https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

Quantitative comparison of three-dimensional bodies using geometrical properties to validate the dissimilarity of a standard collection of 3D geomodels

Friedrich Carl[1], Peter Achtziger-Zupančič [2], Jian Yang[3], Marlise Colling Cassel[4], Florian Wellmann[3]

[1]Chair of Engineering Geology and Hydrogeology, RWTH Aachen University, Lochnerstr. 4-20, 52064 Aachen, Germany
[2]Fraunhofer Research Institution for Energy Infrastructures and Geotechnologies IEG, Kockerellstr. 17, 52062 Aachen, Germany
[3]Chair of Computational Geoscience, Geothermics and Reservoir Geophysics, RWTH Aachen University, Mathieustr. 30, 52074 Aachen, Germany
[4]Chair of Geology and Sedimentary Systems and Geological Institute, RWTH Aachen University, Wüllnerstr. 2, 52062 Aachen, Germany

Correspondence to: Friedrich Carl (carl@lih.rwth-aachen.de)

Abstract. The quantification of 3D structural shapes is a central goal across multiple scientific disciplines, serving purposes such as image analysis and the precise geometric characterization of objects. This study proposes a methodology for the shape quantification based on a set of geometrical parameters in 2D sections of 3D geological shapes and establishes a set of synthetic regular geometries as benchmark models in 3D geomodeling approaches. The proposed methodology is demonstrated on a number of simple geometric bodies and the benchmark models to assess their geometrical dis-/similarity. The dimensions of the structures are measured perpendicular and vertically to their horizontal main axes on a fixed amount of cross sections. Furthermore, gradient and curvature measurements on these cross sections are conducted. A subsequent multi-step data analysis provides insight into the main geometrical characteristics of the structures and visualizes differences between various datasets: Analysis of extension measurements reveals the anisotropy of structures, the existence of overhangs and the character of the top surface of an investigated structure. Analyzing the gradients and curvatures offers information on the slopes of the lateral walls of the structure and its sphericity as well as top surface. Kullback-Leibler divergence is utilized to quantitatively compare individual parameter distributions. Dimensionally reduced cluster analysis groups and systematizes input structures based on the combined statistical parameters and serves for the identification of benchmark models showing large geometrical similarity. It is expected that the methodology and set of benchmark models will aid in advances to model and compare subsurface structures based on sparse data.

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

1 Introduction

1.1 Previous studies

The quantitative comparison of three-dimensional (3D) objects plays a crucial role in various scientific fields, including geology, computer science and engineering (Laga et al., 2019). Accurate shape quantification independent of the objects' orientation is essential for applications such as geological modeling, resource management and structural analysis, where understanding the geometric properties of objects can inform decision making and enhance predictive capabilities. In computer science, numerous methodologies using 2D cross sections have emerged. For instance, Celenk (1995) describes a method involving the alignment of equally-spaced cross sections in two objects via the computation of their respective horizontal main axes. Subsequently, sections are compared through the computation of an averaged shape difference in four directions along the main axes. Recent studies on 3D object recognition using 2D sections apply neural networks: Dumitru et al. (2022) feed 2D input into Convolutional Neural Networks, while Dumitru & Gorgan (2023) enhance this concept by relying on Vision Transformer-based Neural Networks.

The geometrical parameters determined in our study have been applied in variable extent before to describe the shape of objects. Gradient data has been utilized various times: Goh & Chan (2003), for example, propose a shape descriptor derived from the gradient vector field of shapes in binary images. Meanwhile, Ettl et al. (2007) present a method to reconstruct object shape by spatial integration of gradient data, and Çapar et al. (2009) define two gradient-based shape descriptors being applicable to binary and grayscale images. Curvature data have been employed for example by Canul-Ku et al. (2019), proposing two 3D shape descriptors based on a Multi-View approach of curvature features to classify archaeological artifacts. Meanwhile, Mousa (2011) defines a geometric descriptor based on the principle curvature distribution around the surface of objects and Muzahid et al. (2021) approach 3D object recognition by using the principal curvature directions of 3D objects as geometric inputs for a 3D Convolutional Neural Network. In contrast to gradients and curvatures, the direct use of measurements of extent/distance has been less common (e.g. Novotni & Klein, 2001).

In general, many comparative approaches for 3D bodies are predominantly used in medical fields, where (automatic) polygon comparison and pattern recognition play crucial roles in diagnostics: Meyer-Baese & Schmid (2014) provide an extensive overview on methodologies used, for example specialized neural networks or fuzzy clustering algorithms. These methods are also being applied in various ways within geological and geophysical studies (e.g. Demicco & Klir, 2003; Hillier et al., 2021; Liu et al., 2023; Sun & Li, 2015). An alternative way to describe and compare 3D bodies is to use 3D Fourier analysis, as shown by many studies from various scientific fields: Kindratenko et al. (1996) apply it to describe the shape of particles from microscope images. Meanwhile, Li et al. (2000) have studied the irregularity of graphite nodules in cast iron and Drevin (2006) has used Fourier transform results to determine the sphericity of particles. Another approach to the problem of shape comparison is the 3D shape based object class recognition directly from point cloud data, as applied for example by Wohlkinger & Vincze (2011).

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

Aside from the approaches mentioned above, advances for the comparison of geological models are mainly related to uncertainty assessment and quantification within geological models. Schweizer et al. (2017) apply the Jaccard distance and the normalized city-block distance as measures for model dissimilarity, while Suzuki et al. (2008) propose the usage of the Hausdorff distance for the same purpose. Lindsay et al. (2013) developed an approach for model comparison having some similarities to our contribution: To determine the similarities between 101 realizations of a 3D composite geomodel based on the same (perturbed) input dataset, a set of geometrical "geodiversity" parameters (e.g. formation depth, volume, contact surface curvature) was calculated on all stratigraphic units of the case study. The resulting datasets were then analyzed in their ranges to determine endmember model realizations. Furthermore, principal component analysis was employed to determine which geometrical characteristics contribute most to spatial uncertainty and to detect realization outliers for the combined geodiversity metrics.

Despite these advancements in various directions, distinct gaps remain in current research. Many existing methodologies focus on shape characterization for image recognition, while few concentrate on the exact geometrical characterization for robust comparison of 3D structures. This study aims to address this challenge by proposing a novel methodology for the quantitative description, comparison and systematization of explicitly modelled 3D geometries using a set of geometrical parameters. The algorithm is applied to a set of 36 3D geometries approximating subsurface structures of varying rock types, intended to act as benchmark models in geomodeling approaches. By demonstrating the quantification algorithm on these 3D bodies called "standard geometries", their geometrical dis-/similarity is analyzed. Furthermore, the methodology has been applied to a small set of basic 3D geometries (a cube, an ellipsoid, a prism, a pyramid and a sphere) with distinctive and expected divergence of geometrical properties. In what follows, an extension of the concept of "standard geometries" initially described by Carl et al. (2023) is developed as a geometrical systematization to collect and catalogue subsurface geometries of the potential host rocks in the German site selection for a nuclear waste repository (halite rock, claystone and crystalline rocks).

Claystones and shales are clastic sedimentary rocks composed of at least 50 % particles smaller than 4 µm (e.g. Picard, 1971). Since all clastic sediments are initially deposited conformably onto the underlying strata (Selley, 2000), the most typical geometry of claystone/shale sediments is conformable layering (see Fig. 1, upper section). Nonetheless, the appearance of these conformable layers can vary considerably: tilting and folding can result in a range of geometries varying from a flat layered appearance that remain generally conformable. By contrast, faulting, erosion and folding can produce unconformable geometries. Salt rock (i.e., halite) is initially deposited conformably as an evaporitic sediment. Beyond the undeformed, concordant, flat-layer geometry, halite structures are mainly categorized according to two principles: The most common classification is based on the question whether a structure remained concordant in respect to its overlying rocks or intruded into its overburden (Hudec & Jackson, 2007; see Fig. 1, middle section). Following this systematization, salt anticlines, pillows and rollers are categorized as concordant, while salt stocks, sheets and walls are intrusive bodies. In addition, a supplementary subdivision based on the length-to-width ratio of salt bodies is discussed by some authors (e.g., Hudec et al., 2011): Structures showing a length-to-width ratio higher than 2 in map view (thus being considerably

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

anisotropic) are being defined as anticlines or walls, respectively. In contrast, rather isotropic geometries with a length-to-width ratio smaller than 2 are the pillows, stocks and, at least in their early evolutionary stages, sheets. An additional aspect to consider when classifying salt structures is whether the halite is allochthonous or autochthonous. Sheets are the only structural type categorizable as the former: If the bulb of a stock or wall is subhorizontally oriented or moderately dipping above the autochthonous salt source layer, this rock body can be defined as a salt sheet (Hudec and Jackson, 2006). Crystalline rocks considered are both plutonic and high-grade metamorphic rocks (migmatites and gneisses). As the high-grade metamorphic rocks originate from a wide array of protoliths, resulting in diverse geometries, the establishment of a single, coherent classification for both groups is difficult. For instance, orthogneisses and some migmatites originate from plutonic protoliths such as granitoids and exhibit structural characteristics similar to their igneous predecessors. By contrast, paragneisses and the remaining migmatites derive from various sedimentary sources. Their current shape depends not only on the geometry of the original rock body but also on the specific mineral assemblage of the protolith and the pressure-temperature conditions experienced during metamorphism. Overall, most high-grade metamorphic rock bodies in the German subsurface are bounded by either plutonic intrusions or fault zones. Consequently, for the purposes of our geometric approximation, we treat them as discordant rock volumes of varying shape and size (see Fig. 1, lower section). For plutonic rocks, our classification combines the shape of the bodies with their relationship to the overlying strata (concordance or discordance) (Philpotts & Ague, 2009; see Fig. 1, lower section). Among discordant bodies with varying shape, two size-based categories are distinguished: batholiths (exceeding 100 km² in areal extent) and stocks (smaller than 100 km²). Additionally, cylindrical discordant bodies, mainly representing feeder pipes for ascending magma, are recognized. Moreover, two kinds of tabular geometries can be distinguished: discordant dikes and predominantly concordant sills. Beyond these, three concordant geometries are noted: laccoliths (characterized by a roughly flat base and a convex roof), lopoliths (defined by a roughly flat top and a shallow convex base), and phacoliths (lens-shaped bodies lacking any flat boundaries).

