. *Przegląd Geol.* 45, 634–638 ([in Polish]).
- Żelaźniewicz, A., Aleksandrowski, P., Bula, Z., Karnkowski, P.H., Konon, A., Oszczypko, N., Ślązak, A., Zaba, J., Żytko, K., 2011. Tectonic Subdivision of Poland. Komitet Nauk Geologicznych PAN, Wrocław, 60pp. [in Polish].