






## Article

# Decoding Rocks: An Assessment of Geomaterial Microstructure Using X-ray Microtomography, Image Analysis and Multivariate Statistics

Piotr Jan Strzelecki <sup>1,\*</sup> , Anna Świerczewska <sup>1</sup> , Katarzyna Kopczewska <sup>2</sup> , Adam Fheed <sup>1</sup> , Jacek Tarasiuk <sup>3</sup>  and Sebastian Wroński <sup>3</sup>

<sup>1</sup> Faculty of Geology, Geophysics and Environmental Protection, AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Kraków, Poland; swiercze@agh.edu.pl (A.Ś.); fheed@agh.edu.pl (A.F.)

<sup>2</sup> Faculty of Economic Sciences, University of Warsaw, ul. Długa 44/50, 00-241 Warsaw, Poland; kkopczewska@wne.uw.edu.pl

<sup>3</sup> Faculty of Physics and Applied Computer Science, AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Kraków, Poland; tarasiuk@agh.edu.pl (J.T.); wronski@fis.agh.edu.pl (S.W.)

\* Correspondence: piotr.strzelecki@agh.edu.pl

**Abstract:** An understanding of the microstructure of geomaterials such as rocks is fundamental in the evaluation of their functional properties, as well as the decryption of their geological history. We present a semi-automated statistical protocol for a complex 3D characterization of the microstructure of granular materials, including the clustering of grains and a description of their chemical composition, size, shape, and spatial properties with 44 unique parameters. The approach consists of an X-ray microtomographic image processing procedure, followed by measurements using image analysis and statistical multivariate analysis of its results utilizing freeware and widely available software. The statistical approach proposed was tested out on a sandstone sample with hidden and localized deformational microstructures. The grains were clustered into distinctive groups covering different compositional and geometrical features of the sample's granular framework. The grains are pervasively and evenly distributed within the analysed sample. The spatial arrangement of grains in particular clusters is well organized and shows a directional trend referring to both microstructures. The methodological approach can be applied to any other rock type and enables the tracking of microstructural trends in grains arrangement.

**Keywords:** microstructure; sandstone; deformation; X-ray microtomography; multivariate analysis; cluster analysis



**Citation:** Strzelecki, P.J.; Świerczewska, A.; Kopczewska, K.; Fheed, A.; Tarasiuk, J.; Wroński, S. Decoding Rocks: An Assessment of Geomaterial Microstructure Using X-ray Microtomography, Image Analysis and Multivariate Statistics. *Materials* **2021**, *14*, 3266. <https://doi.org/10.3390/ma14123266>

Academic Editor: Eddie Koenders

Received: 19 April 2021

Accepted: 9 June 2021

Published: 13 June 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Rocks are the most widespread material on Earth, a source of mineral resources and host to energy resources. Their properties are strongly related to their composition and microstructure, which constitutes the spatial and geometric configuration of all rock-forming components. Therefore, a full microstructural characterization of granular materials such as rocks requires a description of grain properties, including mineral composition, size, shape and their spatial arrangement [1]. Microstructural analysis is one of the most fundamental examinations in geological sciences as it provides information on conditions of the formation, deformation and alteration processes of rocks. Acquiring such information, in turn, helps to explain the variability of functional properties of geomaterials, including their mechanical or petrophysical parameters and allows for their detailed evaluation.

X-ray microtomography (micro-CT) followed by image analysis has unlocked an emerging possibility for complex 3D material microstructure characterization [2–4]. However, it provides a tremendous amount of information encoded in numerical data as a result. Therefore, the extraction of viable information usually requires the application of

exploratory and multivariate data analysis. By its application, one can observe patterns, trends, and clusters in grains distribution that can reflect, for instance, their origin [5].

Geoscience awaits the advent of rapid algorithms capable of analysing the mineral composition of rocks [6] and corresponding geometry of rock-forming components [7,8]. In all cases, the intention behind developing new solutions is to enhance the reproducibility of the research, minimize subjectivity and reduce time and costs [9]. One of the fast-growing in popularity platforms, R statistical software, has aroused interest especially among academics [10,11], guaranteeing the full reproducibility of the research.

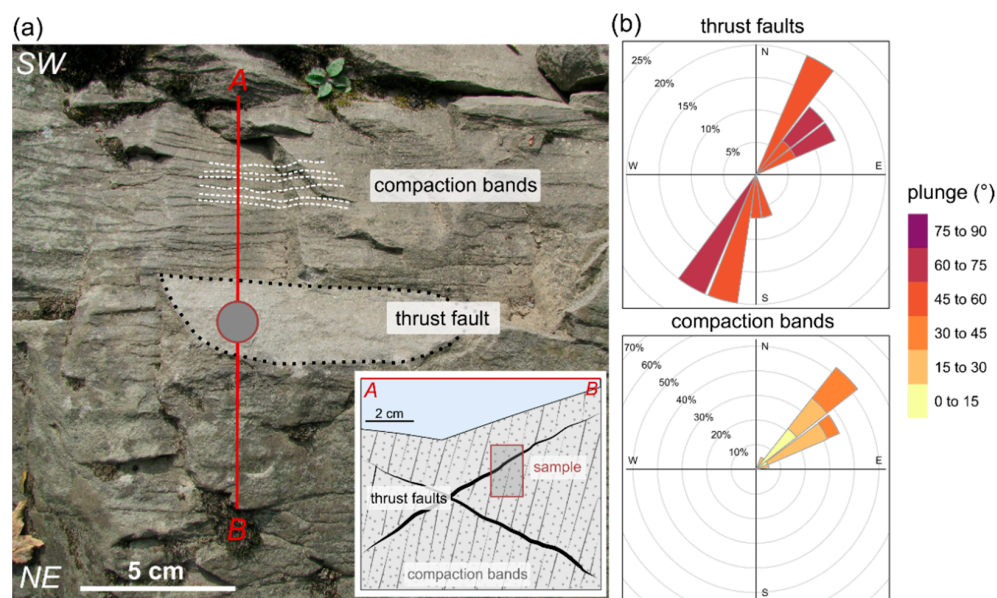
As rocks tend to be highly heterogeneous, manual interpretation of their microstructural parameters would be inefficient, and automated procedures are required. In this contribution, we present a statistical characterization protocol for the microstructural assessment of geomaterials with the fully reproducible code included. The workflow was tested on a naturally deformed sandstone sample. The sample contains two deformational microstructures i.e., the localized fault (shear fracture) with prominent surface and imperceptible macroscopically compaction bands. The study aimed to detect the microstructural changes associated with the microstructures. The protocol consists of data processing and analysis with the use of widely available freeware software of ImageJ for image analysis and the R software for numeric data analysis. It consists of micro-CT images pre-processing, feature extraction and segmentation, measurement of extracted objects and multivariate statistical analysis of the obtained results enabling the microstructural evaluation of the sample.

## 2. Materials

As a test geomaterial, a Carpathian sandstone was used. Carpathian sandstones are popular rock materials, typically used as aggregates, cladding material, and armour stone e.g., [12]. Moreover, Carpathian sandstones serve as regional reservoir rocks, hosting oil and gas deposits [13]. The Carpathian sandstones are characterized by strong internal changeability, which also directly influences e.g., the spread of geomechanical properties [12]. Moreover, the primary microstructure of the sandstones is often overprinted by diagenetic e.g., [14–16] and tectonic processes e.g., [17].

The studied sample was collected at the Nasiczne locality (N49.1732, E22.5980) from a natural exposure in SE Poland. At the exposure, thick-bedded sandstones (the Otryt Sandstone) of the Krosno Beds show pervasively distributed deformation of tectonic microstructures [18]. The micro-structures occur abundantly in the form of thrust faults and compaction bands (Figure 1a). The thrust faults dominantly dip towards NE (ca. 30°) and SE (ca. 210°) under low to moderate angles (15–45°). The compaction bands dip towards SW (ca. 230°) under moderate to high angles (55–85°; Figure 1b). The thrust faults form distinct surfaces, whereas the compaction bands are sometimes macroscopically imperceptible.

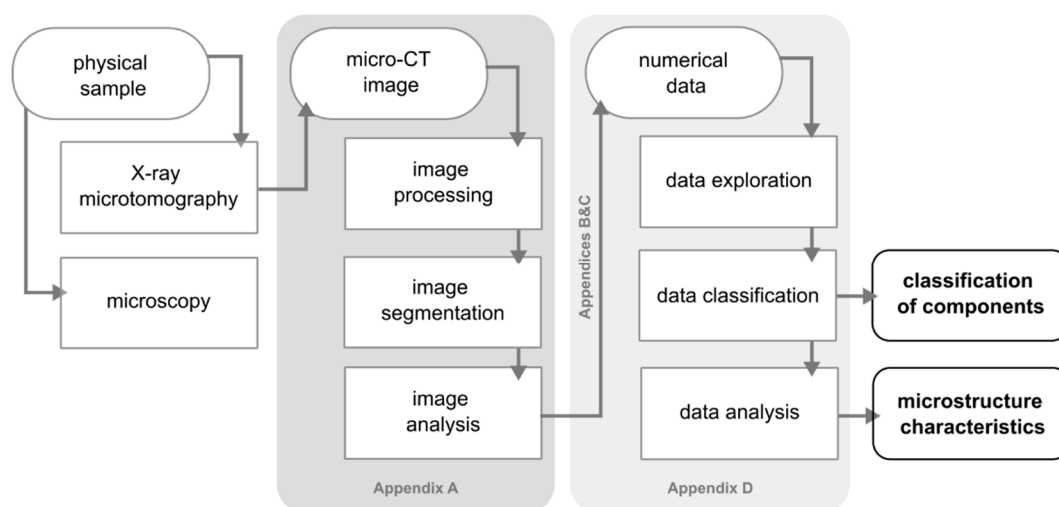
The centimetre-scale samples were collected for microscopical and petrographical analyses. The additional cylindrical core sample of approximately 2 cm in diameter and 2 cm in height bearing the tectonic micro-structures was withdrawn for microstructural assessment with the use of X-Ray microtomography (Figure 1a). This sample was collected with a reference to geographical directions, allowing to reconstruct the original position of the sample after scanning and correlate observed deformation micro-structures in the exposure with the microstructure of the scanned sample.



**Figure 1.** (a) A close view at the sandstones in the exposure and deformational microstructures within them. A situational cross-section of the sampling is presented in the bottom right corner; (b) The orientation of the micro-structures. Note that the orientation of normals to the micro-structures planes is presented.

### 3. Methods

The methodology consists of two main techniques (Figure 2). The first one comprises image processing and analysis and the second one concerns processing and statistical analysis of the resultant numerical data obtained by the former. The protocol and reproducible code described in detail are available in Appendices A–D. The analysis was supplemented by the previous examination of the specimen using polarized light microscopy and cathodoluminescence. The processing and analysis were performed on a notebook equipped with an Intel Core i5-7300HQ (Intel Corporation, Santa Clara, CA, USA) CPU, NVIDIA GeForce GTX 1050 (Nvidia Corporation, Santa Clara, CA, USA) GPU and 16 GB of RAM operating on Windows 10 (64-bit version). The validation of the protocol and calculations of grains ordination on a standard sample was performed (see Appendix E).



**Figure 2.** The workflow of data processing and analysis.

### 3.1. Polarized Light Microscopy and Cathodoluminescence

Supplementary analysis of thin sections corresponding to the core sample was performed. Basic mineralogical composition of the sample was recognised using the point count method with 300 points per thin section utilised. The analysis and microphotographs were taken using an AxioImager.A1m microscope (Carl Zeiss Jena GmbH, Jena, Germany) in both plane- and cross-polarised light. Additionally, petrographic studies using an Eclipse Ci-L microscope (Nikon Instruments Inc., Melville, NY, USA) with a Mk5 cold cathode (Cambridge Image Technology Ltd., Hatfield, UK) under 15 keV voltage with the electron beam current of 400  $\mu$ A were performed. The polarized light microphotographs, as well as cathodoluminescence photos, were taken at the Department of Fossil Fuels, AGH University of Science and Technology (Kraków, Poland).

### 3.2. X-ray Microtomography

The collected cylindrical sample was subjected to X-ray computed microtomography at the Laboratory of Micro and Nano Tomography of the AGH University of Science and Technology. The sample was scanned with a Nanotom S device (GE Sensing & Inspection Technologies GmbH, Wunstorf, Germany) equipped with a nanofocus X-ray tube. The tomograms were registered on a HAM C 7942CA-02 detector (Hamamatsu Photonics K.K., Hamamatsu, Japan). The reconstructions of the measured objects were performed with an aid of the proprietary GE software datosX ver. 2.1.0 and using the Feldkamp algorithm for cone-beam X-ray CT [19]. The sample was scanned at 100 kV of source voltage and 400  $\mu$ A tube current, with a full, 360° rotation of the specimen in 1800 steps. The exposure time equalled 500 ms and the frame averaging of 5 and the image skip of 1 was applied, resulting in a scanning time of 85 min. A stack of images contained approximately 2000 slices of 2300  $\times$  2300 pixel-sized images. The resolution of the image stack was 12  $\mu$ m. The resultant 3D image of the sample represented in grey-scale reflects the X-ray attenuation, which is related to the density ( $\rho$ ) and atomic number ( $Z$ ) of the sample-forming components e.g., [20]. The coefficient of X-ray attenuation is proportional to  $\rho$  and  $Z^4$  at low scanning energies such as these utilized in the present study e.g., [21]. The components with the lowest coefficient of X-ray attenuation are represented by the darkest pixels, while the brightest ones represent those of the highest attenuation e.g., [21].