Building on these classifications, a collection of geometrical end members (standard geometries) that approximate the shape variations of the rock types was set up by Carl et al. (2023). The geometries are intended to act as open source benchmark models for structural geomodeling, as realistic geological models depend on a clear definition of the rock type and the 3D geometries of evaluated rocks. In its initial form, each of the geometrical end members per potential host rock type was represented by a single version of a 3D body. However, as a large share of these initial end members can be represented by a multitude of possible regular geometrical representations, we designed alternative realizations after reviewing literature: Subsurface salt structures have been created after Hudec & Jackson (2007), Hudec et al. (2011) and Jackson & Talbot (1991), claystone geometries have been inspired by Selley (2000) and Nichols (2009), and crystalline rock geometries are based on Markl (2015) and Winter (2013). Additional inspiration was drawn from studying open source 3D models of real subsurface structures (Dutch subsurface models from TNO, available at https://www.dinoloket.nl/en/subsurface-models/map, and Australian subsurface models from Geoscience Australia, available at https://portal.ga.gov.au/3d). The standard geometries were created in blender (https://www.blender.org/) and are visible in Fig. 2. Some standard geometries

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

are non-unique for rock types but can be used in different environments, e.g. stocks/batholiths for salt and crystalline intrusions. This is indicated in the model titles, as in these cases, the names of different structures are separated by an underscore. Blanks in model names are replaced with a period, and in brackets, additional geometrical information are given in some cases, such as the lateral character of the top of a structure (e.g. hourglass-shape) or the roundness of the top surface (rounded or flattened).

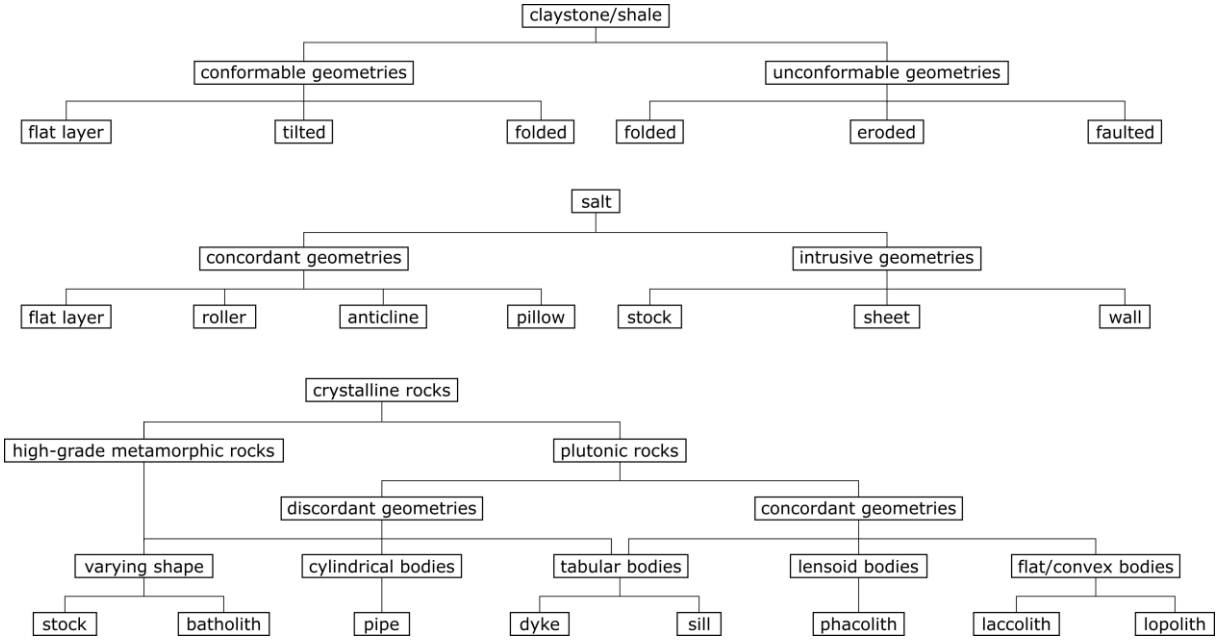


Fig. 1: Geometrical systematization of the rock types considered for the establishment of the catalogue of benchmark models ("standard geometries").

1.2 Content and motivation of this study

Our proposed methodology allows for the quantitative description, comparison and systematization of explicitly modelled structures using a set of geometrical parameters. The horizontal and vertical dimensions as well as gradients and curvatures of 3D geometries are measured on vertical cross sections oriented perpendicular to the two horizontal main axes of the structures. The resulting datasets of these parameters are analyzed statistically, providing insight into the main geometrical characteristics of the input structures: the data analysis yield information about the anisotropy of structures, the potential existence of overhangs, the sphericity and the character of the lateral walls as well as top surface of evaluated structures. Furthermore, clustering is used to systematize the datasets based on the measured parameters. The setup of cross sections perpendicular to the main axes ensures, that the input structures are covered thoroughly with regular-spaced measurements

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

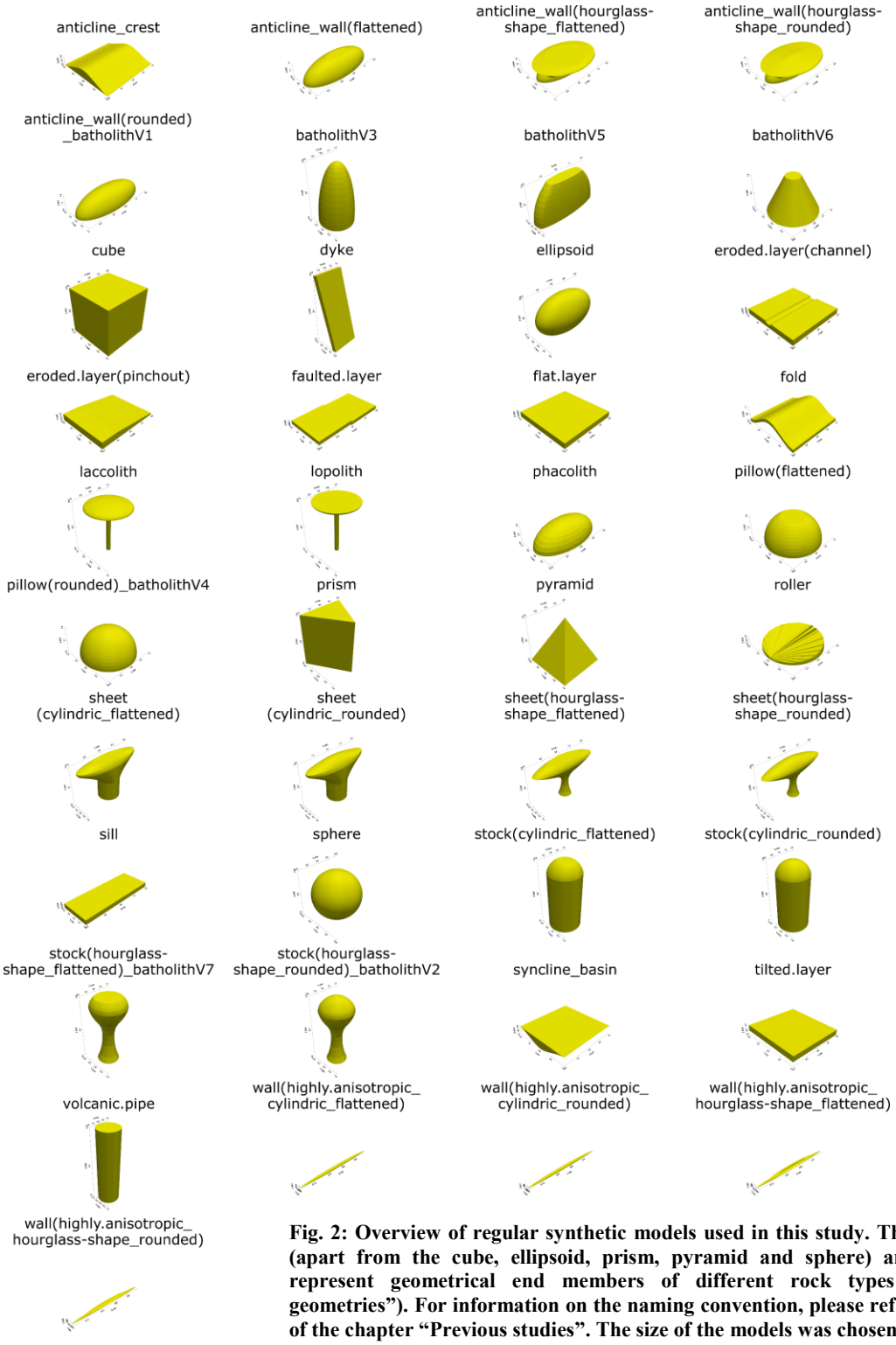


Fig. 2: Overview of regular synthetic models used in this study. The structures (apart from the cube, ellipsoid, prism, pyramid and sphere) are meant to represent geometrical end members of different rock types ("standard geometries"). For information on the naming convention, please refer to the end of the chapter "Previous studies". The size of the models was chosen arbitrarily

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

that follow the 3D contours of the respective geometry. Our method cannot be used to quantitatively compare implicit representations of structures.

The method is applied to the standard geometries established as well in this study, that will be publicly available as benchmarks for structural geomodeling. Applying the method to these models serves two purposes: While the functionality of the approach is validated, we also assess quantitatively, whether the developed benchmark models are geometrically dissimilar or whether some of them can be consolidated for their purpose. We created our quantification method in its simplistic, approximating form in spite of the availability of numerous other methodological approaches for the quantified comparison of 3D structures. The algorithm can reproduce the main geometrical characteristics of input datasets fast but also enhances the interpretability of results, making them accessible to a broader audience. The significance of the benchmark models lies in their ability to facilitate the validation and comparison of different methodologies within geomodeling approaches.

The paper is structured as follows: Sect. 2 outlines the methodology employed in this study, detailing the developed segmentation and measurement algorithm. Sect. 3 presents the results of applying the methodology to the benchmark models and a single subsurface dataset, while Sect. 4 discusses the implications of these findings in the context of existing research. Finally, Sect. 5 concludes with future research directions.

2 Methods

2.1 Segmentation and measurement algorithm

For our approach, we aimed at a high grade of automation and easy integration in a model analysis process. The method requires the dataset to either be a mesh with extractable vertices or a data frame of vertex coordinates themselves (the input file format has to be changed if it is not .vtk).