### 3.3. Image Processing and Analysis

The image processing and analysis were performed using the Fiji software (ImageJ version 1.53c, Java 1.8.0\_66, 64-bit) [22]. The procedure of the processing and analysis of the image is described in detail with a sequence of steps to perform in Fiji in Appendix A. Firstly, the raw image (2284  $\times$  2284  $\times$  1674 pixels) was converted into the 8-bit format resulting in a file size of 8 GB. Subsequently, image enhancement was performed (Appendix A, step 5). Consequently, corrections of brightness and contrast were applied, improving the legibility of the image. Considering the cylindrical shape of the sample, the image was cropped and limited to 1000 pixels (12 mm)-sided cube constituting a volume of interest (VOI; file size 0.9 GB). On the VOI correction of the X-ray attenuation using BaSiC [23] was applied. Feature extraction was performed using the global thresholding method [24]. The granular components of the rock above the global threshold value were extracted and binarized. The binarized image was filtered using a 3D median filter (kernel size 2  $\times$  2  $\times$  2 pixels) and the objects at the image edges were removed. Lastly, a size filter, removing the objects below 100 voxels in volume (approximately <0.07 mm in diameter), was applied to eliminate the artefacts and imprint of the smallest insignificant objects with the size closest to the resolution of the image. Finally, the resultant image was segmented to attribute individual objects with a unique label. The total time of image processing was approximately 30 min. The prepared image was subjected to the measurements. They were performed using the ImageJ 3D suite [25] and MorphoLibJ [26] plugins and lasted approx. 40 min in total. The results of these measurements were combined and a data set consisting of 44 parameters was prepared using R (Appendix B). The parameters were prepared based



on these available by default in the image analysis software. Their choice was dictated by the simplicity of their interpretation as well as their relatively short computation time. The parameters fall into 4 categories describing the chemical composition, size, shape, and spatial orientation and arrangement of each object representing a grain. Each category is represented by several unique parameters (Appendix C, Table A1). By introducing a higher number of parameters i.e., 44, one obtains a possibility of tracing detailed variability of grains properties is provided. Parameters concerning the spatial orientation of grains were transformed into a spherical coordinate system (Appendix C, Figure A1). The calculated parameters are input variables for further analyses which were performed in R software with additional libraries (Appendix C, Table A2).

### 3.4. Exploratory Data Analysis

For exploration and visualization of the data set, principal component analysis (PCA) [27–29] and factor analysis (FA) [30] were performed. PCA and FA are statistical techniques used for multivariate data exploration and allow for dimension reduction, simultaneously retaining most of the variation in the data set. As a result of PCA, principal components (PCs) are calculated which are the eigenvalues and eigenvectors of the correlation matrix of the original variables. PCs are mutually uncorrelated and can delineate different characteristic features of the data set objects. Each principal component (PC) explains a portion of the total variance in a data set. Because variables are measured in a mixture of units and ranges, the data were standardized. To determine the number of PCs, a scree plot [31] was used. To uncover the underlying structure of PCs, FA was performed. The structure of PCs was examined using standardized loadings and indexes of uniqueness, communality and complexity [32] (Appendix D, step 4).

### 3.5. Cluster Analysis

Cluster analysis is a method of unsupervised learning and pattern recognition that allows for classifying and identifying groups in a data set. Cluster analysis allows similar observations to be grouped into homogenous subsets with no previously known pattern. The data is grouped based on the similarity between the observations concerning the values of variables for each observation. The distance between the pair of observations is a measure of the dissimilarity. Observations within the same cluster are similar and different from other clusters. The clusters themselves are dissimilar. In the studied case, cluster analysis based on a k-medoid method using the Clustering Large Applications (CLARA) algorithm [33,34] was performed. The CLARA algorithm is a variant of the Partitioning Around Medoids (PAM) algorithm dedicated to large data sets. Clustering was performed on the standardized original data set using the Manhattan distance matrix. For the classification of the components, the variables describing properties that are independent of the spatial orientation of components, i.e., these describing their chemical properties, size and shape were employed. The optimal number of clusters was determined based on silhouette statistics [35] (Appendix D, step 5).

### 3.6. Microstructure Assessment

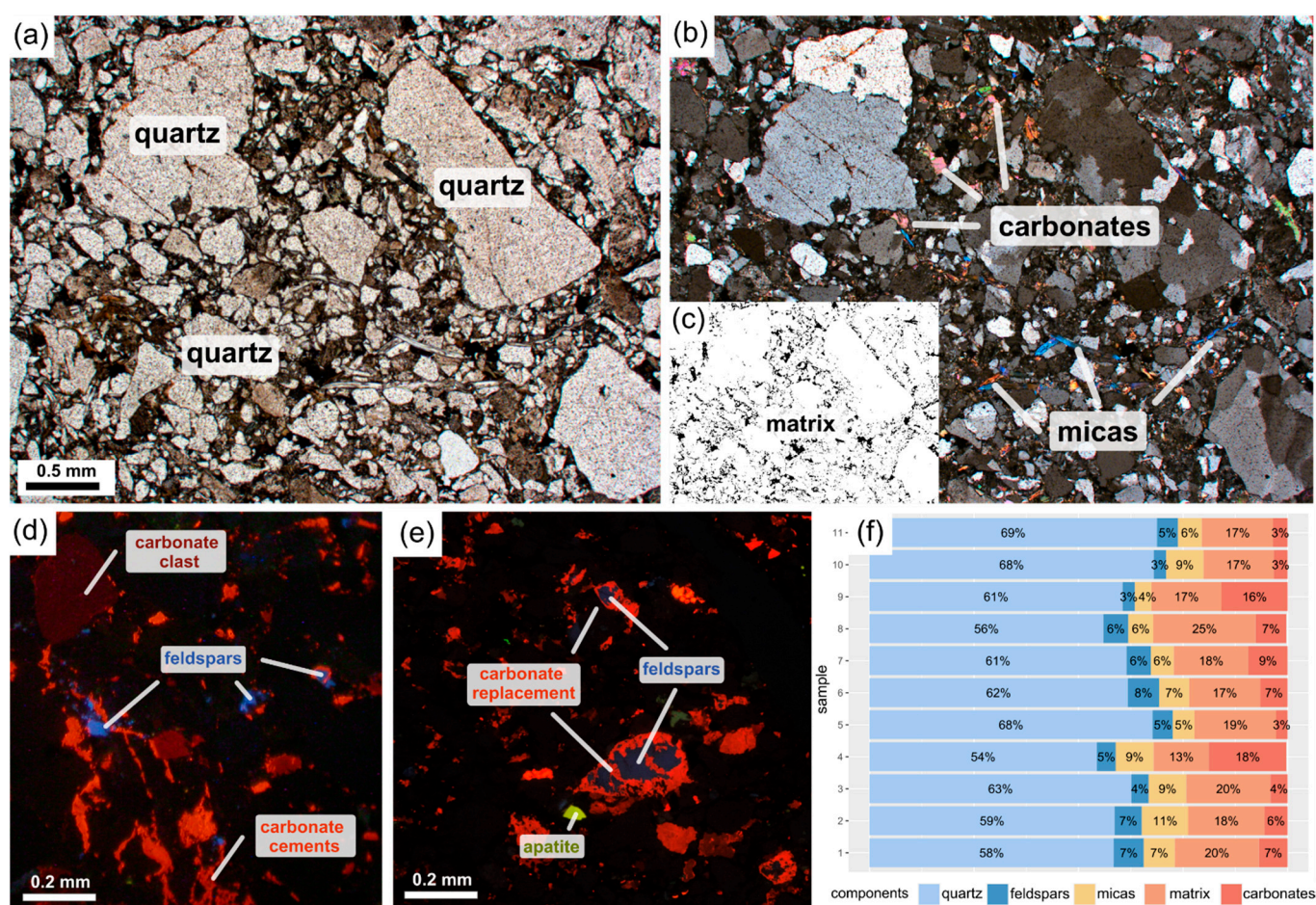
Microstructure assessment was performed based on the results of the cluster analysis. The distribution of parameters was examined in the previously determined groups. The distribution of variables describing the chemical composition, size, and shape of grains were illustrated on violin plots cf. [36]. For the variables describing the spatial configuration of objects, the radial distribution function was calculated [37]. The radial distribution function finds the centre of grain in a given position at a radial distance from the centre of a reference sphere, which was set to the centre of the VOI. This measurement characterizes packing structures and contains information about interparticle correlations and their organization. To examine the spatial orientation of grains, rose plots were constructed. The rose plots provide information on the direction and dip angle of particular grains. The orientation of

the grains was projected based on the orientation of fitted ellipsoid semi-axes (Appendix D, step 6).

## 4. Results

### 4.1. Sample Characterization

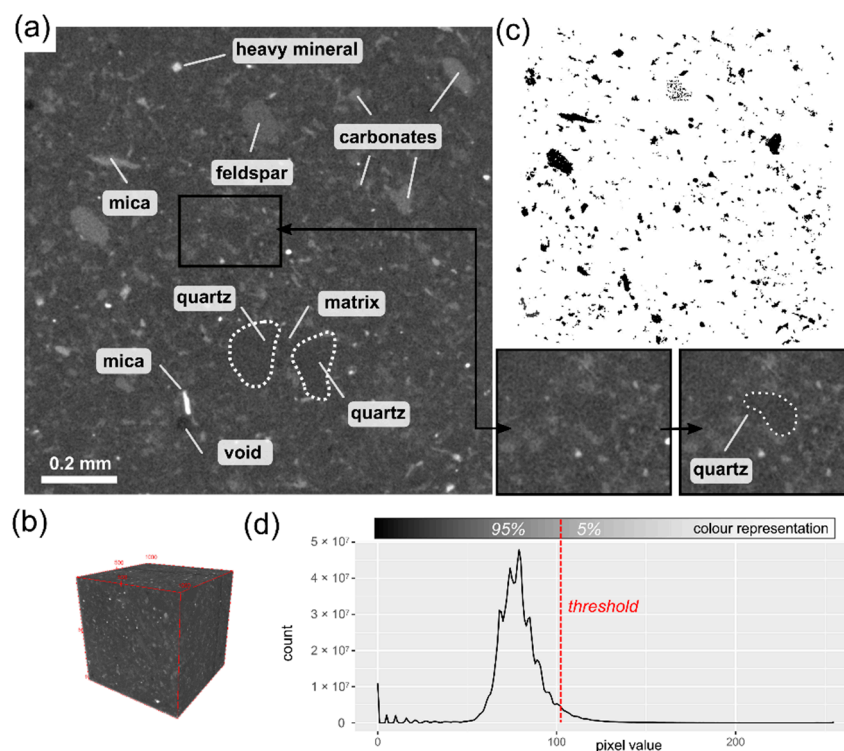
The test material is a poorly sorted sandstone. It is composed predominantly of quartz grains (54–69%), subordinately of micas (4–11%), including a substantial amount of muscovite and a subordinate share of biotite, feldspars (3–7%), and carbonates (3–18%) in the form of grains and cement (Figure 3a–f). The last mentioned sometimes replace feldspars (Figure 3d–f). The intergranular space of the sandstone is filled with matrix (Figure 3c), composed of a mixture of silt-sized minerals of the granular framework and clay minerals. The matrix constitutes 13–25% of the rock's volume (Figure 3a–f). Other accessory minerals occur in a small proportion (<1%) and are represented by heavy minerals such as pyrite or apatite.



**Figure 3.** (a) Microphotographs showing rock composition and microstructure corresponding to the sampled rock shown in the plane-polarised light and (b) cross-plane polarized light; (c) A mask of the matrix extracted from Figure 3a; (d,e) the cathodoluminescence microphotographs showing the mineral composition of the sample; (f) Results of the mineral composition determination using the point-count method on the corresponding thin sections.

In the micro-CT image, the mineral composition is harder to recognise as some minerals show similar properties in terms of X-ray absorption and consequently attain the same intensity of pixel values. Quartz, feldspars and matrix together constitute 90% of the rock's volume and in the micro-CT and appear as grey-coloured background (Figure 4a,b). Some minerals are distinguishable from the background as they show higher absorption of X-rays

and thus lighter shades of grey. These minerals represent carbonates, micas and heavy minerals, and their granular form can be extracted from the micro-CT image. These grains constitute 5% of the VOI (Figure 4c). The remaining 95% of the VOI are made up of voids, which constitute unfilled spaces within the rock. These volume shares are comparable with the results of mineral composition recognition from the point count method. The extracted components are pervasively distributed within the VOI, with a total number of 25,015 individual grains.