To retrieve characteristic statistics, a geometrical segmentation algorithm (see e.g. Shamir, 2008) has been established, which first rasters the 3D model into 22 equidistant cross sections with the normal direction parallel to the longer horizontal axis of the mesh´s bounding box. As measurements are conducted perpendicular to the two horizontal main axes of the structures, two sets of cross sections need to be determined separately. Orientation of sections normal to the longitudinal axis of the structure (first direction) have been determined by a minimization of the cross-sectional area (Stephenson, 2018; Fig. 3, Part 1). The cross sections normal to the first set are set up by rastering the established sections vertically, then first connecting raster lines of consecutive sections and lastly the resulting segments (Fig. 3, Part 3). After their respective setup, the cross sections of both directions are corrected automatically and/or manually for artifacts (Fig. 3, Part 2). Extensional measurements are conducted on each cross section at 5 equidistant transects (Fig. 3, Part 4). Since the very first and last cross section of both directions are excluded from the measurements as they would (undesirably) slice irregular polygons several times, 20 intervals are considered for every input structure. This results in 100 measurements being conducted respectively for each of the two horizontal parameters as well as 200 values for the vertical extent. Please note, that the

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

assumption that a cross section of the first set is perpendicular to the longitudinal horizontal main axis only applies to the center point of the given section. The same limitation applies to a given cross sectional segment (trapezoidal segment) of an orthogonal section and the secondary horizontal main axis.

In addition to the extensional measurements, gradient and curvature calculations are carried out (see Fig. 3, Part 4). Both parameters are determined on all cross sections between consecutive vertices of a cross section. The curvature in 2D is defined as the reciprocal of the circumradius of a triangle. Therefore, it is calculated between three consecutive vertices in either the xz- or yz-plane, by first determining the side lengths (a, b and c) of the triangle between the points, then the semiperimeter of the triangle and the area through Heron´s formula, before calculating the curvature as the reciprocal of the circumradius of the triangle through:

$$curvature = \frac{4 \times \text{area}}{\text{a} \times \text{b} \times \text{c}} \quad (1)$$

The selected method measuring the lateral extents of input meshes normal to their horizontal main axes is advantageous over approaches analyzing an input body using parallel sections as applied in various medical imaging techniques like e.g. MRI (see e.g. Meyer-Baese & Schmid, 2014). Such an approach would have resulted in a dissimilar amount of output measurements for the two horizontal extents for many input structures as well as for different structures overall, both in case of a uniform grid for all datasets as well as an individual regular grid per dataset. Only the usage of an anisotropic grid, depending on the bounds of the input mesh, would have resulted in an equal amount of measurements per horizontal direction. However, using a supplementary grid would have generally resulted in the problem, that irregular structures would have often been cut several times along a horizontal measuring line. This would have created subordinate polygons that are completely disconnected from each other.

In contrast, covering every input structure with a constant number of measurements as also applied similarly by Celenk (1995) comes with an advantage and a disadvantage: while it ensures that the quantification of input datasets with our method is scale-independent as datasets of different structures have the same amount of data, the geometrical spatial variability of larger bodies might not be captured equally well as the shape of smaller ones. The potential impact of this matter is currently being analyzed in a follow-up study that applies the methodology to a database of over 300 structural models of subsurface structures from various geological settings. The question whether structures shall be represented by equal or dissimilar data quantities also concerns the gradient and curvature data: Orthogonal sections are created from a set of 19 trapezoidal segments (i.e. 40 vertices), while cross sections in the first direction are based on a varying, most often higher number of vertices. As gradients and curvatures are being calculated between neighboring vertices, the potentially larger edges between vertices in the orthogonal sections lead to a less-rounded appearance of the cross sections, directly affecting the values of both parameters.

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

```
Part 1:
Input: initial cross sections
Output: cross sections rotated perpendicular to longer horizontal main axis by minimization of cross section area
for each initial section do
    for each of 38 rotation steps (0° to 180° in 5° increments) do
        Compute rotation angle (theta)
        Apply rotation matrix to original normal ([1,0,0] or [0,1,0]) to get rotated normal
    end for
    for each rotated normal do
        Slice mesh using rotated normal and center point of section
        Retrieve vertices from rotated slice
        Project cross section into YZ plane to calculate area
        Compute centroid of this projected section
        Sort points by angle relative to centroid
        Compute area of polygon using sorted points and shoelace formula
    end for
end for

Part 2:
Input: rotated cross sections
Output: rotated cross sections after artifact correction
Step 1:
for each rotated cross section do
    set vertex with lowest z-value as vertex index=0
    perform normalized nearest neighbour algorithm
end for

Step 2:
for each section after normalized nearest neighbour algorithm do
    apply correction criterion
    if correction criterion = True then
        initialize manual vertex order correction in plotly.dash
        for each section with artifacts do
            correct vertex order by clicking on previous (correct) vertex then incorrect vertex
        end for
    end if
end for

Part 3:
Input: corrected cross sections of 1st direction
Output: corrected cross sections of orthogonal direction
for each section of 1st direction do
    raster section vertically into 22 vertical lines
    for each vertical line do
        retrieve X, Y and Zmin+Zmax - coordinates
    end for
end for

for 2 consecutive sections of 1st direction do
    for all vertical lines in both sections do
        extract X, Y and Zmin+Zmax (=2 points per section)
        combine 4 points into trapezoidal segment
    end for
end for

for every index of vertical lines do
    combine trapezoidal segments to assemble uncorrected orthogonal section
end for

for uncorrected cross sections of orthogonal direction
    repeat Part 2

Part 4:
Input: all cross sections
Output: horizontal and vertical dimensional measurements
for each cross section of 1st direction except index 0 & 21 do
    rotate & project section onto YZ plane
    create 5 horizontal and vertical measurement transects
    measure horizontal and vertical dimensions between intersections of transect and polygon
end for

for each cross section of orthogonal direction except index 0 & 21 do
    rescale sections (corresponds to rotation & projection onto YZ plane)
    create 5 horizontal and vertical measurement transects
    measure horizontal and vertical dimensions between intersections of transect and polygon
end for

for all cross sections do
    compute gradients
    compute curvatures
end for
```

Fig. 3: Pseudocode of the algorithm that creates the cross sections of both directions and measures the dimensional extents, gradients and curvatures on these sections. For further information, see chapter 2

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

2.2 Data analysis

The individual geometrical measurements were combined into a database and analyzed by the first five statistical moments and visualized as histograms and cumulative distribution functions (CDF´s). Comparative analyses of data distributions and a cluster analysis were carried out on the measured data, to demonstrate that 3D bodies can be quantitatively compared based on the statistical distributions of geometrical properties and to assess their dis-/similarity.

Semi-quantitative comparison of histograms was done for the statistical data, analyzing the vertical extension measures, combined horizontal extension measurements, the gradients and the curvatures. For gradient data, the frequency of infinite values was counted separately, since they represent vertical segments between two consecutive vertices. As those values cannot be plotted together with the remaining data as a separate bin, their frequency was visualized as a horizontal line. For gradients and curvatures, overflow bins were established: for the gradient data at the 5^{th} and 95^{th} percentile and for the curvatures only at the 95^{th} percentile. This aimed at facilitating the interpretability of the histograms, since for most datasets, a small percentage of values (<5%) was considerably larger than the rest, thereby spreading the measurements to a large number of additional histogram bins. The Kullback–Leibler divergence (Kullback and Leibler, 1951) was calculated on normalized data between the individual distributions of the geometrical parameters of the different input models, for quantification of the similarity between the structures. Cluster analysis followed data normalization to a range of -1 to 1 and principal component analysis (PCA; see Jolliffe, 2002). As variables (“features”) for PCA, 20 percentiles of the probability density functions (PDF´s) of the combined horizontal data, vertical data, gradients and curvatures were chosen. As the first two principle components only explained 40% of the variance, a matrix plot for the principal components 1 to 12 was created, to cover 90% of the variance. The number of clusters used in the K-means clustering algorithm was determined using an elbow plot and the silhouette score.

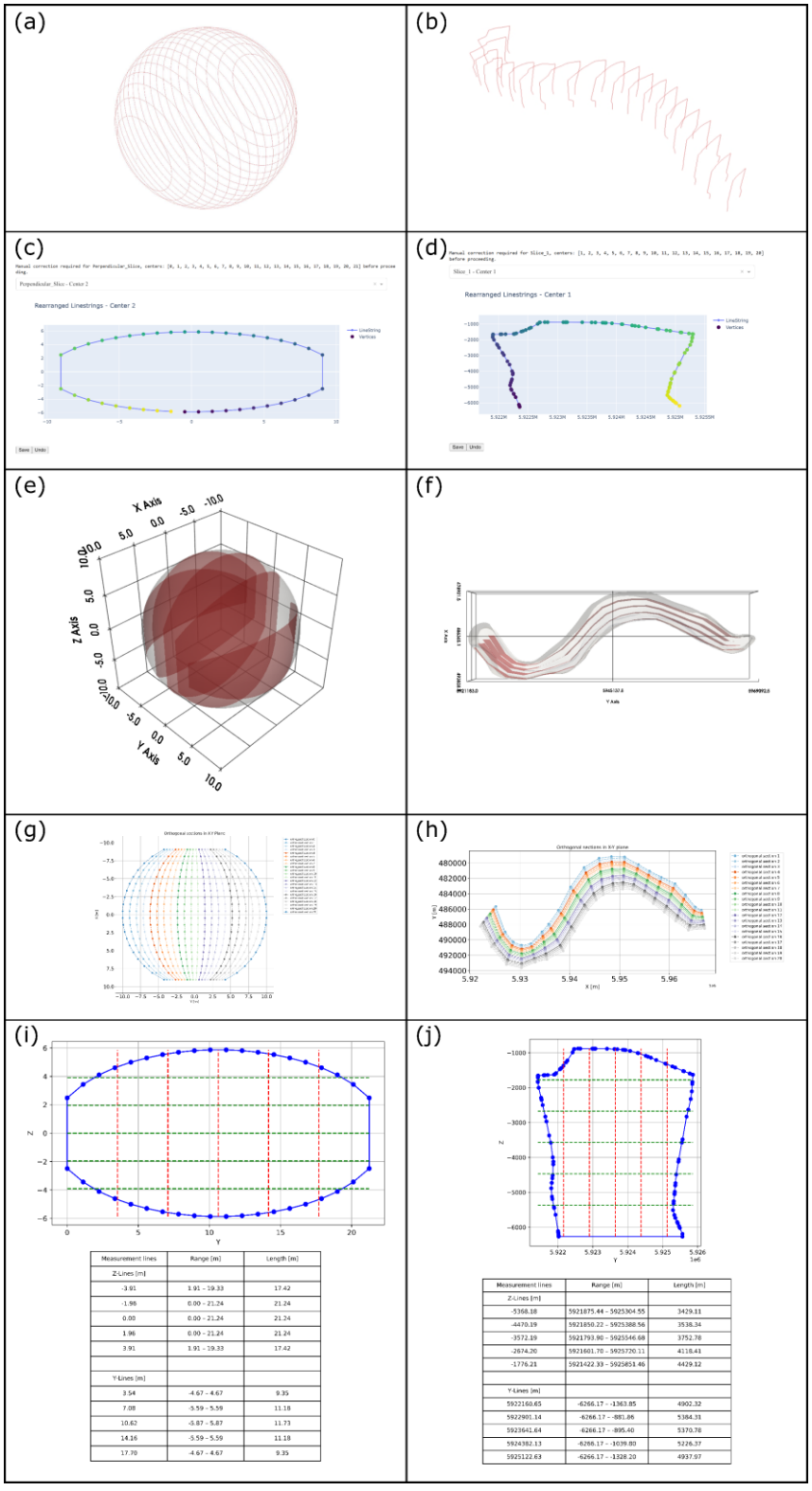
3 Results

Results of the segmentation and measurement algorithm as well as the data analysis are demonstrated using a sphere and a real, explicitly modeled subsurface salt body: the intrusive structure “Altenbruch-Beverstedt” from Lower Saxony, Germany, taken from BGR et al. (2022). Subsequently, the results of the cluster analysis are presented.

3.1 Segmentation and measurement algorithm

The initial subdivision of the input mesh (Fig. 4a & b) is followed by the stepwise rotation of the initial cross sections. The respective rotation step showing the minimal cross-sectional area is optimally oriented normal to the longitudinal main axis of the structure (first direction). Optimal orientation of all sections of the first direction of the sphere corresponds to 0° rotation, unlike when running the algorithm on an irregular mesh like Altenbruch-Beverstedt. This is the case due to the regularity and symmetry of all test models of this study. After subsequent artifact correction (Fig. 4c & d), the second set of cross sections is assembled from trapezoidal segments (for illustration, a subset of sections is shown in Fig. 4e & f).

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025



Measurement lines	Range (m)	Length (m)
Z-Lines (m)		
-3.91	1.91 - 19.33	17.42
-1.98	0.00 - 21.24	21.24
0.00	0.00 - 21.24	21.24
1.96	0.00 - 21.24	21.24
3.91	1.91 - 19.33	17.42
Y-Lines (m)		
3.54	-4.67 - 4.67	9.35
7.08	-5.59 - 5.59	11.18
10.62	-5.87 - 5.87	11.73
14.16	-5.59 - 5.59	11.18
17.70	-4.67 - 4.67	9.35

Measurement lines	Range (m)	Length (m)
Z-Lines (m)		
-5368.18	5921875.44 – 5925304.55	3429.11
-4470.39	5921850.22 – 5925388.56	3538.34
-3572.19	5921793.90 – 5925546.68	3752.78
-2674.20	5921601.70 – 5925720.11	4118.41
-1776.21	5921422.33 – 5925851.46	4429.12
Y-Lines (m)		
5922160.65	-6266.17 – -1363.85	4902.32
5922901.14	-6266.17 – -881.86	5384.31
5923641.64	-6266.17 – -895.60	5370.78
5924382.13	-6266.17 – -1039.80	5226.37
5925122.63	-6266.17 – -1328.20	4937.97

Fig. 4: Visual representation of the segmentation and measurement algorithm for a sphere model (left column) and the German intrusive salt structure "Altenbruch -Beverstedt" (right).
a & b) initial segmentation of the input meshes.
c & d) Plotly.dash app for vertex-order correction.
e & f) Orthogonal cross sections.
g & h) Coverage of input structure with cross sections (top view)
i & j) Example of extensional measurement results

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

Following potential artifact corrections of the orthogonal sections, both sets of cross sections are finalized (Fig. 4g & h) and extensional measurements as well as gradient and curvature calculations are carried out (Fig. 4i & j). The computational power required by the algorithm is low (runtime without varying artifact corrections: ca 60 s, using as the CPU an AMD Ryzen 7 PRO 5850U at max. 50% capacity at 3.5 GHz speed and the integrated GPU at 0.6 GB usage).

3.2 Data analysis

The results of the first analysis step, the first five statistical moments per parameter and the data visualized as histograms and CDF´s, are seen in Table 1 and Fig. 5, respectively. The size of the sphere was chosen arbitrarily, as the subsequent Kullback-Leibler divergence and cluster analysis are based on normalized data. Both the statistical moments for the sphere and the distributions in Fig. 5 (left column) reveal differences for the three parameters, although individual extents should be the same in all three dimensions, if a sphere would be measured equally in all directions. This is due to compromises of the algorithm ensuring its universal applicability. For Altenbruch-Beverstedt, the large variance and standard deviation of the combined horizontal data and the difference between the mean values of both individual horizontal parameters reflect the strong anisotropy of the structure, while the statistics for the vertical data indicate a moderate variation in vertical measurements.

Gradient and curvature histograms of the example cases are visible in Fig. 6. For the sphere, the distribution of the gradient histogram is symmetric (Fig. 6a). The curvature histogram (Fig. 6c) shows a prevalence of very small values and subordinate maxima around 0.1, 0.2 and in the overflow bin that contains 394 values (5% of all data) above 0.37. For Altenbruch-Beverstedt, the gradient distribution is asymmetric and the number of infinite gradients is higher (Fig. 6b). In comparison to the curvature distribution of the sphere, the curvature data (Fig. 6d) is monomodal apart from the overflow bin.

The Kullback-Leibler (KL-)divergence (Kullback and Leibler, 1951) was calculated to quantitatively determine the similarity between the tested geometries. The distributions of the six parameters (the individual horizontal extents along both horizontal main axes, the combined horizontal data, the vertical data, gradients and curvatures) were compared between the models. The similarity of two distributions is larger, the smaller the KL divergence is, with a value of 0 indicating equality of the distributions. In addition to KL divergences of individual parameters, an averaged KL divergence was calculated: by taking the mean of the values between two models, the overall dis-/similarity between models was assessed. According to the averaged KL divergence, the sphere is most similar to “pillow(rounded)_batholithV4”, while the closest model to Altenbruch-Beverstedt is “wall(highly.anisotropic_hourglass-shape_rounded)”. The result of the calculation of the individual KL divergences for the example cases is visualized in Fig. 7. For the sphere, the most similar models regarding the respective distributions of the six parameters are the “sheet(cylindric_rounded)” for the horizontal data of the first direction, the “prism” for the orthogonal horizontal data, the “batholithV3” for the combined horizontal data, the “anticline_wall(rounded)_batholithV1” for the vertical data, the “phacolith” model for the gradients and the “ellipsoid” for the curvatures (compare Fig. 2 for the model appearances). For Altenbruch-Beverstedt, “batholithV6” is most similar regarding the horizontal data of the first direction, “roller” for the orthogonal horizontal data,

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

280 "wall(highly.anisotropic_hourglass-shape_flattened)" for the combined horizontal data, "wall(highly.anisotropic_hourglass-shape_rounded)" for the vertical data, "pillow_flattened" for the gradients and "roller" for the curvatures.

Table 1: a) First five statistical moments per parameter for the sphere model (note: the dimensions of the sphere are chosen arbitrarily). b) First five statistical moments per parameter for the model of the real subsurface structure (Altenbruch-
285 **Beverstedt). Statistics for Altenbruch-Beverstedt reflect the strong anisotropy of the structure**

(a) sphere	mean [m]	variance [m]	std_dev [m]	skew	median [m]
Horizontal length	14	13	4	-0.4	14
Horizontal length orthogonal	18	4	2	0	18
Vertical length	14	13	4	-0.5	15
Horizontal data combined	16	13	4	-0.8	17

(b) Altenbruch-Beverstedt	mean [m]	variance [m]	std_dev [m]	skew	median [m]
Horizontal length	3825	635352	797	-0.2	3795
Horizontal length orthogonal	48644	8956682	2993	-2.2	49845
Vertical length	4766	486158	697	-2.9	4856
Horizontal data combined	26235	5.07E+08	22516	0.02	21696