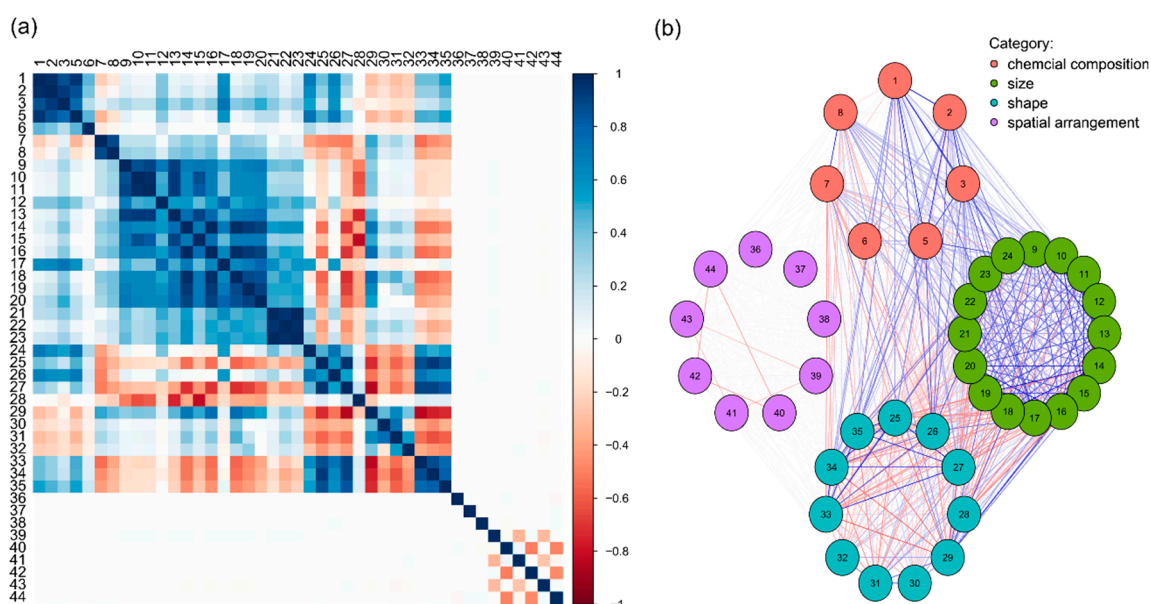
**Figure 4.** (a) Selected micro-CT image of the sample with mineral components marked. Note that outlines of quartz grains are visible; (b) 3D visualization of the VOI; (c) The extracted and analysed mineral components, the image corresponds to this shown in Figure 4a; (d) Histogram of colour distribution (pixel value) within the analysed volume of the sample.

#### 4.2. Data Exploration

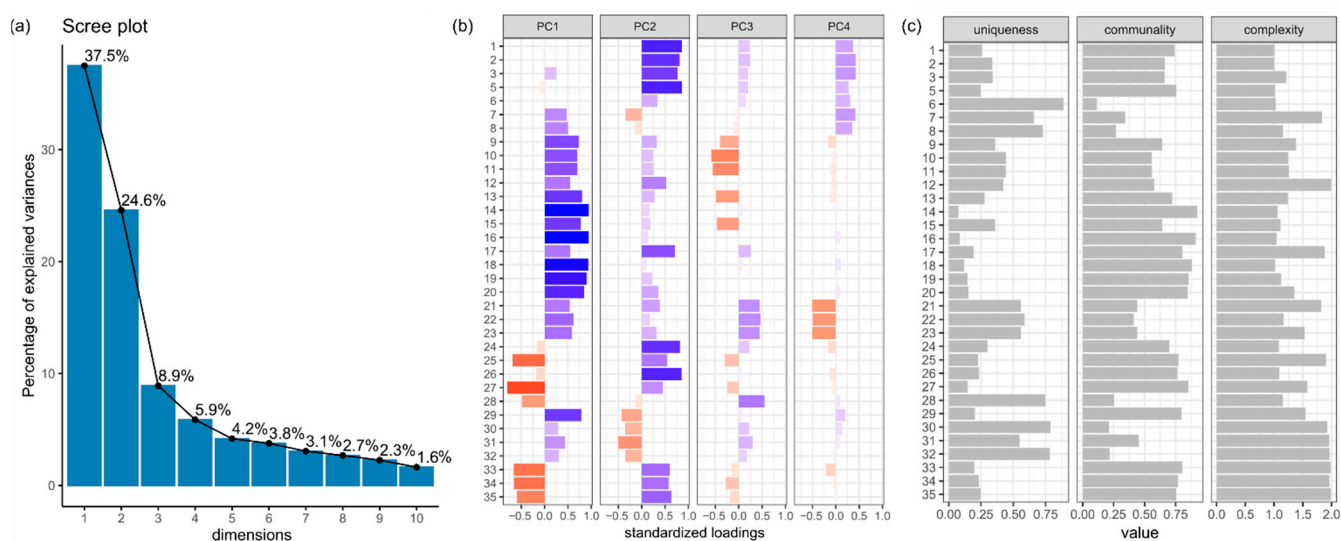
Each grain was described with 44 parameters. The minimal intensity parameter was excluded from the further analysis as its variance equals zero. The initial variables are correlated. The correlation coefficients between variables are high, absolute values above 0.5, especially in the case of variables describing similar properties within the category (Figure 5a). However, correlations between variables of a different category are also noticeable. The mutual correlation between variables is strong in the case of chemical properties, and size and shape variables. The variables in the category of spatial arrangement show either weak or no correlation, both between the variables within that category and the variables from other categories (Figure 5b).

For the correlated variables (all except the spatial arrangement category), PCA was conducted. The scree plot (Figure 6a) shows that the initial number of variables can be reduced to a lower number of dimensions. The first four principal components (dimensions) explain 76.9% of the total variance in the data set, whereas the first two account for the majority of it i.e., 37.5% and 24.6%, respectively. The first two PCs depict the general variability in the data set.





**Figure 5.** (a) Heatmap of the correlation matrix between the variables. The colour designates the value of the correlation coefficient between particular variables; (b) A graph showing correlation between variables in a form of network. The colour of the nets corresponds to this presented in Figure 5a and designates the correlation coefficient. The variables are grouped regarding their categorical subdivision. Numbers (1–44) correspond to the names of variables (see Appendix C, Table A1).



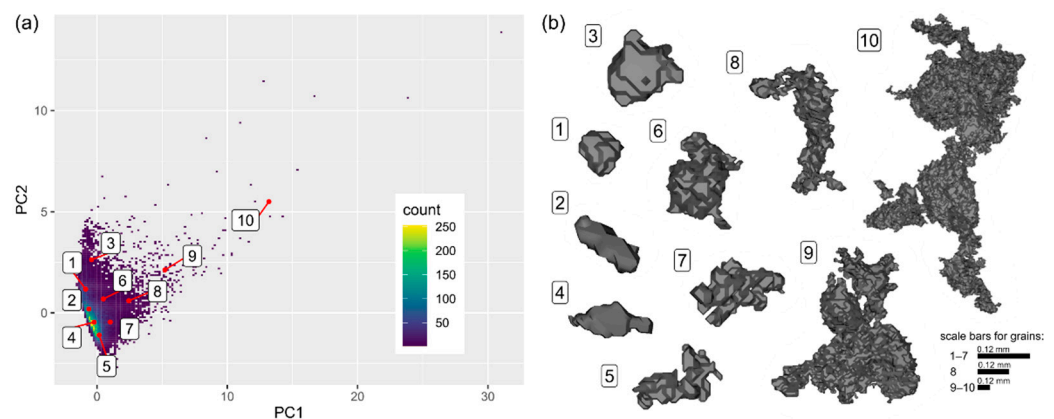
**Figure 6.** (a) The scree plot; (b) The structure of the first four PCs. Colours correspond to the value of standardized loading; (c) The statistics for the first two PCs. Numbers (1–35) correspond to the names of the variables (see Appendix C, Table A1).

Further dimensions explain less than 5% of the variance. The structure of the first four PCs indicates an ordered structure of the initial variables underlying the dimensions. The first PC shows an input of variables describing geometrical properties of the grains such as the size and shape (Figure 6b; variables no. 9–35), whereas the second PC is loaded from the variables describing the chemical composition (Figure 6b; variables no. 1–8) and geometrical properties. However, in the case of PC2 the loading values from the geometrical variables are lower, and in PC3 and PC4 they are mainly below 0.5 of the absolute value. The communality index for the first two PCs indicates that variables such as describing intensity distribution statistics, Euler number or ratios of ellipsoid semi-axes



(Figure 6c; variables no. 6–8, 28, 30–32 respectively) are more unique as they show a low proportion of common explained variance. The average complexity equals 1.4.

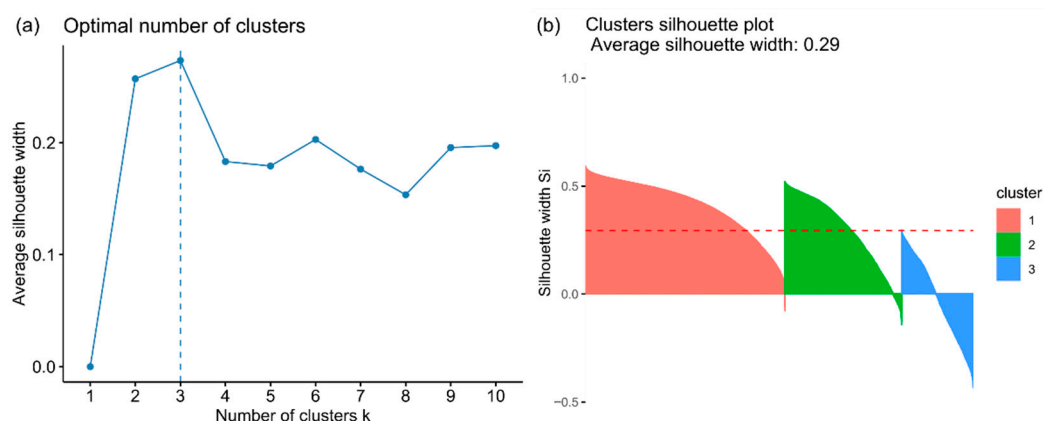
The projection of the data onto the PC coordinate system suggests that the data do not show any distinctive clusters and are projected densely and continuously closer to the origin of the PC coordinate system. Towards increasing PCs, the tail of outliers is observed (Figure 7a). The outliers are represented by the largest individual grains or aggregates of mutually connected grains (Figure 7b).



**Figure 7.** (a) Projection of the objects into the PC coordinate system using a density map; (b) The visual sample of 10 grains (objects) from the dataset. The labels correspond to these in (a).

#### 4.3. Classification of Components

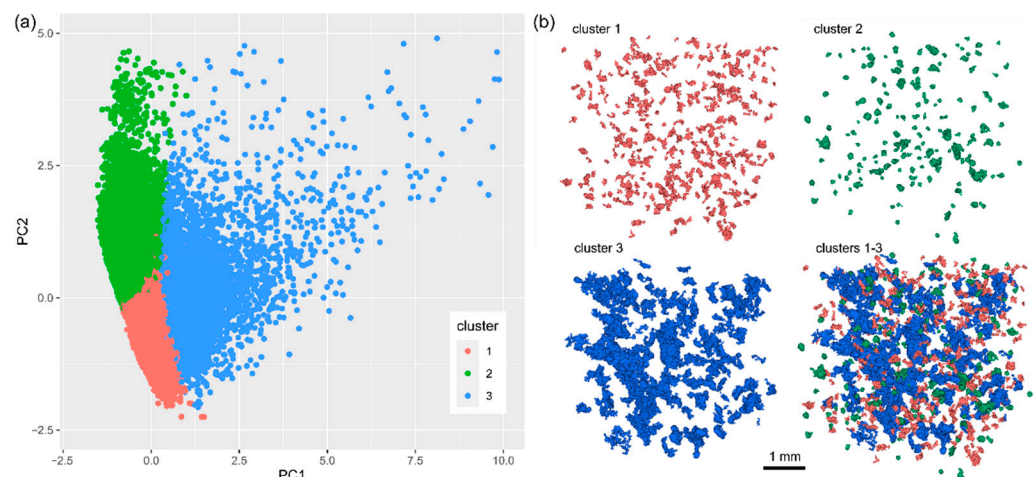
The average silhouette width statistics indicate 3 as the optimal number of subsets within the data (Figure 8a), with an average width of cluster equalling 0.29. The clustering of the data set returns three clusters with 12,878, 7519 and 4618 objects, respectively. The average widths of the three obtained clusters are 0.42, 0.30 and 0.01, respectively. The first two clusters show a major contribution of the positive values indicating that the objects belonging to these two clusters are similar within their cluster, whereas in the third one, a great number of negative values points on a general dissimilarity and diversification of the objects within that cluster (Figure 8b).



**Figure 8.** (a) The optimal number of clusters determined using average silhouette width statistics; (b) The quality of clustering using 3 as the optimal number of clusters based on average silhouette width statistics.

Observations within the first and second cluster are concentrated, whereas, in the third one, they are more dispersed. The third cluster also consists of outlier observations (Figure 9a). The results of the clustering appear to be satisfactory upon visual inspection.

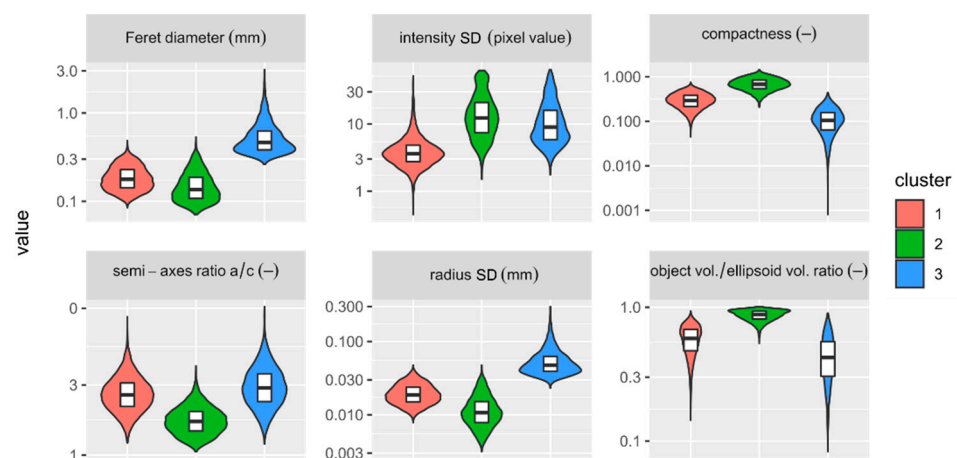
The first and second clusters represent relatively small grains as compared to those from the third cluster. The grains in the first cluster appear to be more irregular in comparison to these of the second cluster. The grains of these two clusters are secondary and fill the intergranular space of the specimen, corresponding to the main (largest) grains of the third cluster (Figure 9b).



**Figure 9.** (a) Results of the cluster analysis: data classification in the PC coordinate system and (b) visual representation of the clustering results presented as a fragment of the VOI.

#### 4.4. Microstructure Assessment

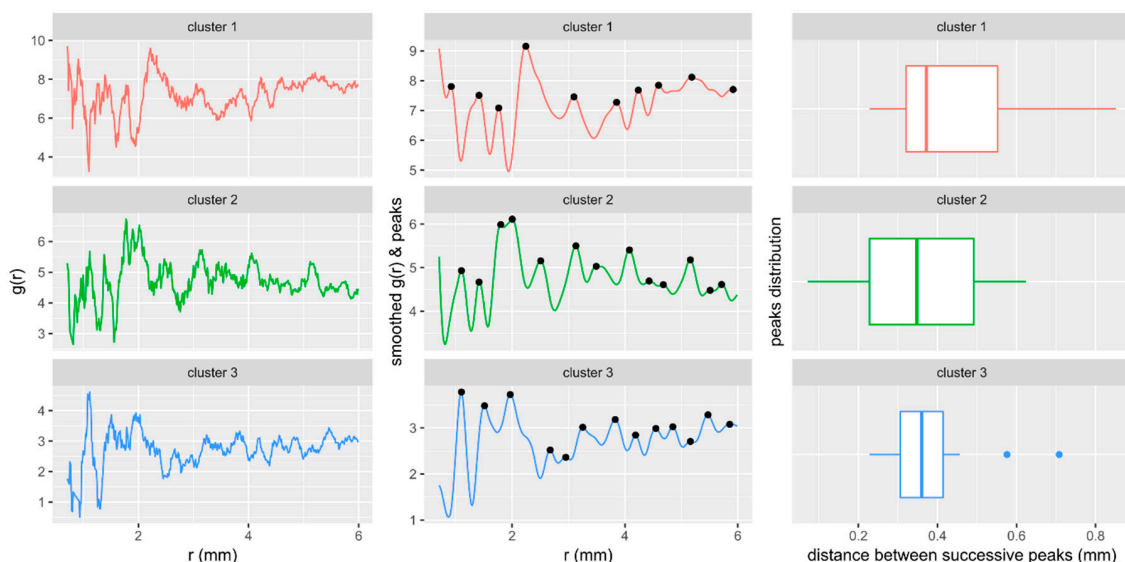
The general observations highlighted in the previous section find confirmation in the numeric data. The clusters show diversified grain properties (Figure 10). In general, grains in the first and second clusters are smaller in comparison to those of the third one (Figure 10; e.g., the median Feret diameter equals 0.18, 0.16, 0.46 mm, respectively). The second and third clusters are more chemically diversified in comparison to the first one (Figure 10; e.g., median intensity SD equals 13, 9, 3, respectively). The second cluster consists of regular grains of ellipsoidal and spherical shapes (Figure 10; e.g., high compactness and low semi-axes ratio  $a/c$ ), whereas the first and the third ones contain more irregular grains (Figure 10; e.g., higher radius SD values and lower object to ellipsoid volume ratios).



**Figure 10.** Distribution of the selected parameters in the clusters.

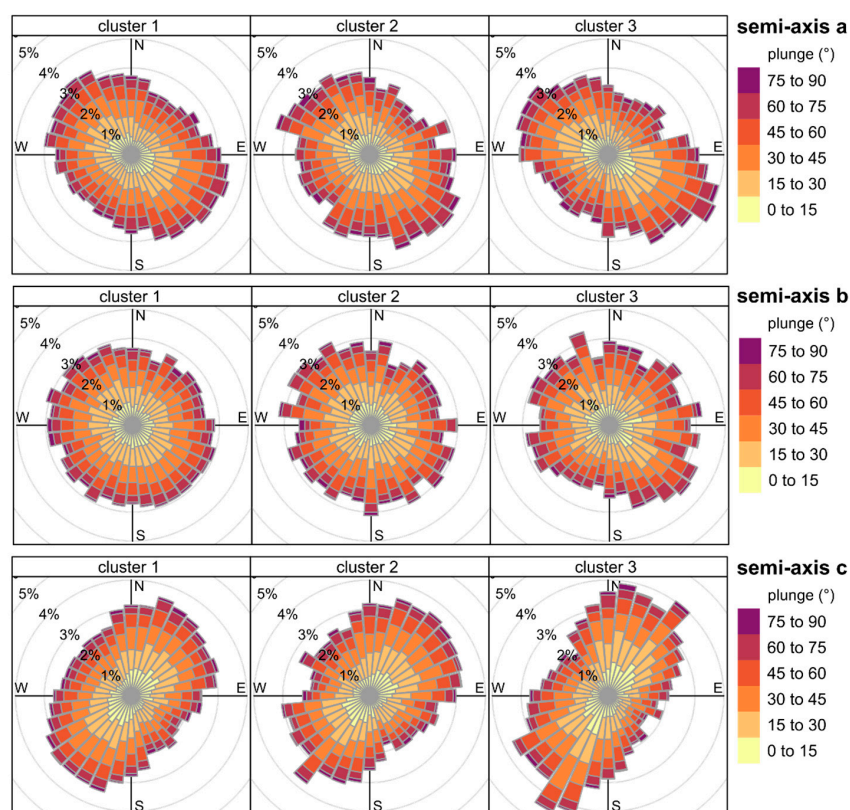
The grains are evenly distributed in the specimen. With increasing distance from the centre of the VOI, the periodicity and density of the distribution of grains' geometrical centres remain stable (Figure 11). The specimen shows a dense and compact packing of

the granular framework. The distance between the successive peaks, i.e., the distances between the geometrical centres of grains in particular clusters, corresponds to the average grain size in the third cluster. Moreover, all clusters show a similar median value which ranges from 0.35 to 0.37 mm, indicating that the third cluster played a determinative role in controlling the microstructure of the sample.



**Figure 11.** Radial distribution function  $g(r)$  calculated for the geometrical centres of grains in the clusters.

The grains are well arranged in all clusters and the distribution of their orientation delineates distinct trends, indicating the presence of directional microstructure (Figure 12).



**Figure 12.** Spatial orientation of the grains in the clusters. The width of a bin is  $10^\circ$ ; N equals  $0^\circ$ .

The distribution of the longest semi-axes (a) orientation in the grains is mainly NE-SW-oriented, whereas the orientation of the intermediate semi-axes (b) is uniform in all directions in all clusters. Conversely, the shortest semi-axes (c) are NW-SE-oriented. The number of observations in the dominant direction is higher by 1 to 3 per cent points in comparison to other directions. The orientation of the shortest semi-axes (c semi-axes), which are normal to the ab surfaces, indicates that the grains are predominantly tilted in the NE or SW direction under high angles. However, the dominant directions are slightly shifted between the clusters. The spread is best visible between the second and third cluster considering the orientation of semi-axes c. In the second cluster, the dominant direction aligns with  $50^\circ$ , whereas in the third cluster with  $30^\circ$ . It is noteworthy that within the third cluster there is a considerable number of observations trending  $50^\circ$  under low plunge values, as well (Figure 12; the third cluster c semi-axes orientation). In the first cluster, the dominant direction is the resultant of these two directions. The frequency of observations in the opposite directions remains at a similar level.

## 5. Discussion

### 5.1. Image Processing and Analysis

Segmentation of the image remains challenging as rock-forming minerals typically display similar attenuation properties [20]. Ideally, minerals should be characterized by a diversified density. Therefore, not all granular materials or rocks would be suitable for the analysis. Likewise, in this case, only a limited content of the whole sample was analysed. Although the selected components constitute only 5% of the total rock's volume, the grains subjected to the analysis are widespread and dispersed across the entire VOI. Therefore, in our interpretation, they are representative and reflect the general microstructure of the sample. As a consequence, even an analysis of a limited volume can be sufficient to obtain proper information on the microstructure of a geomaterial as a whole. Some of the grains seem to be connected (Figure 7b), which may be seen as an artefact. However, as revealed by the supplementary analysis of the thin sections, carbonates may form such structures. Nonetheless, the connected grain chains are inevitable as some of them might be naturally contacting. Therefore, more advanced methods for grain segmentation could be considered cf. [21].

### 5.2. Data Exploration

The first two PCs explain 62.1% of the total variability in the data set. This value is slightly above the minimum recommended threshold of 60% [38]. In general, it is advised to maximize the value of the total explained variance retaining subsequent PCs until they account for at least 95% or to the point when PCs explain at least 5% of the variance [38]. Therefore, the first two PCs allowed for projecting the data onto a two-dimensional space and tracing the basic properties responsible for data variabilities such as grain size, shape and chemical composition. Following the mentioned rule, and adding the next two PCs would not significantly increase the explained variance. Conversely, this would complicate projecting the data by imposing four dimensions in total. The value loadings should be at least 0.3–0.4 of the absolute value. In this case, they were usually above 0.7 which indicates a strong correlation of the initial variables with the PCs [38]. Some of the variables load more than one PC (Figure 6c; complexity index), but the mean complexity index remains low (1.4 on average) and the structure remains clear. In the case of high complexity index, a rotation of the PCs to simplify the interpretability of the underlying structure can be considered cf. [29]. Usually, the projection of the data depends on the purpose of the analysis. It is a common practice to only utilise selected single variables.

### 5.3. Data Classification

In this study, a k-medoid-based clustering algorithm was chosen. The advantage of this is the following. Firstly, it minimizes the sum of dissimilarity in the pairs of observations, contrary to the k-means algorithm which minimizes the total mean standard



error. The k-medoid algorithm is less sensitive to both the outliers and noise. Additionally, cluster centres are designated based on real observation (medoid) and can handle the clusters of different density and size cf. [39]. In the case of high dimensional data, the Manhattan distance matrix usage is recommended [40]. In the case of this study, the statistically optimal number of clusters was designated as three with the average silhouette width statistics equalling 0.29. The value of the last quantity indicates that the data might be structured and support the validity of cluster analysis application [34]. However, the designation of the number of clusters is subjective and depends on the method applied cf. [41]. The choice of the cluster number can be based on silhouette statistics, internal or external criteria [42]. The latter assumes a comparison of the cluster solution with external results. For such a solution, manual and visual assessment of the cluster solution (see Figure 9b) and expert knowledge should be considered for the validation of the obtained results.

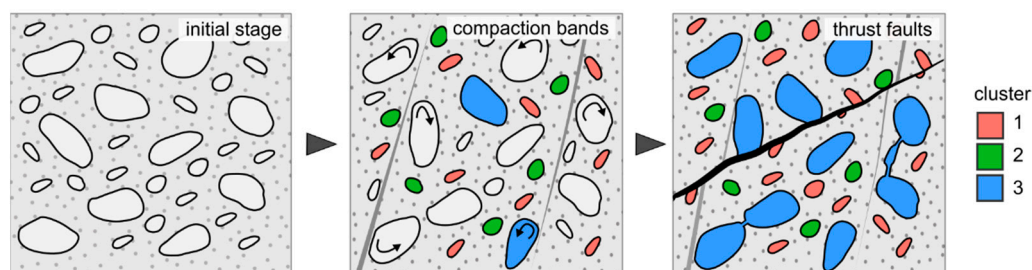
Comparison of clustering results with a different number of clusters can be applied as a data exploration method to gain an understanding of the patterns and features of a data set. Initially introducing a higher number of clusters might also result in a more detailed classification of grains allowing for peculiar grain clusters detection. The choice of the initial variables (grain parameters) as input for cluster analysis is also important, as their different combination may lead to significantly different results. Therefore, their choice should be based on the purpose of the analysis and the classification of the selected aspects of data. By using only selected parameters e.g., only from one category, distinct classification of chemical composition, size or shape of grains could be obtained. Therefore, introducing a relatively high number of parameters i.e., 44 was aimed to provide as much information about the variability of the sample as possible (e.g., Figure 6c). The role and significance of particular parameters depend on the nature of geomaterial and its grains and it may vary depending on the material.

#### 5.4. Microstructure Interpretation: A Case Study

The preferred arrangement of grains in sedimentary rocks such as sandstones can be either a result of sedimentary or deformational processes cf. [43]. In the case of the studied sample, their deformational origin seems to be responsible for the resultant microstructure of the sample. The grains mostly dip under moderate to high angles, indicating their rearrangement and steepening during the deformation. Moreover, the alignment of the grains is consistent with the direction of micro-faults and compaction bands. In the case of the sedimentary origin of grains arrangement, the grains are tilted under low to moderate angles and the presence of directional trends is a result of their directional transportation cf. [43]. The direction of the paleo-current inferred from the surrounding strata indicates a NE-SW-directed material transport [44]. This direction coincides with the orientation of the observed deformation, but the arrangement of grains indicates that the deformation overprinted the primary sedimentary microstructure.