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

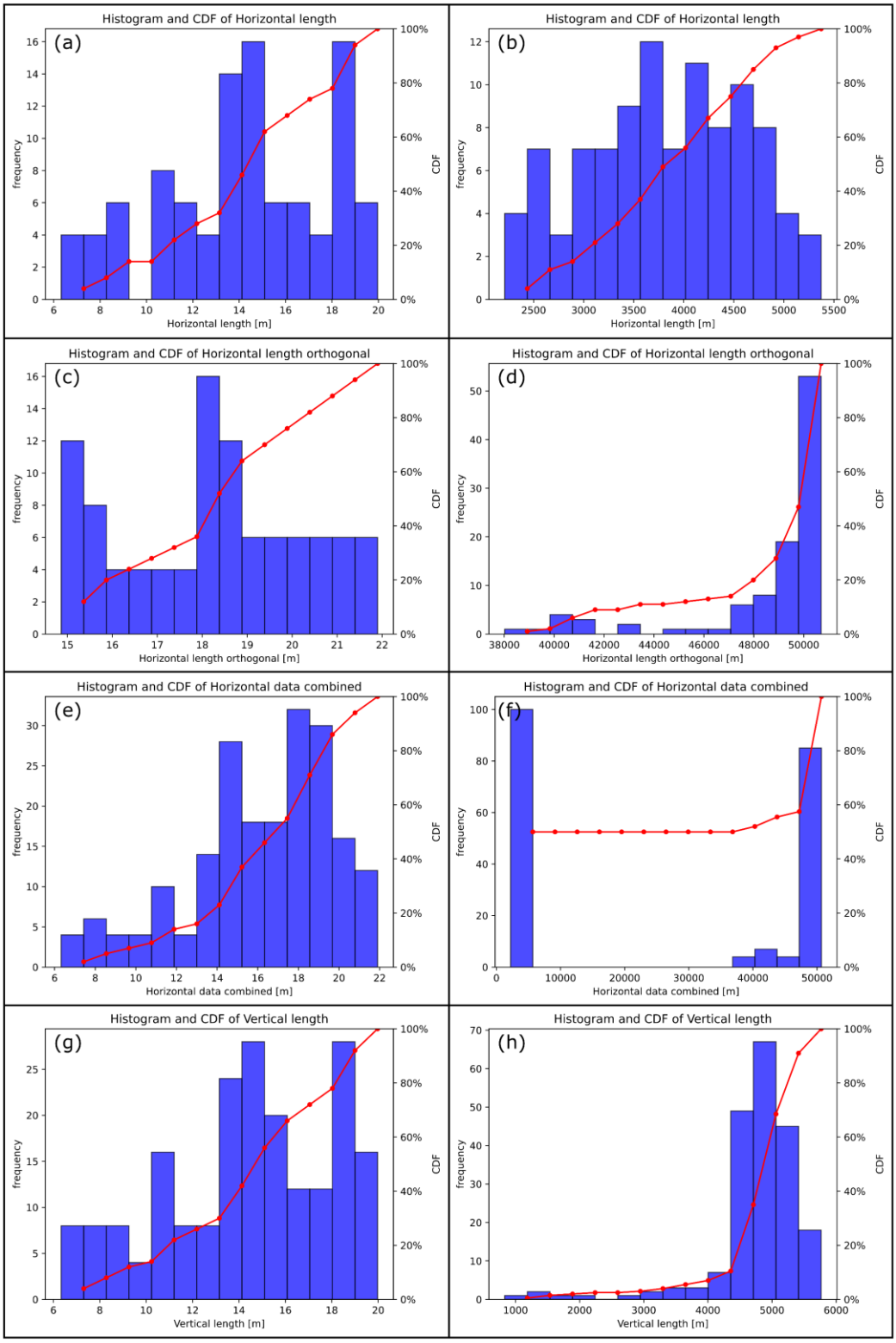


Fig. 5: Data distributions and cumulative distribution functions (CDF´s) for the extensional parameters. Left column: sphere, right column: Altenbruch-Beverstedt.
a & b) Horizontal data from the first direction.
c & d) Horizontal data from the orthogonal direction.
e & f) Combined horizontal data.
g & h) Vertical data

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

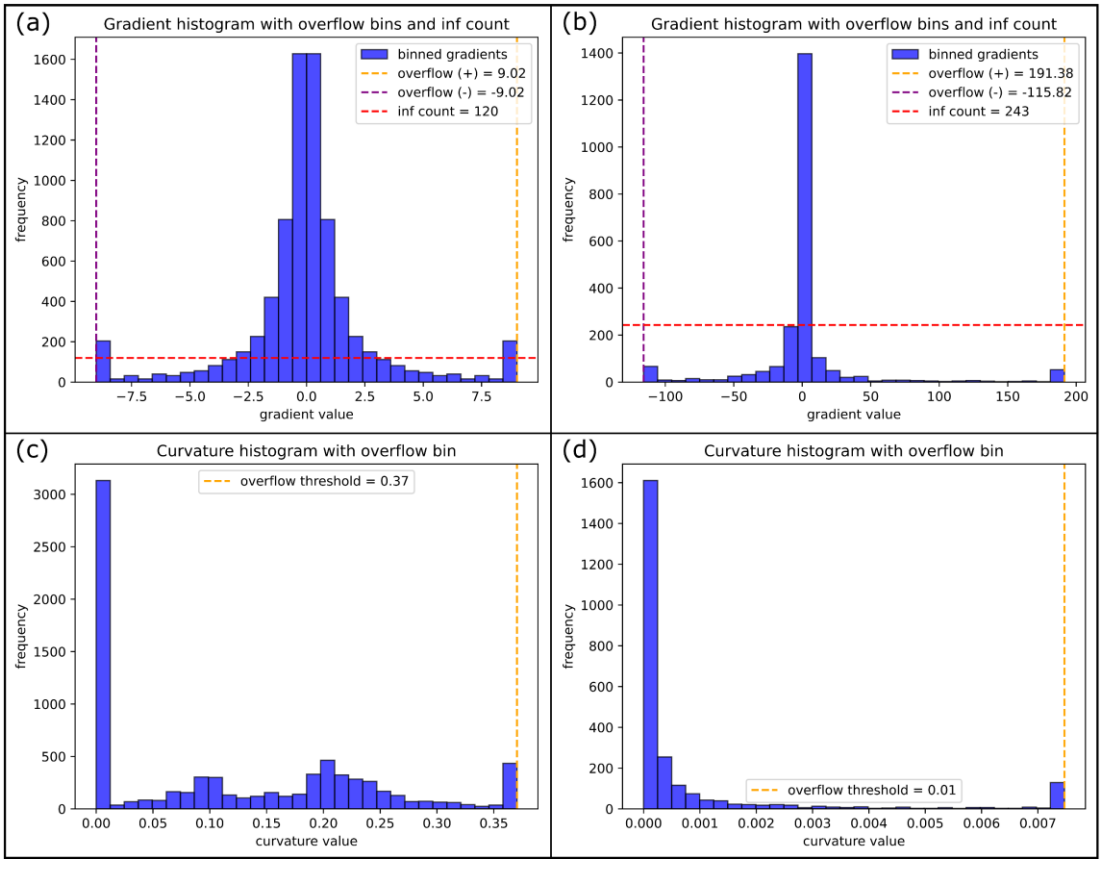


Fig. 6: Gradient and curvature data for the sphere (a & c) and Altenbruch-Beverstedt (b & d). Amount of data in overflow bins: Gradient diagram (sphere) 384 values (4.8% of all data), Curvature diagram (sphere) 394 values (5%); Gradient diagram (Altenbruch-Beverstedt) 112 values (4.6%); Curvature diagram 101 values (4.1%)

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

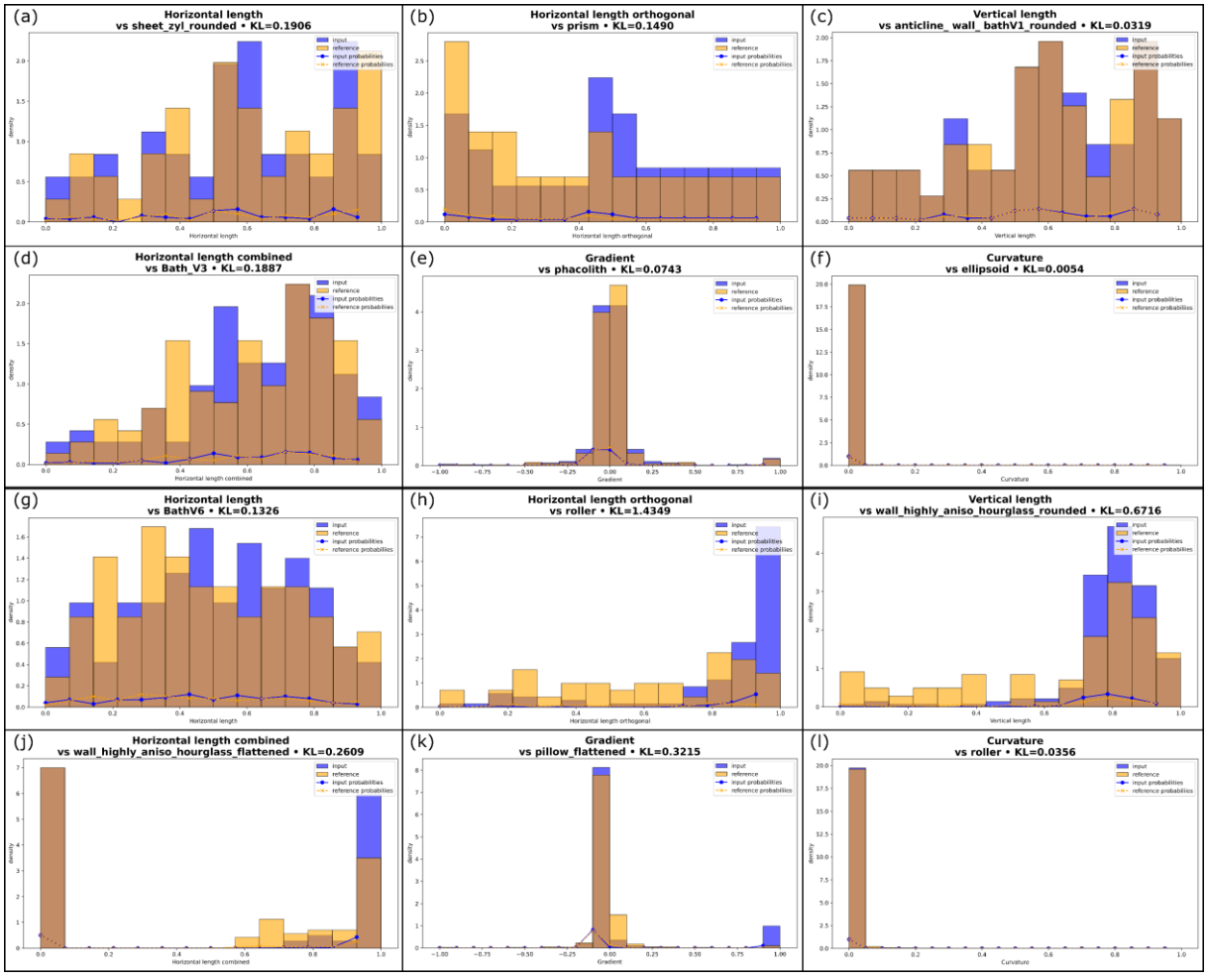


Fig. 7: Visualized Kullback-Leibler divergences. a-f: sphere, g-l: Altenbruch-Beverstedt. Visualized is the most similar data distribution (orange) compared to the respective distribution of the two example models (blue). The calculated Kullback-Leibler divergences are noted in the headlines of individual figures

3.3 Cluster analysis

Cluster analysis on all measured data of the regular geometries resulted in 7 clusters considering the combined analysis of the elbow plot and silhouette score (Fig. 8b). With the first two principle components (PC´s) only explaining 40% of the variance (see Fig. 8a), the number of PC´s necessary to cover more than 90% of the variance was determined to be 12. The feature contribution matrix (Fig. 8c) reveals the contribution ("loadings") of the binned PDF´s to the principal components, with bright yellowish colors indicating a strong positive contribution and dark blue colors a substantial negative contribution. In the contribution matrix, percentiles 0 to 19 represent the PDF of the combined horizontal data, followed by the vertical data (20 to 39), gradients (40 to 59) and curvatures (60 to 79). The principal component matrix plot is visible in Fig. 9.