The application of cluster analysis grouped grains into groups among which two of them are distinctive i.e., cluster no. 2 and 3. The second cluster can be related to primary grains of the rock, heavy minerals such as biotite, which were deposited during sedimentation and were involved in deformation from the beginning. Their alignment is in accordance with the compaction bands. The third cluster consists of grains of mixed origin. It seems that initially they were involved in the reorientation recorded by grains of the second cluster and then overprinted by diagenetic processes during which some of the original grains were dissolved and replaced by carbonate cement e.g., (Figure 3d,e). The carbonate cementation and the replacement of feldspars by carbonates is commonly observed in Carpathian sandstones e.g., [15,16]. The introduction of carbonate cement seems to be associated with the formation of micro-faults. Hence, the correspondence between the micro-fault orientation and the trend in cluster 3 is observed. The faulting can cause micro-fracturing of the material, thus promoting the formation of elongated voids, usually in accordance with the faulting direction. As result, a space for cement

crystallisation is formed and thus the presence of mutually connected grains can be the result of this process (Figure 13).



**Figure 13.** The evolution of the Krosno Beds from the Nasiczne site decoded from the studied sample.

## 6. Summary and Conclusions

The presented protocol allows for the complex 3D characterization of granular geomaterials microstructure utilizing 3D microCT images, image analysis and multivariate statistics. The protocol consists of image processing analysis and the multivariate analysis of its results. The outcome of the image analysis is a detailed description of grains with 44 unique parameters describing compositional properties, size, shape, morphology, and spatial arrangement of each grain. By the application of multivariate statistics and data exploration, an understanding of the characteristics of the grain can be obtained. Based on the parameters, cluster analysis is conducted which classifies the objects based on their similarity, allowing for characteristics of groups to be distinguished within a data set. The methodological approach can be applied to any other rock type and other geomaterials. The proposed protocol is useful in deciphering the nature of geological processes recorded in rocks, as well as being used for the explanation of the anisotropy of material properties. The detailed numerical characterization of the microstructure enables the tracking of even the gentlest trends within geomaterials. The application of freely available software for the analysis increases its availability and broadens the audience. The attached protocol with the included code can be reused in the presented form or modified by means of the implementation of proprietary solutions and thus promotes the reproducibility of the research.

The results confirm the presence of deformational microstructures in the studied sample related to both deformations. The grains were clustered into groups representing different compositional and geometrical aspects of the granular framework. With the use of supplementary microscopical studies, the origin of grains was possible to be determined and the sequence of recorded processes recognized. Firstly, the grains in the sample underwent deformation and reorientation leading to a steepening of the grains and their compact packing, with the formation of compaction bands. Subsequently, micro-faults were formed and this was associated with carbonate cement intrusion leading to the replacement of some grains and their connection, probably by micro-fracturing of the material during faulting.

Summing up, two major conclusions concerning the studied material can be presented: (1) the grains are pervasively and evenly distributed within the analysed sample and show compact packing and, (2) the spatial arrangement of grains is well organized and show directional trends referring to the orientation of compaction bands and micro-faults.

**Author Contributions:** Conceptualization, P.J.S.; Formal analysis, P.J.S. and K.K.; Investigation, P.J.S.; Methodology, P.J.S., J.T. and S.W.; Resources, P.J.S., A.Ś., J.T. and S.W.; Visualization, P.J.S.; Writing—original draft, P.J.S.; Writing—review & editing, A.Ś., K.K. and A.F.; Funding acquisition, P.J.S., A.Ś. and A.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Science Centre, Poland (grant number: 2018/31/N/ST10/02486) and supported by AGH statutory funds (subsidy number: 16.16.140.315).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data set and the codes included in appendices are also available on GitHub (<https://github.com/piotrstrzelecki/materials>, accessed on 5 June 2021). The other data presented in this study are available on request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. The Protocol of Image Processing and Analysis in Fiji

ImageJ version 1.53c Java 1.8.0\_66 (64-bit)

Additional plugins required: MorphoLibJ, 3D image suite, BaSiC

Legend:

> - next step

\* - operations in a pop-up window

# - a comment

1. Launch Fiji Software
2. File > Open # open a micro-CT image - read into RAM (File > Import > TIFF Virtual Stack – read an image directly from drive storage)

### Image processing

3. Image > Type > 8-bit # convert image to 8-bit type
4. Image > Crop # resize and crop the image
5. Image > Adjust > Brightness/Contrast # adjust brightness and contrast
6. Plugins > BaSiC # apply a X-ray attenuation correction
7. File > Save As > Tiff # save the corrected image
8. Image > Adjust > Threshold > \* (check stack histogram) > Apply > \* (uncheck calculate threshold for each image) # choose and extract chosen components to a binary image format at appropriate threshold level
9. Plugins > 3D > 3D Fast Filters > \* median # apply 3D median filter on the binary image
10. Plugins > MorphoLibJ > Kill Borders # remove objects at the edges of the image
11. Plugins > MorphoLibJ > Binary Images > Size opening 2D/3D # remove objects below chosen size
12. Plugins > 3D > 3D Simple Segmentation # assign ID to each separate object
13. File > Save As > Tiff # save the segmented image

### Image analysis

14. Plugins > MorphoLibJ > Analyze > Analyze Regions 3D > \* File > Save As > regions.csv # perform and save the measurements; the segmented image is required
15. Plugins > MorphoLibJ > Analyze > Equivalent Ellipsoid > \* (eigenvectors table) File > Save As > eigenvectors.csv # perform and save the measurements; opened the segmented image is required
16. Plugins > MorphoLibJ > Analyze > Geodesic Diameter 3D > \* File > Save As > geodesic.csv # perform and save the measurements; the segmented and corrected images are required
17. Plugins > MorphoLibJ > Analyze > Intensity Measurements 2D/3D > \* File > Save As > intensity.csv # perform and save measurements, opened the segmented and corrected images are required
18. Plugins > 3D > 3D Manager > \* Settings > \* (in the measurements section check all but Convex Hull) > OK > Add Image > Select All > Measure 3D \* File > Save As > measure.csv # perform and save the measurements; the segmented image is required

### Data visualization

19. Plugins > MorphoLibJ > Binary Images > Assign Measure to Label # assign values from a table to the objects in the segmented image i.e., cluster id. (Open > clusters.csv)
20. Plugins > 3D Viewer # visualize the image in 3D

## Appendix B. The Code for Raw Data Processing and Data Set Preparation

```

# 1. Required packages and functions -----
# Install libraries; remove '#' to execute the line
# install.packages("DescTools")

# Load libraries
library(DescTools)
# Load functions
RotationMatrix <- function(x, y, z) {
  # Creates rotation matrix for 3D vectors in Euclidean space
  # x - rotation around x axis in degrees
  # y - rotation around y axis in degrees
  # z - rotation around z axis in degrees
  x <- -x / 180 * pi
  y <- -y / 180 * pi
  z <- -z / 180 * pi
  R <- cbind(c(1, 0, 0), c(0, cos(x), sin(x)), c(0, -sin(x), cos(x))) %*%
    cbind(c(cos(y), 0, -sin(y)), c(0, 1, 0), c(sin(y), 0, cos(y))) %*%
    cbind(c(cos(z), sin(z), 0), c(-sin(z), cos(z), 0), c(0, 0, 1))
  return(R)
}

ImageJtoGeo <- function(x, y, z) {
  # Function for coordinate system conversion:
  # ImageJ (Cartesian) to Geographical (spherical) coordinates
  # x - the x component of a 3D vector
  # y - the y component of a 3D vector
  # z - the z component of a 3D vector
  temp <- as.data.frame(CartToSph(x, y, z, up = T))
  temp[, 2:3] <- temp[, 2:3] * 180 / pi
  for (i in 1:nrow(temp)) {
    if (temp$theta[i] < 0) {
      if (temp$phi[i] < 90) {
        temp$phi[i] <- temp$phi[i] + 270
      }
      else {
        temp$phi[i] <- temp$phi[i] - 90
      }
    }
    else {
      temp$phi[i] <- temp$phi[i] + 90
    }
  }
  temp$theta <- abs(temp$theta)
  return(temp[, 2:3])
}

# 2. Load data -----
# Set your working directory
setwd("C:/Folder/...") # COMPLETE!

# Load raw data
Intensity <- read.csv("intensity.csv", header = T, sep = ",", dec = ".")
Region <- read.csv("regions.csv", header = T, sep = ",", dec = ".")
Geodesic <- read.csv("geodesic.csv", header = T, sep = ",", dec = ".")
Vectors <- read.csv("eigenvectors.csv", header = T, sep = ",", dec = ".")
Measure <- read.csv("measure.csv", header = T, sep = ",", dec = ".")

# Set a resolution of the measurements. To preserve pixel units set to 1
resolution <- 1 # e.g., 1 pix = 0.012 mm

# 3. Prepare a final data set -----
# (OPTIONAL) Rotation of components.
# Relevant if different position of the sample orientation is required
RZ <- RotationMatrix(0, 0, 0) # Rotation around z axis in ImageJ
RY <- RotationMatrix(0, 0, 0) # Rotation around y axis in ImageJ
RX <- RotationMatrix(0, 0, 0) # Rotation around x axis in ImageJ
# Rotation
Vectors[, 2:4] <- as.matrix(Vectors[, 2:4]) %*% RZ %*% RY %*% RX # rotate semi-axis a
Vectors[, 5:7] <- as.matrix(Vectors[, 5:7]) %*% RZ %*% RY %*% RX # rotate semi-axis b
Vectors[, 8:10] <- as.matrix(Vectors[, 8:10]) %*% RZ %*% RY %*% RX # rotate semi-axis c

# Prepare the final data set (parameters in pixel units)
data <- data.frame(row.names = 1:nrow(Intensity))[1:nrow(Intensity), ]
{
  data$`mean intensity` <- Intensity$Mean
}

```



```

data$`intensity sd` <- Intensity$StdDev
data$`max intensity` <- Intensity$Max
data$`min intensity` <- Intensity$Min
data$`median intensity` <- Intensity$Median
data$`mode intensity` <- Intensity$Mode
data$`intensity skewness` <- Intensity$Skewness
data$`intensity kurtosis` <- Intensity$Kurtosis
data$`volume object` <- Measure$Vol..pix.
data$`volume ellipsoid` <- Measure$Vol..pix. / Measure$RatioVolEllipsoid
data$`volume box` <- Measure$VolBounding..pix.
data$`volume ball` <- Geodesic$Radius^3 * 4 / 3 * pi
data$`surface area` <- Measure$SurfCorr..pix.
data$`Feret diameter` <- Measure$Feret..unit. / resolution
data$`mean breadth` <- Region$MeanBreadth / resolution
data$`geodesic diameter` <- Geodesic$Geod..Diam.
data$`ball radius` <- Geodesic$Radius
data$`semi-axis a` <- Region$Elli.R1 / resolution
data$`semi-axis b` <- Region$Elli.R2 / resolution
data$`semi-axis c` <- Region$Elli.R3 / resolution
data$`mean radius` <- replace(Measure$DCMean..unit., is.na(Measure$DCMean..unit.), 0) / resolution
data$`radius sd` <- replace(Measure$DCSD..unit., is.na(Measure$DCSD..unit.), 0) / resolution
data$`max radius` <- replace(Measure$DCMax..unit., is.na(Measure$DCMax..unit.), 0) / resolution
data$`min radius` <- replace(Measure$DCMin..unit., is.na(Measure$DCMin..unit.), 0) / resolution
data$`compactness` <- Measure$CompCorr..pix.
data$`discrete compactness` <- Measure$CompDiscrete
data$`sphericity` <- Measure$SpherCorr..pix.
data$`Euler number` <- Region$EulerNumber
data$`geodesic elongation` <- Geodesic$Geod..Elong.
data$`semi-axes ratio a/b` <- Region$Elli.R1.R2
data$`semi-axes ratio a/c` <- Region$Elli.R1.R3
data$`semi-axes ratio b/c` <- Region$Elli.R2.R3
data$`object vol./ellipsoid vol. ratio` <- Measure$RatioVolEllipsoid
data$`object vol./box vol. ratio` <- Measure$RatioVolbox
data$`ball vol./object vol. ratio` <- Geodesic$Radius^3 * 4 / 3 * pi / Measure$Vol..pix.
data$`x coordinate` <- Measure$CX..pix.
data$`y coordinate` <- Measure$CY..pix.
data$`z coordinate` <- Measure$CZ..pix.
data$`semi-axis a trend` <- ImageJtoGeo(Vectors[, 2], Vectors[, 3], Vectors[, 4])[, 2]
data$`semi-axis a plunge` <- ImageJtoGeo(Vectors[, 2], Vectors[, 3], Vectors[, 4])[, 1]
data$`semi-axis b trend` <- ImageJtoGeo(Vectors[, 5], Vectors[, 6], Vectors[, 7])[, 2]
data$`semi-axis b plunge` <- ImageJtoGeo(Vectors[, 5], Vectors[, 6], Vectors[, 7])[, 1]
data$`semi-axis c trend` <- ImageJtoGeo(Vectors[, 8], Vectors[, 9], Vectors[, 10])[, 2]
data$`semi-axis c plunge` <- ImageJtoGeo(Vectors[, 8], Vectors[, 9], Vectors[, 10])[, 1]
}

# Save the data set in pixel units
write.csv(data, "dataset_pixel.csv", row.names = F)

# Save the data set in the desired unit
resolution <- 0.012
unit <- "mm"

# Prepare the final data set (parameters in desired units)
data[, 9:12] <- data[, 9:12] * resolution^3
data[, 13] <- data[, 13] * resolution^2
data[, c(14:24, 36:38)] <- data[, c(14:24, 36:38)] * resolution

# Save the data set in mm units
write.csv(data, paste0("dataset_", unit, ".csv"), row.names = F)

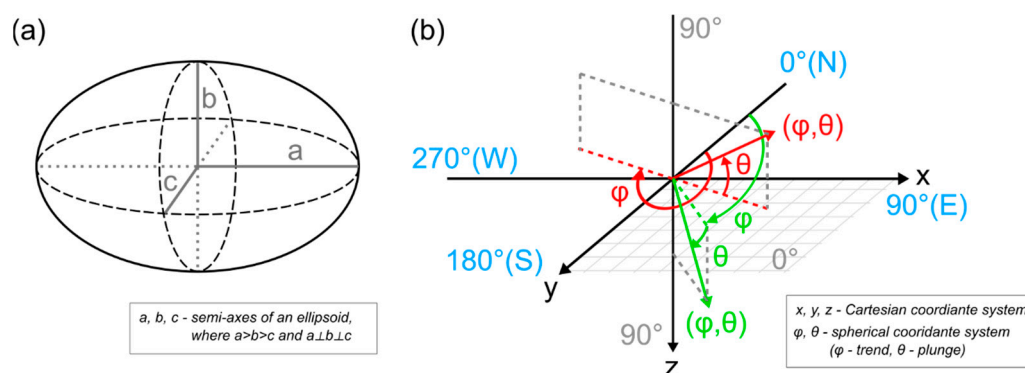
# Clear the workspace
rm(list = ls())
gc()

```

## Appendix C. Data Description and Manipulation

**Table A1.** The description of the object parameters resulting from the image analysis based on [45,46] and modified by the authors.