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

Positive contributions to PC1 (Fig. 9a) are dominated by the 50 to 55% bins of the vertical data and gradient data, while there is no percentile with a strong negative contribution to PC1. This effectively separates the bluish-green cluster at high positive PC1 scores from the rest of the data. All four models ("flat.layer", "sill", "cube" and "prism"), share a distinct geometrical similarity: When segmenting them with our algorithm, cross sections are always flat at the top and of exactly the same vertical extent throughout the entire structure. This results in a step-wise appearance of the respective PDF´s, with the step being in the middle of the functions. For PC2, the 45 to 50% gradient bin has the highest positive loading, while there are stronger negative loadings for the 0 to 5% as well as 95 to 100% gradient bins. Therefore, models of the black cluster (highest positive PC2 scores) are characterized by many comparatively small gradient measurements. This reflects the presence of many low to moderately inclined surfaces in a geometry (depending on the variance in a data distribution) and an overall more rounded appearance (see e.g. the highlighted black example model "pillow(rounded)_batholithV4" in Fig. 9a). Meanwhile, models at high negative PC2 scores, like the blue cluster, show many large gradient data (positive and negative). This represents the abundance of steep-dipping to vertical surfaces for a dataset. Thus, PC2 is an indicator for the overall steepness of the lateral parts of a structure or, on the other hand, its sphericity. For PC3 (Fig. 9b), large positive contributions are spread among the 0 to 5% and 95 to 100% horizontal bins as well as the 95 to 100% vertical bin and the 0 to 5% and 45 to 50% gradient bins, while the only considerable negative loading is exhibited by the 10 to 15% gradient bin. Therefore, datasets with very positive PC3 scores (vermilion cluster and all but one blue model) represent anisotropic bodies with rather flat top surfaces and steep-dipping to vertical lateral walls (see e.g. the highlighted vermilion "dyke"). In contrast, however, datasets of largest negative PC3 scores (like the reddish-purple cluster), cannot be linked to very high data percentages in that 10 to 15% gradient bin; its loading (-0.27) not being the main cause of the observed negative PC3 scores. PC4 shows considerable positive loadings for the 50 to 55% bin of the vertical data and the 0 to 5% and 95 to 100% gradient bins. Meanwhile, large negative loadings are seen for the 0 to 5% and 95 to 100% horizontal bins, the 95 to 100% vertical bin and the 80 to 85% gradient bin. These contributions mainly drive the differentiation of the reddish-purple and vermilion clusters (negative PC4 scores) from the other clusters apart from some sky-blue models. Since the horizontal and vertical bins contributing very negatively are the same horizontal and vertical bins contributing particularly positively to PC3, it can be deduced that the overall position of the vermilion models in the PC3 vs. PC4 diagram is more driven by these horizontal and vertical bins. Meanwhile, the datasets from the reddish-purple and sky-blue models are comparatively influenced more by the 80 to 85% gradient bin also showing a considerable negative loading. Still, most datasets from these clusters at negative PC4 scores can be considered as rather anisotropic geometries with mainly steeper (but not vertical) lateral walls, while models at higher positive PC4 scores exhibit uniform vertical extents and steep-dipping to vertical lateral walls. This explains the position of the isolated blue model at highest positive PC4 scores ("volcanic.pipe"; highlighted in Fig. 9b; see also Fig. 10h for the gradient distribution), completing the separation of the blue cluster from the rest of the data. PC5 (Fig. 9c) shows strong positive contributions for the 0 to 5% vertical bin and the 50 to 55% and 95 to 100% gradient bins, while stronger negative loadings are given by the 50 to 55% vertical bin and the 0 to 5% gradient bin. This separates the majority of the sky-blue cluster (highest positive PC5 scores) from the rest of the datasets. As this corresponds to the first appearance

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

of the 0 to 5% vertical bin among considerable contributing bins, most of the associated models are characterized by widespread low vertical extents and few much larger ones, as seen in overhang configurations (for example, see the model "laccolith" in Fig. 2). PC6 is mainly influenced by the gradient data, where the 10 to 15% bin contributes the most negatively and the 80 to 85% bin contributes positively. Once again (as for PC3), the 10 to 15% gradient bin, however, does not seem to be the main reason for the separation of the reddish-purple cluster at very negative PC6 scores. Similarly, the sky-blue models at higher positive PC6 scores do not exhibit particularly large high percentages in the respective bin. In the remaining visualized PC-cluster plots, the variance explained by the PC´s does not allow any major discrimination anymore and partially repeats patterns seen in the more informative plots.

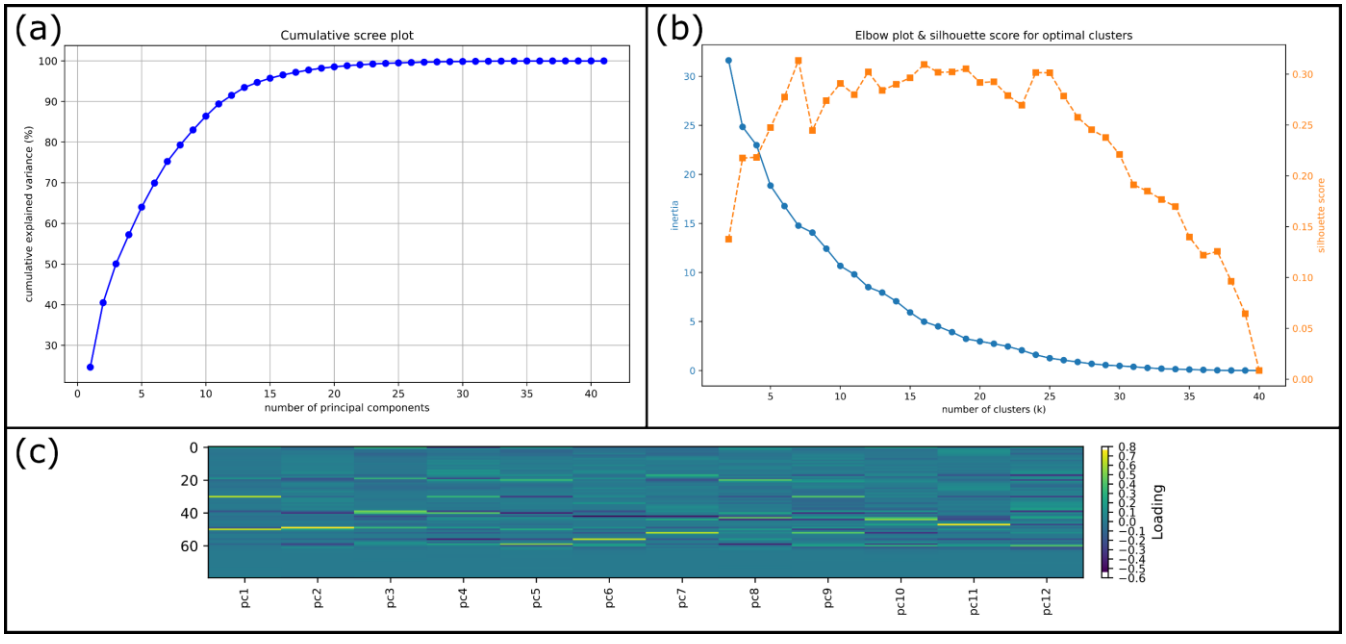


Fig. 8: Calculated supplementary information for the setup and interpretation of the cluster analysis after principal component analysis. a) Cumulative scree plot, showing the explained variance with increasing number of principal components. b) Elbow plot and silhouette score to determine number of clusters. c) Contribution matrix showing the contribution of the input data (percentiles of the probability density functions of measured parameters) to the principal components

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

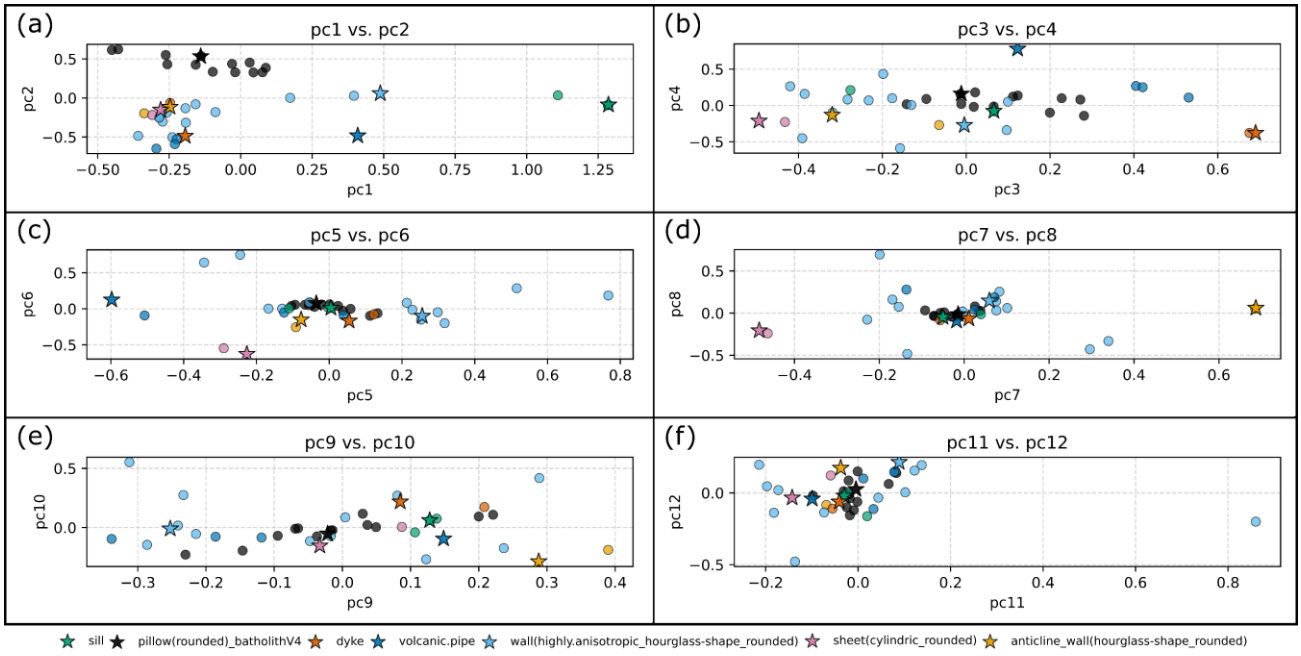


Fig. 9: Matrix plot of principal components (PC´s) explaining 90% of the variance in the data (see also Fig. 8a). As the first PC´s explain the most amount of variance (decrease of explained variance with increasing number of PC´s), most information can be gained from the first 3 cluster plots (see text). Larger stars mark example models for clusters (see legend)

4 Discussion

By applying a set of defined geometrical descriptors to systematically generated 3D benchmark models, this study establishes a framework for the quantitative comparison of shape properties. The analysis highlights how key attributes such as anisotropy, surface morphology, and sphericity vary across models, offering a structured perspective on their geometric dis-/similarities. These outcomes prompt a deeper discussion of how well the proposed descriptors capture meaningful shape differences and how this quantitative framework advances the analysis of 3D geological structures.