Category	No.	Parameter	Description
Chemical Composition	1	mean intensity	mean of voxel values in the object
	2	intensity SD	standard deviations of voxel values in the object
	3	max intensity	maximum of voxel value in the object
	4	min intensity	minimum of voxel value in the object
	5	median intensity	median of voxel values in the object
	6	mode intensity	mode of voxel values in the object
	7	intensity skewness	the skewness of voxel values distribution in the object
	8	intensity kurtosis	kurtosis of voxel values distribution in the object
Size	9	object volume	the volume of the object
	10	ellipsoid volume	the volume of the ellipsoid fitted to the object
	11	box volume	the volume of the bounding box encompassing the object
	12	ball volume	the volume of the largest ball inscribed in the object
	13	surface area	the sum of the surface area of contouring voxel faces
	14	Feret diameter	the largest calibrated distance between two contour voxels in the object
	15	mean breadth	the mean of the Feret diameter measured in all directions of the object
	16	geodesic diameter	the length of the longest geodesic path within the object
	17	ball radius	the radius of the largest inscribed ball
	18	semi-axis a	the length of the longest semi-axis of the ellipsoid
	19	semi-axis b	the length of the intermediate semi-axis of the ellipsoid
	20	semi-axis c	the length of the shortest semi-axis of the ellipsoid
	21	mean radius	mean distance from the geometrical centre of the object to the surface
	22	radius SD	the standard deviation of distance from the geometrical centre of the object to surface
Shape	23	max radius	maximum distance from the geometrical centre of the object to the surface
	24	min radius	minimum distance from the geometrical centre of the object to the surface
	25	compactness	the normalized ratio between the surface and the volume of the object
	26	discrete compactness	the measure of compactness for porous and fragmented objects
	27	sphericity	basing on the face-connectivity of voxels within the object
	28	Euler number	root square of the compactness
	29	geodesic elongation	the number of connected fragments of the object, minus the number of holes in the object, plus the number of bubbles within it
	30	semi-axes ratio a/b	the ratio between the geodesic diameter and the diameter of the largest inscribed ball
	31	semi-axes ratio a/c	the ratio between the length of the largest and medium semi-axis of the ellipsoid
	32	semi-axes ratio b/c	the ratio between the lengths of the largest and smallest semi-axis of the ellipsoid
	33	object vol./ellipsoid vol. ratio	the ratio between the length of the medium and smallest semi-axis of the ellipsoid
	34	object vol./box vol. ratio	the ratio between the volume of the object and the ellipsoid
	35	ball vol./object vol. ratio	the ratio between the volume of the object and the bounding box
			the ratio between the volume of the ball and the object
Spatial Arrangement and Orientation	36	x coordinate	x coordinate of the geometrical centre for each object
	37	y coordinate	y coordinate of the geometrical centre for each object
	38	z coordinate	z coordinate of the geometrical centre for each object
	39	semi-axis a trend	the direction of the ellipsoid's semi-axis a on the xy plane measured clockwise from the north
	40	semi-axis a plunge	the vertical angle, measured downwards, between the xy plane and the ellipsoid's semi-axis a
	41	semi-axis b trend	the direction of the ellipsoid's semi-axis b on the xy plane measured clockwise from the north
	42	semi-axis b plunge	the vertical angle, measured downwards, between the xy plane and the ellipsoid's semi-axis b
	43	semi-axis c trend	the direction of the ellipsoid's semi-axis c on the xy plane measured clockwise from the north
	44	semi-axis c plunge	the vertical angle, measured downwards, between the xy plane and the ellipsoid's semi-axis c



**Figure A1.** (a) An ellipsoid and dependency between its semi-axes; (b) The rule of coordinates transformation from the Cartesian coordinate system in Fiji software to the spherical coordinate system of vectors representing semi-axes of the ellipsoids.

**Table A2.** The libraries utilised for data manipulation, visualization and analysis in R.

Library Name and Version	Usage	Reference
base_4.0.3	data manipulation	[47]
openair_2.8-1	data visualization	[48]
ggpmisc_0.3.7	data manipulation	[49]
cluster_2.1.0	cluster analysis	[50]
viridis_0.5.1	data visualization	[51]
Morpho_2.8	data manipulation	[52]
ggrepel_0.9.0	data visualization	[53]
gridExtra_2.3	data visualization	[54]
tidyverse_1.3.0	data manipulation	[55]
reshape2_1.4.4	data manipulation	[56]
factoextra_1.0.7	cluster analysis	[57]
qgraph_1.6.5	data visualization	[58]
ggplot2_3.3.3	data visualization	[59]
corrplot_0.84	data visualization	[60]
psych_2.0.12	data exploration	[61]
DescTools_0.99.39	data manipulation	[62]

## Appendix D. The Code for Data Set Analysis

```
# 1. Required packages and functions -----
# Install libraries; remove '#' to execute the line
# install.packages("dplyr")
# install.packages("psych")
# install.packages("qgraph")
# install.packages("corrplot")
# install.packages("ggplot2")
# install.packages("factoextra")
# install.packages("ggrepel")
# install.packages("tidyverse")
# install.packages("viridis")
# install.packages("plotly")
# install.packages("reshape")
# install.packages("Morpho")
# install.packages("cluster")
# install.packages("gridExtra")
# install.packages("ggpmisc")
# install.packages("openair")

# Load libraries
library(psych)
```

```

library(corrplot)
library(ggplot2)
library(qgraph)
library(factoextra)
library(reshape2)
library(tidyverse)
library(gridExtra)
library(ggrepel)
library(Morpho)
library(viridis)
library(cluster)
library(ggpmisc)
library(openair)

# 2. Load data -----
# Set your working directory
setwd("C:/Folder/...") # COMPLETE!
data <- read.csv("https://raw.githubusercontent.com/piotrstrzelecki/materials/main/dataset_mm.csv",
header = T, sep = ",", dec = ".", check.names = F) # load the data
res <- 0.012 # set the resolution
unit <- "mm" # set the unit

## 2.1 Column names preparation
# Prepare the names of the variables: 3 columns including original names, numerical id, names with units
{
  coln <- cbind(original = colnames(data), id = c(1:ncol(data)), ggplot = c(1:ncol(data)))
  coln[1:6, 3] <- paste0(coln[1:6, 1], " (pixel value)")
  coln[9:12, 3] <- paste0(coln[9:12, 1], " (", unit, "^3)")
  coln[13, 3] <- paste0(coln[13, 1], " (", unit, "^2)")
  coln[c(14:24, 36:38), 3] <- paste0(coln[c(14:24, 36:38), 1], " (", unit, ")")
  coln[c(7:8, 25:35), 3] <- paste0(coln[c(7:8, 25:35), 1], " (over(phantom(0)))")
  coln[39:44, 3] <- paste0(coln[39:44, 1], " (degree)")
  coln[, 3] <- gsub(" ", "~", coln[, 3])
}

# 3.0 The data set overview - basic statistics -----
# Basic statistics
describeBy(data)

# "min intensity" shows a zero variance. It was excluded from further analyses
data <- data[, -4]
coln <- coln[-4, ]

# 4.0 Correlation between variables -----

colnames(data) <- coln[, 2] # switch to numerical id of variables names
# Figure 5a
corrplot(cor(data), method = "color", tl.col = "black") # corrplot::
# Figure 5b
qgraph(cor(data), posCol = "blue3", negCol = "brown3", minimum = 0, cut = 1, vsize = 5,
legend.cex = .5, label.scale.equal = T, legend = T, borders = T, palette = "ggplot2",
Glratio = 5, groups = list(`chemical composition` = 1:7, `size` = 8:23, `shape` = 24:34,
`spatial arrangement` = 35:43)) # qgraph::

# 4.1 Principal component analysis -----
# variables from the spatial arrangement category were excluded
data.pca <- prcomp(data[, 1:34], center = T, scale. = T)
summary(data.pca)

# 4.2 Number of principal components determination -----
# Figure 6a
fviz_eig(data.pca, addlabels = T, ggtheme = theme_classic()) # factoextra::

# 4.3 Factor analysis -----

data.fa <- principal(data[, 1:34], rotate = "none", nfactors = 4, scores = TRUE) #psych::

# data manipulation
fa.loadings <- data.fa$loadings[1:34, ]
fa.loadings <- data.frame(fa.loadings)
fa.loadings$var <- factor(rownames(fa.loadings), levels = rev(rownames(fa.loadings)))
loadings.m <- melt(fa.loadings, id.vars = "var")

# Figure 6b
ggplot(loadings.m, aes(x = value, y = var, fill = value)) +
  facet_wrap(~variable, nrow = 1, scales = "fixed") + # place the factors in separate facets
  geom_bar(stat = "identity") +

```



```

scale_fill_gradient2(name = "Loading", high = "blue", mid = "white", low = "red",
midpoint = 0, guide = F) +
theme_bw(base_size = 10) +
xlab(label = "Standardized loadings") +
ylab(label = "")

# statistics for the first and second PCs
# data manipulation
data.fa <- principal(data[, 1:34], rotate = "none", nfactors = 2, scores = TRUE)
fa.loadings <- data.frame(data.fa$uniquenesses)
colnames(fa.loadings) <- "uniqueness"
fa.loadings$communality <- data.fa$communality
fa.loadings$complexity <- data.fa$complexity
fa.loadings$var <- factor(rownames(fa.loadings), levels = rev(rownames(fa.loadings)))

# Figure 6c
loadings.m <- melt(fa.loadings, id.vars = "var")
ggplot(loadings.m, aes(x = value, y = var, fill = variable)) +
geom_bar(stat = "identity", fill = "grey") +
facet_wrap(~variable, nrow = 1, scales = "free_x") +
theme_bw() +
theme(axis.title.y = element_blank())

# 4.4 Data in PC coordinates -----

pca.xy <- data.frame(data.fa$scores[, 1:2])

# Sample of the data set for visual assessment
set.seed(1)
sample <- pca.xy[mcNNindex(target = as.matrix(pca.xy), query = as.matrix(kmeans(pca.xy,
centers = 10, nstart = 25, iter.max = 100)$centers), k = 1), ]

# Figure 7a
ggplot(pca.xy, aes(x = PC1, y = PC2)) +
geom_bin2d(bins = 200) +
scale_fill_continuous(type = "viridis") +
geom_point(sample, mapping = aes(x = PC1, y = PC2), size = 1, color = "red") +
geom_label_repel(sample[order(sample[, 1]), 1:2], mapping = aes(x = PC1, y = PC2,
label = 1:nrow(sample)), max.overlaps = 15, box.padding = 0.4, segment.color = "red") +
theme(legend.position = c(0.8, 0.3))

# 5.0 Cluster analysis -----
# 5.1 Optimal number of clusters determination

# Figure 8a
fviz_nbclust(scale(data[, 1:34]), FUNcluster = clara,
diss = dist(scale(data[, 1:34]), method = "manhattan"), method = "gap_stat")

# 5.2 Data clustering

clust <- clara(scale(data[, 1:34]), k = 3, metric = "manhattan", samples = 50, sampsize = 500)

# Figure 9a
ggplot() +
geom_point(pca.xy, mapping = aes(x = PC1, y = PC2, color = as.factor(clust$clustering))) +
labs(col = "cluster") + xlim(-2, 10) + ylim(-2.5, 5) + theme(legend.position = c(0.9, 0.2))

write.csv(clust$clustering, "clusters.csv", row.names = FALSE)

## 5.3 Quality of the clustering - silhouette statistics

# Figure 8b
clust <- pam(scale(data[, 1:34]), k = length(clust$i.med),
metric = "manhattan", medoids = clust$i.med, do.swap = F)
fviz_silhouette(clust)

# 6.0 Microstructural assessment -----
# 6.1 Chemical composition, size and shape of components in clusters

data.ggplot <- data[, 1:34]
colnames(data.ggplot) <- coln[1:34, 3]
data.ggplot$cluster <- as.factor(clust$clustering)
data.ggplot <- melt(data.ggplot, id = "cluster")

# Figure 10
ggplot(data.ggplot, aes(x = cluster, y = value, fill = cluster)) +
geom_violin() +

```

```

geom_boxplot(width=0.2,outlier.shape = NA, coef = 0,fill = "white") +
scale_y_continuous(trans = "log10") +
facet_wrap(~variable, ncol = 5, scale = "free", labeller = label_parsed) +
theme(legend.position = c(0.9, 0.05), axis.title.x = element_blank(),
axis.text.x = element_blank(),axis.ticks.x = element_blank())

# 6.2 Spatial arrangement of components

# Radial distance calculation
colnames(data) <- coln[, 2]
rad.dist <- data.frame(sqrt((data$`36` - ceiling(max(data$`36`)) / 2)^2 +
(data$`37` - ceiling(max(data$`37`)) / 2)^2 +
(data$`38` - ceiling(max(data$`38`)) / 2)^2))
rad.dist$cluster <- as.factor(clust$clustering)

# Radial distribution function g(r)
dr <- 15 * res #
gr <- matrix(ncol = length(unique(clust$clustering)), nrow = ceiling(max(data$`38` / 2 / res))

for (i in 1:length(unique(clust$clustering))) {
  rd <- rad.dist[clust$clustering == i, 1]
  for (j in 1:nrow(gr)) {
    gr[j, i] <- length(rd[rd > (j * res) & rd < (j * res + dr)]) / (4 * pi * (j * res)^2 * dr)
  }
  rm(list = "rd")
}

# Smoothed radial distribution function g(r) calculation
gr.sm <- gr
for (i in 1:ncol(gr)) {
  gr.sm[, i] <- smooth.spline(gr[, i], spar = 0.4)$y
}

# Periodicity of peaks in g(r) determination
gr.his <- list()
for (i in 1:(ncol(gr.sm))) {
  gr.his[[i]] <- ((1:nrow(gr)) * res)[ggpmisc::find_peaks(gr.sm[, i])]
  for (j in seq_along(gr.his[[i]])) {
    gr.his[[i]][j] <- gr.his[[i]][j + 1] - gr.his[[i]][j]
  }
}
for (i in 1:length(gr.his)) {
  names(gr.his)[i] <- paste0("cluster ", i)
}

# Visualization and data manipulation
gr.vis <- data.frame(gr)
for (i in 1:ncol(gr.vis)) {
  colnames(gr.vis)[i] <- paste0("cluster ", i)
}
gr.vis$x <- (1:nrow(gr)) * res
gr.vis <- melt(gr.vis, id = "x")
gr.vis$smooth <- melt(gr.sm)$value
gr.his <- melt(gr.his)

# Figure 11
r1 <- ggplot(gr.vis[gr.vis$x > 0.7, ]) +
  geom_line(mapping = aes(y = value, x = x, color = variable)) +
  theme(legend.position = "none") +
  labs(y = "g(r)") +
  labs(x = paste0("r ", "( ", unit, " )")) +
  facet_wrap(~variable, ncol = 1, scale = "free_y")
r2 <- ggplot(gr.vis[gr.vis$x > 0.7, ]) +
  geom_line(mapping = aes(y = smooth, x = x, color = variable)) +
  stat_peaks(mapping = aes(y = smooth, x = x), span = 11) +
  theme(legend.position = "none") +
  labs(x = paste0("r ", "( ", unit, " )"), y = "smoothed g(r) & peaks") +
  facet_wrap(~variable, ncol = 1, scale = "free_y")
r3 <- ggplot(gr.his, aes(y = L1, x = value, colour = L1)) +
  geom_boxplot() +
  theme(legend.position = "none",axis.text.y = element_blank(),
axis.ticks.y = element_blank()) +
  labs(x = paste0("distance between successive peaks ", "( ", unit, " )"), y = "peaks distribution") +
  facet_wrap(~L1, ncol = 1, scales = "free_y")
grid.arrange(r1, r2, r3, ncol = 3)

## 6.3 Spatial orientation of components ----

```

```

# data manipulation
data.dir <- data[, 39:44]
colnames(data.dir) <- coln[39:44, 1]
data.dir$cluster <- as.factor(paste0("cluster ", clust$clustering))

# Figure 12
windRose(data.dir,
  type = "cluster", ws = "semi-axis a plunge", wd = "semi-axis a trend",
  layout = c(nlevels(data.dir$cluster), 1), angle = 10,
  annotate = F, breaks = seq(0, 90, 15), paddle = F,
  grid.line = 1, offset = 0, cols = rev("inferno")(4)[2:4]),
  border = 8, key.header = "semi-axis a plunge (degrees)",
  key.footer = "", key.position = "right")

windRose(data.dir,
  type = "cluster", ws = "semi-axis b plunge", wd = "semi-axis b trend",
  layout = c(nlevels(data.dir$cluster), 1), angle = 10,
  annotate = F, breaks = seq(0, 90, 15), paddle = F,
  grid.line = 1, offset = 0, cols = rev("inferno")(4)[2:4]),
  border = 8, key.header = "semi-axis b plunge (degrees)",
  key.footer = "", key.position = "right")

windRose(data.dir,
  type = "cluster", ws = "semi-axis c plunge", wd = "semi-axis c trend",
  layout = c(nlevels(data.dir$cluster), 1), angle = 10,
  annotate = F, breaks = seq(0, 90, 15), paddle = F,
  grid.line = 1, offset = 0, cols = rev("inferno")(4)[2:4]),
  border = 8, key.header = "semi-axis c plunge (degrees)",
  key.footer = "", key.position = "right")

```

## Appendix E. The Protocol Validation on a Standard Sample

The validity of measurements and computations was verified on a standard sample. The sample was built in the Fiji software (the macro for the standard sample creation in Fiji is available at <https://github.com/piotrstrzelecki/materials>, accessed on 5 June 2021) and consists of 180 artificially generated grains of known properties in the image of  $700 \times 700 \times 600$  pixel in size. Each grain's centre is equidistant (100 pixels) from the closest neighbours. The sample is composed of five layers and each layer consists of 36 grains. Within each layer, grains have the same shape. The uppermost layer consists of spherical grains. Toward the bottom of the sample, the shape of grains changes and grains acquire more ellipsoidal up to discoidal shapes. The lowermost layer consists of rod-shaped grains (Figure A2a). Each object has unique properties. The properties gradually change with increasing x, y and z coordinates of objects. With increasing y coordinate, the length of objects' semi-axis a increases from 5 to 45 pixels and its trend increases from 0 to  $330^\circ$  in a  $60^\circ$  step. With increasing x coordinate, the plunge of grains' semi-axis a increases from 0 to  $90^\circ$ .

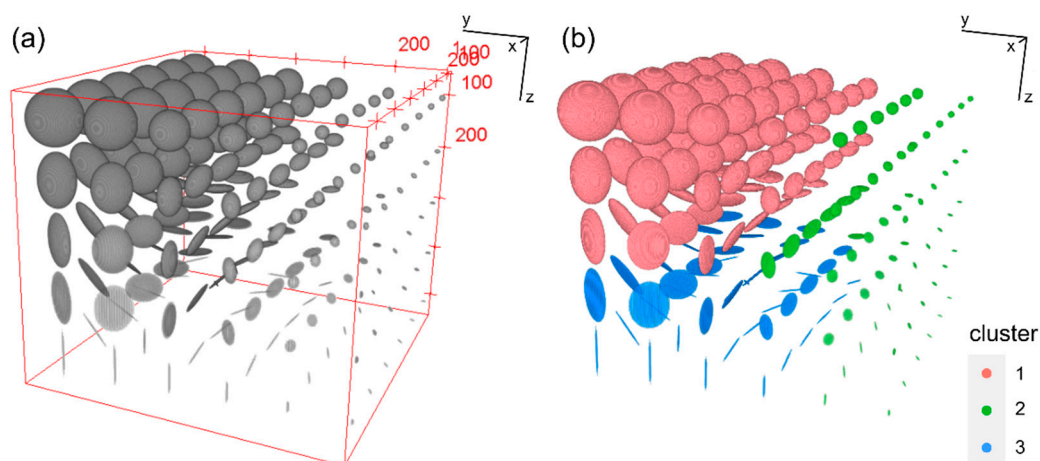
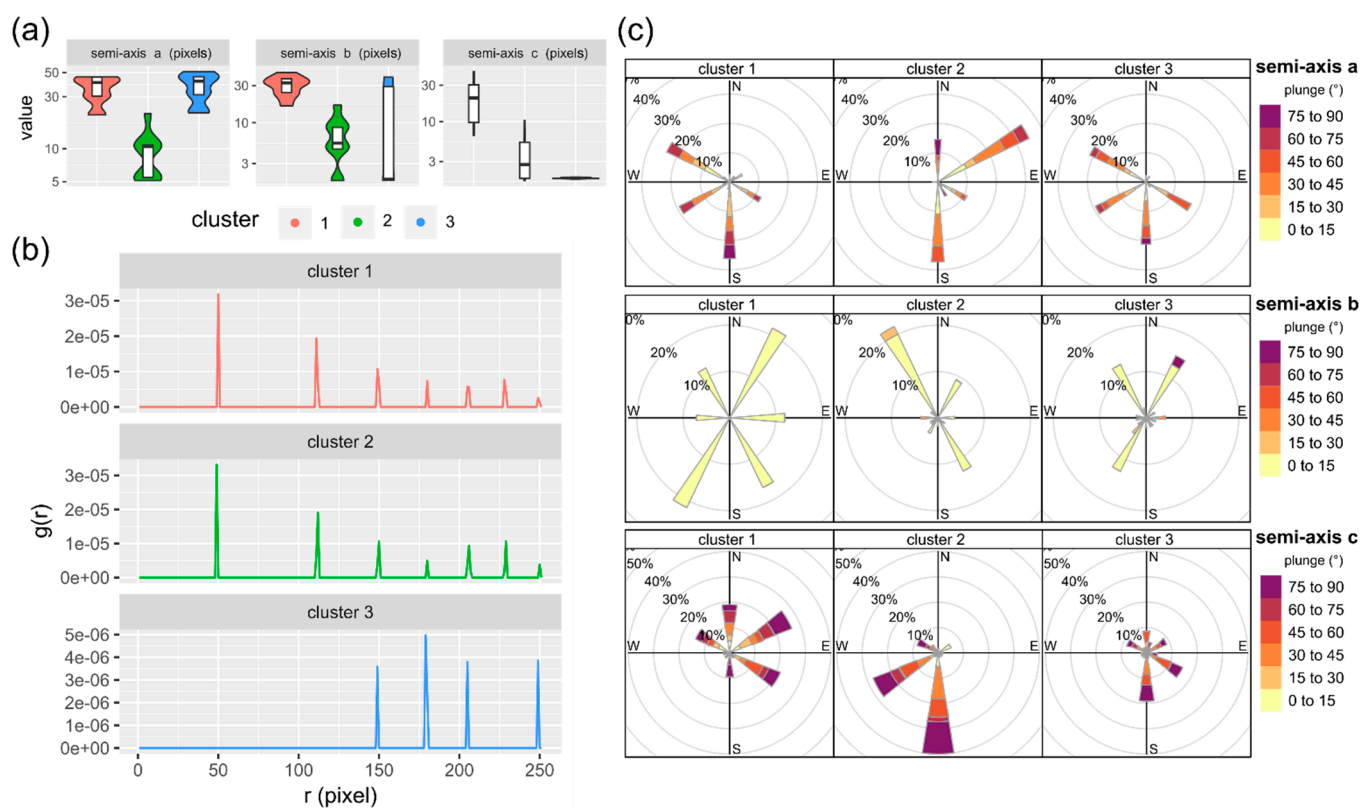


Figure A2. 3D visualization of the standard sample (a) and clustered grains (b).

The artificial grains were clustered into 3 groups based on the parameters resulting from image analysis and calculations (Figure A2b). The first cluster is composed of the biggest grains of spherical and ellipsoidal shapes (Figure A3a; the highest values of all semi-axes), whereas the second one consists of the smallest grains of regular shape (Figure A3a; the lowest values of semi-axes a and b). The third cluster is composed of discoidal and rod-shaped grains. shape (Figure A3a; the lowest values of semi-axes c). The grains belonging to the first and second clusters are present near the geometrical centre of the sample, whereas the grains of the third cluster are distant from the sample's centre (Figure A2b, Figure A3b). The trends of semi-axis a in the first and second clusters are similar as the grains within these clusters occupy similar xy position in the sample and have larger angles in comparison to the grains in the second cluster, which are localised closer to the origin of the xy coordinate system (Figure A2c). The plunge angles are diversified in all clusters and show values from 0 to 90°. The expected properties find confirmation in the numerical data, hence confirming the correctness of the calculations.



**Figure A3.** (a) Distribution of the selected parameters; (b) Radial distribution function  $g(r)$  calculated for the geometrical centres of grains; (c) Spatial orientation of the grains in the clusters.