4.1 Comparison of existing methods for 3D shape characterization with the proposed workflow

The workflow of this study differs from many approaches developed for the description of 3D objects, as the application for most of these methods lies in the shape characterization for image recognition rather than for exact geometrical characterization of a body. Therefore, the majority of studies (e.g. Canul-Ku et al., 2019; Çapar et al., 2009; Dumitru et al., 2022; Ettl et al., 2007; Goh & Chan, 2003; Mousa, 2011; Muzahid et al., 2021; Wohlkinger & Vincze, 2011) define shape descriptors based on varied input data to automatically detect specific shapes in binary, grayscale and/or colored images or

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

point cloud data from large databases. Meanwhile, our approach aims at characterizing 3D structures through precise measurements of geometrical parameters, thereby providing datasets suited for quantitative comparison.

Studies presenting approaches that show similarities to ours are Celenk (1995), Schweizer et al. (2017) and Lindsay et al. (2013). Celenk (1995) determines the horizontal main axes of equally-spaced cross sections as well, but does so to align sections of two different objects. Comparison is then achieved by computing the averaged shape difference of sections between the objects in four directions along the main axes. Key differences of our approach therefore involve the segmented assemblage of cross sections in the orthogonal direction (following the respective segmented horizontal main axis) and the exact measurement of the dimensional extents on the sections. Hence, our method opts for the determination of larger datasets of absolute measurements on a single object, that are compared to other bodies in subsequent steps. Meanwhile, Celenk (1995) computes the relative measure that is the averaged shape difference, representing a faster, but more approximate approach of object comparison between two objects, as the author does not segment the horizontal main axis along the larger extent. Schweizer et al. (2017) do not try to compare the dimensions of individual 3D structures, but use the Jaccard distance and the normalized city-block distance as measures for model dissimilarity instead. The two parameters are being applied as measures for the similarity in position of certain geological units between two model realizations of the same study site. In a similar fashion, the Hausdorff distance has been used before (see e.g. Suzuki et al., 2008). These dissimilarity distances were not applied in our study, as they could only act as size indicators rather than shape descriptors and would not give any indication on where two structures differ spatially. Meanwhile, our approach provides insight into both shape and size differences of objects, which is crucial for geological modeling. Lindsay et al. (2013) explore geometric uncertainty across multiple realizations of a study site, evaluating parameters like depth, volume, and curvature, which parallel those in our study. However, their parameters are often tied to stratigraphic units and may not apply directly to individual 3D structures. Both studies utilize PCA to analyze geometric variability and model differences, although executed differently. Despite these existing methodologies, we opted for a straightforward approach, allowing us to efficiently replicate the main geometric characteristics of input datasets. Our algorithm is computationally efficient, easily interpretable with basic geological knowledge, and accessible to a non-specialist audience.

4.2 Assumptions and compromises of the algorithm ensuring its universal applicability

Despite the strengths of our methodology, certain limitations must be acknowledged. The reliance on discrete differential geometries (Bobenko et al., 2008) means that the input dataset must represent a single, compact, and topologically connected structure (Thiele et al., 2016). For objects separating from one to the next cross section into multiple strands, split algorithms are available. However, this comes at the cost of interpretability of the statistics of geometric properties. Moreover, the method functions optimally for convex hulls (Rockafellar, 1970), focusing on the restricted set of geometries which are relevant for the site selection of nuclear waste deposits. Still, this may restrict its applicability to more complex geological formations. These assumptions should be considered when interpreting results in other domains.

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

The focus of this study was to establish a generalized algorithm to infer dis-/similarity between geometries. Given the wide range of potential and available models, the algorithm requires some trade-offs to be universally applicable. Discussion of the data distributions for the geometrical parameters (see Fig. 5) focuses on the results from measuring the sphere, representing a comprehensible case with distinct expected data distributions: The nature of a sphere is a similar shape of any section through the center, eventually resulting in a normal distribution of the levelled distance measurements in both horizontal and vertical directions. This expected distribution is not produced in our case due to the generation approach of the orthogonal cross sections: The assembled sections follow the contour of the structure (see Fig. 4g & h), which results in larger measurements for the orthogonal horizontal data and a slightly tailed distribution of the combined horizontal data, similar to an ellipsoid with a low contrast in the main axes. As this situation is rarely seen in geological modeling, the impact is small since anisotropic geometries are measured accurately with our segmentation algorithm. The gradient and curvature data reflect the effects of our approach as well: While the gradient diagram of the sphere shows a symmetric distribution as expected, the relative elongation of the orthogonal sections increases the frequency of lower gradient measurements. Due to this accumulating effect, the presence of low-dipping surfaces of a structure is overestimated by the data. Furthermore, the exclusion of marginal cross sections leads to vertical clipping that introduces infinite gradient measurements (representing two consecutive vertices being exactly vertical) that would not exist when measuring the sections in a rounded, unclipped state. The curvature data is influenced by this clipping as well, that results in few large values where the three consecutive vertices form a large angle. These measurements increase the variance of curvature data considerably, with the majority of data for most datasets being located within the 0 to 5% and 5 to 10% bins.

4.3 Analysis of parameter distributions and model dis-/similarity

Analyzing the data distributions of a structure visually already reproduces distinct geometrical characteristics of an input dataset. The distribution of the combined horizontal data indicates whether a pronounced anisotropy is present for an analyzed structure: if the data is separated into two clearly distinguishable subordinate distributions (see Fig. 10a), the geometry is considerably anisotropic (the farther apart the two maxima, the more anisotropic a body is). Caution is advised for a distribution with two close maxima (Fig. 10b): this could be the consequence of the inflated extent in the orthogonal direction (see above). Analyzing the combined horizontal data and the vertical data together reveals whether a structure shows substantial variations in its horizontal extent over its vertical range. Such a shape, in the subsurface more often present as overhangs rather than as upward tapering, is indicated by the simultaneous presence of multimodal distributions for both parameters (Fig. 10c & d). The vertical data distribution also characterizes the top surface of a geometry: if the distribution is monomodal, with a) the maximum being the bin representing the highest measurements, and b) the frequency in lower bins being substantially smaller, then the presence of a flat top surface is indicated. The existence of a flat top surface can be verified by analyzing the gradient and curvature data: a high frequency of very small measurements for both parameters supports such an analysis (Fig. 10e-g). Gradient data also indicates the steepness of lateral surfaces of a body: as high and infinite gradient data stem from steep to vertical faces of a structure, the presence of steep-dipping lateral surfaces can be

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

recognized (Fig. 10h). Combining the inferences from analyzing top and lateral surfaces therefore provides insight into the overall sphericity of an input dataset: a more spherical structure is represented by larger quantities of intermediate gradient measurements and of moderate to high curvature data (Fig. 10i & j).

The Kullback-Leibler (KL-)divergence is used to quantitatively compare the individual parameter distributions of two structures: by providing a single value for a given statistical parameter and pair of input models, an intuitive and quick way for assessing the similarity of two distributions is given, as the similarity is higher, the smaller the KL divergence. The average of the respective values of KL divergence was employed as a measure for the dis-/similarity of full models as well. However, informational content of this parameter is limited, as there is no indication regarding which parameters two compared structures are most similar or differ more. Therefore, principal component analysis and K-means clustering have been employed as well, providing this information based on all combined parameters. In general, values of KL divergence show an error for the gradient distributions: infinite values had to be converted to the highest finite gradient value of a given dataset to enable the computation, inflating the highest bin. Furthermore, the large variance of curvature data for most input models (see for example Fig. 7f &l) decreases the applicability of the KL divergence for that parameter, as most models show very similar normalized distributions. To assess the impact of the large variance on individual KL divergences of curvature data and smallest averaged KL divergences, they were also calculated using a 95th percentile overflow bin (see Table 2). Smallest KL divergences for the curvatures of the two example models are notably higher, especially for the sphere, reflecting the dissimilarity of data distributions when applying the filter (column 1 & 2). The impact on the smallest averaged KL divergence (column 3 & 4) is smaller, yet still considerable.

Table 2: Comparison of KL divergences with and without the usage of a 95th percentile overflow bin for the curvature distributions.

structure	smallest KL divergence for curvature without overflow bin	smallest KL divergence for curvature with overflow bin	smallest averaged KL divergence for all properties without overflow in curvature	smallest averaged KL divergence for all properties with overflow in curvature
sphere	0.0054 (“ellipsoid”)	0.16 (“wall(highly.anisotropic_cylindric_rounded)”)	0.45 (“pillow(rounded)_batholithV4”)	0.56 (“batholithV3”)
Altenbruch-Beverstedt	0.036 (”roller”)	0.05 (“batholithV5”)	1.2 (“wall(highly.anisotropic_hourglass-shape_rounded)”)	1.3 (“wall(highly.anisotropic_hourglass-shape_rounded)”)