## References

1. Lindqvist, J.E.; Åkesson, U.; Malaga, K. Microstructure and functional properties of rock materials. *Mater. Charact.* **2007**, *58*, 1183–1188. [\[CrossRef\]](#)
2. Strzelecki, P.J.; Fheed, A.; Radzik, N.; Świerczewska, A. X-ray microtomography and digital image analysis application to a petrological characterization of carbonates. *Przegląd Geol.* **2019**, *67*, 195–197. [\[CrossRef\]](#)
3. Promentilla, M.A.B.; Cortez, S.M.; Papel, R.A.D.C.; Tablada, B.M.; Sugiyama, T. Evaluation of microstructure and transport properties of deteriorated cementitious materials from their X-ray computed tomography (CT) images. *Materials* **2016**, *9*, 388. [\[CrossRef\]](#)
4. Rajczakowska, M.; Stefaniuk, D.; Łydzba, D. Microstructure Characterization by Means of X-ray Micro-CT and Nanoindentation Measurements. *Studia Geotech. Mech.* **2015**, *37*, 75–84. [\[CrossRef\]](#)
5. Schmith, J.; Höskuldsson, Á.; Holm, P.M. Grain shape of basaltic ash populations: Implications for fragmentation. *Bull. Volcanol.* **2017**, *79*. [\[CrossRef\]](#)



6. Bishwal, R.M. Potential use of R-statistical programming in the field of geoscience. In Proceedings of the 2017 2nd International Conference for Convergence in Technology (I2CT), Mumbai, India, 7–9 April 2017; IEEE: New York, NY, USA, 2017; pp. 979–982. [\[CrossRef\]](#)
7. Rabbani, A.; Jamshidi, S.; Salehi, S. An automated simple algorithm for realistic pore network extraction from micro-tomography images. *J. Pet. Sci. Eng.* **2014**, *123*, 164–171. [\[CrossRef\]](#)
8. Strzelecki, P.J.; Fheed, A.; Raczynski, P. 3D geometric evaluation of porosity types in carbonates. In *AIP Conference Proceedings*; AIP Publishing LLC: Melville, NY, USA, 2020; p. 030002. [\[CrossRef\]](#)
9. Cheng, G.; Yue, Q. Rock Images Analysis of FCM Clustering Algorithm Based on Weighted Color Texture Features. *J. Phys. Conf. Ser.* **2018**, *1*, 012185. [\[CrossRef\]](#)
10. Lai, J.; Lortie, C.J.; Muenchen, R.A.; Yang, J.; Ma, K. Evaluating the popularity of R in ecology. *Ecosphere* **2019**, *10*, e02567. [\[CrossRef\]](#)
11. Royo, H. Programming Tools: Adventures with R. *Nature* **2015**, *517*, 109–110.
12. Hydzik-Wiśniewska, J.; Pękala, A. The evaluation of the physico-mechanical properties of selected carpathian sandstones in terms of their use as a armourstone. *Arch. Min. Sci.* **2019**, *64*, 65–77. [\[CrossRef\]](#)
13. Karnkowski, P. *Oil and Gas Deposits in Poland*; Geosynoptics Society: Cracow, Poland, 1999.
14. Hurai, V.; Marko, F.; Tokarski, A.K.; Świerczewska, A.; Kotulová, J.; Biroň, A. Fluid inclusion evidence for deep burial of the Tertiary accretionary wedge of the Carpathians. *Terra Nova* **2006**, *18*, 440–446. [\[CrossRef\]](#)
15. Pszonka, J.; Wendorff, M. Carbonate cements and grains in submarine fan sandstones—The Cergowa Beds (Oligocene, Carpathians of Poland) recorded by cathodoluminescence. *Int. J. Earth Sci.* **2017**, *106*, 269–282. [\[CrossRef\]](#)
16. Jarmolowicz-Szulc, K.; Karwowski, Ł.; Marynowski, L. Fluid circulation and formation of minerals and bitumens in the sedimentary rocks of the Outer Carpathians—Based on studies on the quartz-calcite-organic matter association. *Mar. Pet. Geol.* **2012**, *32*, 138–158. [\[CrossRef\]](#)
17. Świerczewska, A.; Tokarski, A.K. Deformation bands and the history of folding in the Magura nappe, Western Outer Carpathians (Poland). *Tectonophysics* **1998**, *297*, 73–90. [\[CrossRef\]](#)
18. Strzelecki, P.J.; Tokarski, A.K.; Świerczewska, A. Structural evolution of the eastern part of the Silesian Nappe recorded in deformation bands, Polish segment of the Outer Carpathians. In Proceedings of the 16th Meeting of the Central European Tectonic Groups, Ryto, Poland, 18–21 April 2018; Tomaszczyk, M., Głuszyński, A., Sikora, R., Eds.; AGH University of Science and Technology: Cracow, Poland, 2018; Volume 44, p. 196. [\[CrossRef\]](#)
19. Feldkamp, L.A.; Davis, L.C.; Kress, J.W. Practical cone-beam algorithm. *J. Opt. Soc. Am. A* **1984**, *1*, 612. [\[CrossRef\]](#)
20. Bam, L.C.; Miller, J.A.; Becker, M. A mineral X-ray linear attenuation coefficient tool (MXLAC) to assess mineralogical differentiation for X-ray computed tomography scanning. *Minerals* **2020**, *10*, 441. [\[CrossRef\]](#)
21. Guntoro, P.I.; Ghorbani, Y.; Koch, P.H.; Rosenkranz, J. X-ray microcomputed tomography (μct) for mineral characterization: A review of data analysis methods. *Minerals* **2019**, *9*, 183. [\[CrossRef\]](#)
22. Schindelin, J.; Arganda-Carreras, I.; Frise, E.; Kaynig, V.; Longair, M.; Pietzsch, T.; Preibisch, S.; Rueden, C.; Saalfeld, S.; Schmid, B.; et al. Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **2012**, *9*, 676–682. [\[CrossRef\]](#)
23. Peng, T.; Thorn, K.; Schroeder, T.; Wang, L.; Theis, F.J.; Marr, C.; Navab, N. A BaSiC tool for background and shading correction of optical microscopy images. *Nat. Commun.* **2017**, *8*, 83–142. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Iassonov, P.; Gebrenegus, T.; Tuller, M. Segmentation of X-ray computed tomography images of porous materials: A crucial step for characterization and quantitative analysis of pore structures. *Water Resour. Res.* **2009**, *45*, 1–12. [\[CrossRef\]](#)
25. Ollion, J.; Cochenec, J.; Loll, F.; Escudé, C.; Boudier, T. TANGO: A generic tool for high-throughput 3D image analysis for studying nuclear organization. *Bioinformatics* **2013**, *29*, 1840–1841. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Legland, D.; Arganda-Carreras, I.; Andrey, P. MorphoLibJ: Integrated library and plugins for mathematical morphology with ImageJ. *Bioinformatics* **2016**, *32*, 3532–3534. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1901**, *2*, 559–572. [\[CrossRef\]](#)
28. Hotelling, H. Analysis of a complex of statistical variables into Principal Components. *J. Educ. Psychol.* **1933**, *24*, 417. [\[CrossRef\]](#)
29. Abdi, H.; Williams, L.J. Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 433–459. [\[CrossRef\]](#)
30. Spearman, C. ‘General Intelligence’, Objectively Determined and Measured. *Am. J. Psychol.* **1904**, *15*, 201–292. [\[CrossRef\]](#)
31. Cattell, R. The Scree Test for the number of factors. *Multivar. Behav. Res.* **1966**, *1*, 245–276. [\[CrossRef\]](#)
32. Hofmann, R.J. Indices Descriptive of Factor Complexity. *J. Gen. Psychol.* **1977**, *96*, 103–110. [\[CrossRef\]](#)
33. Kaufman, L.; Rousseeuw, P.J. *Clustering Large Data Sets*; Elsevier Science Publishers B.V.: Amsterdam, The Netherlands, 1986.
34. Kaufman, L.; Rousseeuw, P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1990; ISBN 3175723993.
35. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [\[CrossRef\]](#)
36. Hintze, J.L.; Nelson, R.D. Violin plots: A box plot-density trace synergism. *Am. Stat.* **1998**, *52*, 181–184. [\[CrossRef\]](#)
37. Scott, G.D. Radial Distribution of the Random Close Packing of Equal Spheres. *Nature* **1962**, *194*, 956–957. [\[CrossRef\]](#)
38. Hair, J.F.; Black, W.C.; Babin, B.; Anderson, R.E. *Multivariate Data Analysis*, 7th ed.; Pearson: New York, NY, USA, 2010; ISBN 9780138132637.

39. Arora, P.; Deepali; Varshney, S. Analysis of K-Means and K-Medoids Algorithm for Big Data. *Phys. Procedia* **2016**, *78*, 507–512. [CrossRef]
40. Aggarwal, C.C.; Hinneburg, A.; Keim, D.A. On the surprising behavior of distance metrics in high dimensional space. In Proceedings of the International Conference on Database Theory, London, UK, 4–6 January 2001; Springer: Berlin/Heidelberg, Germany, 2001; Volume 1973, pp. 420–434. [CrossRef]
41. Charrad, M.; Ghazzali, N.; Boiteau, V.; Niknafs, A. Nbclust: An R package for determining the relevant number of clusters in a data set. *J. Stat. Softw.* **2014**, *61*, 1–36. [CrossRef]
42. Theodoridis, S.; Koutroumbas, K. *Pattern Recognition*, 4th ed.; Academic Press: Cambridge, MA, USA, 2008; ISBN 9781597492720.
43. Pettijohn, F.J.; Potter, P.E.; Siever, R. *Sand and Sandstone*; Springer: New York, NY, USA, 1987; ISBN 978-0-387-96350-1.
44. Stachowska, A.; Łoziński, M.; Śmigielski, M.; Wysocka, A.; Jankowski, L.; Ziółkowski, P. Anisotropy of magnetic susceptibility as an indicator for palaeocurrent analysis in folded turbidites (Outer Western Carpathians, Poland). *Sedimentology* **2020**, *67*, 3783–3808. [CrossRef]
45. Legland, D.; Arganda-Carreras, I. MorphoLib User Manual. 2018. Available online: <https://github.com/ijpb/MorphoLib/releases/download/v1.4.0/MorphoLib-manual-v1.4.0.pdf> (accessed on 15 January 2021).
46. Cordelieres, F.P. The 3D Object Counter Plugin a.k.a 3D-OC. 2009. Available online: [https://imagejdocu.tudor.lu/\\_media/plugin/analysis/3d\\_object\\_counter/3d-oc.pdf](https://imagejdocu.tudor.lu/_media/plugin/analysis/3d_object_counter/3d-oc.pdf) (accessed on 15 January 2021).
47. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
48. Carslaw, D.C.; Ropkins, K. Openair—An r package for air quality data analysis. *Environ. Model. Softw.* **2012**, *27*, 52–61. [CrossRef]
49. Aphalo, P.J. *Ggpmisc: Miscellaneous Extensions to 'Ggplot2'*; R Package Version 0.3.7; R Foundation for Statistical Computing: Vienna, Austria, 2020.
50. Maechler, M.; Rousseeuw, P.; Struyf, A.; Hubert, M.; Hornik, K. *Cluster: Cluster Analysis Basics and Extensions*, R package version 2.1.0; R Foundation for Statistical Computing: Vienna, Austria, 2019.
51. Garnier, S. *Viridis: Default Color Maps from 'Matplotlib'*; R Package Version 0.5.1; R Foundation for Statistical Computing: Vienna, Austria, 2018.
52. Schlager, S. Morpho and Rvcg—Shape Analysis in {R}. In *Statistical Shape and Deformation Analysis*; Zheng, G., Li, S., Szekel, G., Eds.; Academic Press: Cambridge, MA, USA, 2017; pp. 217–256. ISBN 9780128104934.
53. Słowikowski, K. *Ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'*; R Package Version 0.9.0; R Foundation for Statistical Computing: Vienna, Austria, 2020.
54. Auguie, B. *Gridextra: Miscellaneous Functions for 'Grid' Graphics*; R Package Version 2.3; R Foundation for Statistical Computing: Vienna, Austria, 2017.
55. Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; et al. Welcome to the Tidyverse. *J. Open Source Softw.* **2019**, *4*, 1686. [CrossRef]
56. Wickham, H. Reshaping Data with the reshape Package. *J. Stat. Softw.* **2007**, *21*, 1–20. [CrossRef]
57. Kassambara, A.; Mundt, F. *Factoextra: Extract and Visualize the Results of Multivariate Data Analyses*; R Package Version 1.0.7; R Foundation for Statistical Computing: Vienna, Austria, 2020.
58. Epskamp, S.; Cramer, A.O.J.; Waldorp, L.J.; Schmittmann, V.D.; Borsboom, D. Qgraph: Network visualizations of relationships in psychometric data. *J. Stat. Softw.* **2012**, *48*. [CrossRef]
59. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016.
60. Wei, T.; Simko, V. *R Package 'Corrplot': Visualization of a Correlation Matrix*; R Package Version 0.84; R Foundation for Statistical Computing: Vienna, Austria, 2017.
61. Revelle, W. *Psych: Procedures for Psychological, Psychometric, and Personality Research*; R Package Version 2.0.12; R Foundation for Statistical Computing: Vienna, Austria, 2020.
62. Andri, S.; Aho, K.; Alfons, A.; Anderegg, N.; Aragon, T.; Arachchige, C.; Arppe, A.; Baddeley, A.; Barton, K.; Bolker, B.; et al. *DescTools: Tools for Descriptive Statistics*; R Package Version 0.99.39; R Foundation for Statistical Computing: Vienna, Austria, 2020.