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

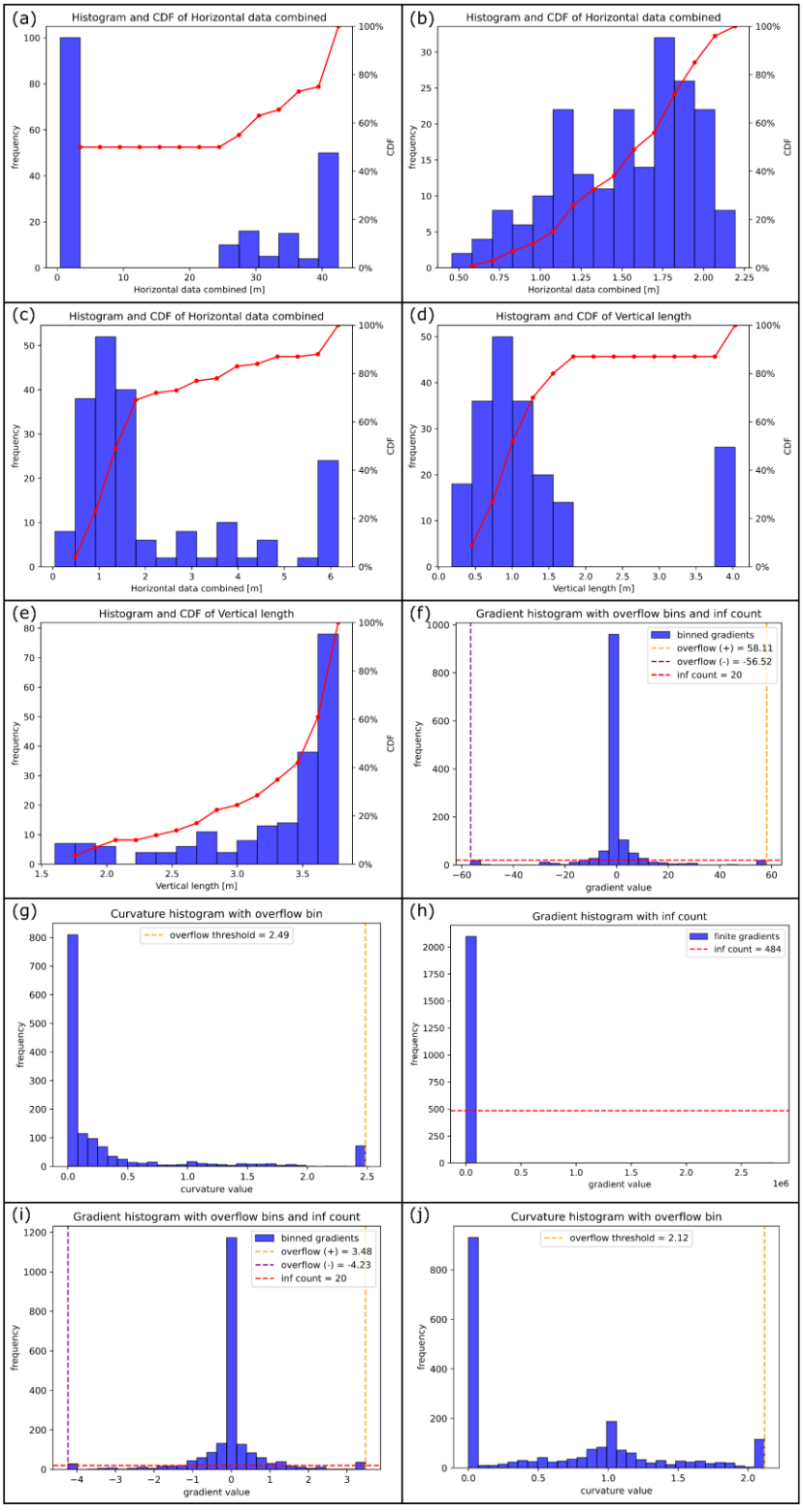


Fig. 10: Analysis of data distributions to reproduce geometric characteristics of input models. Please compare with model appearances in Fig. 2.
a) Combined horizontal data of "wall(highly.anisotropic_hourglass-shape_rounded)", reflecting strong anisotropy.
b) Combined horizontal data of "batholithV3" incorrectly indicating slight anisotropy.
c) & d) Combined horizontal data and vertical data of "sheet(hourglass-shape_rounded)" indicating presence of overhangs.
e)-g) Vertical data, gradients and curvatures of "batholithV5", revealing the presence of a flat top surface
h) Gradient data of "volcanic.pipe", reflecting the prominence of vertical lateral walls)
i) & j) Gradient and curvature data of "pillow(rounded)_batholithV4" showing the spherical character of the input model

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

The discriminability of standard geometries in the cluster analysis is ensured despite the compromises made in the methodology: structures of varying anisotropy plot in different parts of cluster diagrams showing contribution of the combined horizontal data, as the highest contribution of horizontal data comes from the first and last distribution bins (see Fig. 8c). Similarly, as horizontal and vertical data distributions of flattened geometries show the discussed characteristic properties, they differ in their PC scores from their rounded counterparts. Furthermore, the discussed increased frequency of gradient measurements around 0 does not change cluster patterns as it applies to all datasets. The same is true for the artifact-influenced curvature data and its impact on the general clustering of similar structures. Still, its squeezed nature shows an effect on the overall clustering results, as the curvature data does not show any considerable contribution in the PCA.

The overall cluster results validate that the flat and/or cuboidal geometries (vertical extent ≤ horizontal extent and/or exclusively straight lateral surfaces) mostly differ considerably from the other structures designed to represent intrusive subsurface bodies: the flat/cuboidal geometries are mainly distributed among the bluish-green, blue and vermilion clusters (see Fig. 9). As recognized in the results, these three clusters can be differentiated from the other models within the first four PC´s. Only the “eroded.layer(pinchout)” is located outside, in the sky-blue cluster, although representing a flat geometry. The K-means cluster analysis furthermore indicates that some standard geometries are similar across all parameters. Therefore, it was assessed whether certain standard geometries are redundant to simplify the benchmark selection process. The pairs of flattened and rounded versions show high similarities, leading to the exclusion of the flattened models while retaining the "uneroded" structures. The models "batholithV3" and "pillow(rounded)_batholithV4" cluster closely, differing only in vertical elongation; thus, "batholithV3" is excluded. Although similarities exist between models with varying lateral characteristics, both "cylindric" and "hourglass" shape variations are retained. This also applies to various "sheets" and the "anticline_wall(hourglass-shape_rounded)”, which exhibit similar PC scores in some, but not all cluster plots. Lastly, some of the flat and cuboidal bodies in the bluish-green cluster ("flat.layer", "sill" and "cube") are nearly identical in position. The "cube" is excluded from the benchmark collection, while the other two geometries are merged, keeping the shape of the "sill”.

4.4 Online collection of benchmark geometries

Following the application of the established methodology to the standard geometries, the collection can be effectively condensed from 36 to 25 benchmark models (see Fig. 11). Decreasing the database by validating the bodies’ geometrical dissimilarity facilitates the decision making on the best suitable benchmark for a case study. Despite our reduction efforts, this list is not expected to be exhaustive: we would like to encourage users to suggest additional geometries based on their expertise and/or literature, to ensure that suitable benchmark models are available for as many geomodeling applications as possible.

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

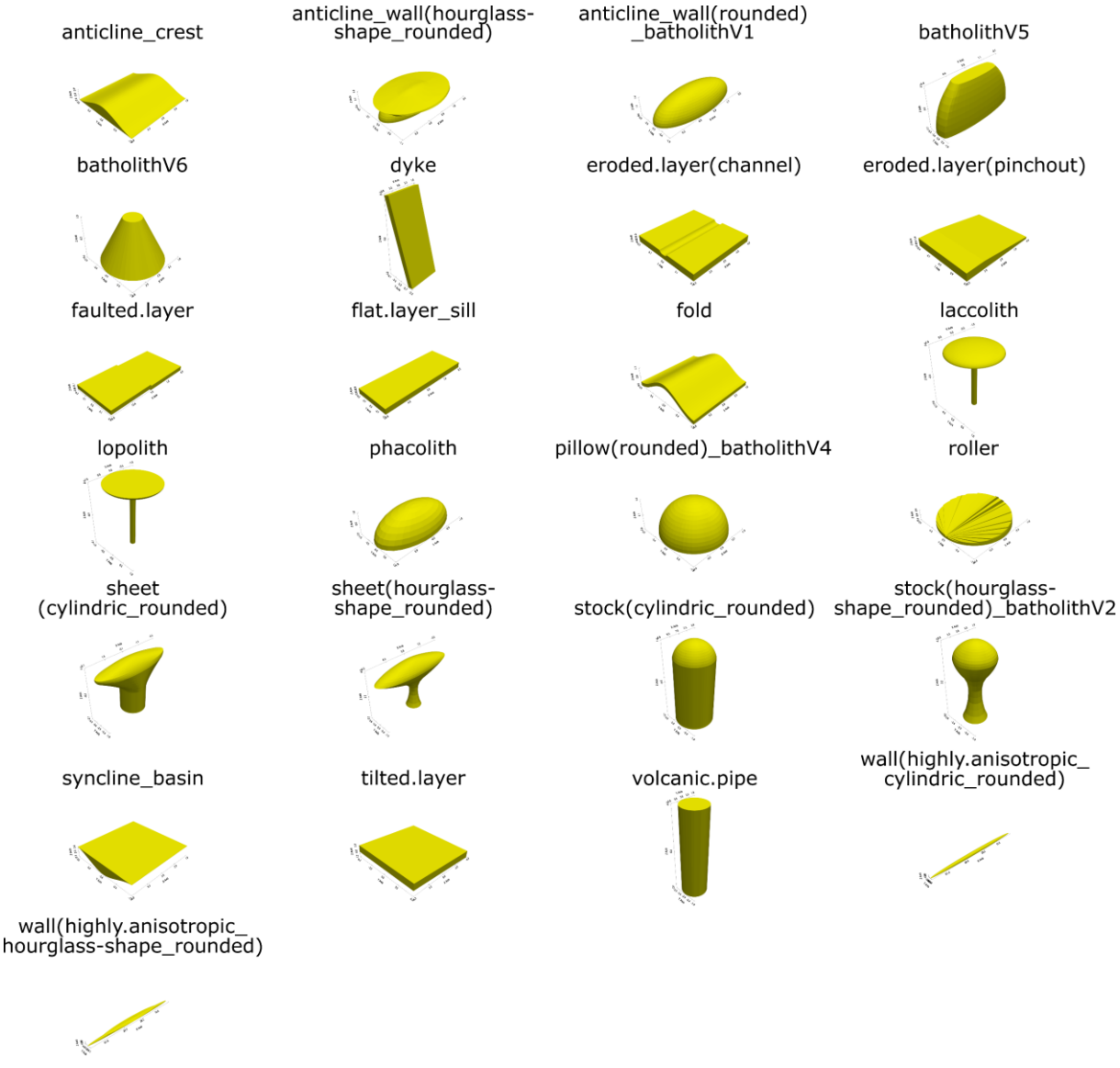


Fig. 11: Condensed collection of standard geometries after application of the quantification method.

4.5 Potential usage of the quantification method and the standard geometries

The methodology will be a part of a larger framework to model and compare geological structures based on sparse data. For most regions of interest for nuclear waste disposal, seismic 2D data are available frequently with a few boreholes. This is similar to the cross sections established through the benchmark models here, allowing for a fast model selection based on the geometrical properties and potentially further constraining hyper parameter selection for interpolation. Since the

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

interpolation will result in a series of stochastic prior realizations, the method will be used for falsification by data (e.g. boreholes). Furthermore, applying our methodology also supports testing for the minimum amount of data necessary for geological modeling, as different data densities and configurations can be inserted into the algorithm. The open-access collection of benchmarks for geomodeling is also a convenient tool to visualize the range of three-dimensional geometries of the different rock types to a broader audience, which aids in the communication of uncertainties and decisions for geoscientists and stakeholders in various settings (see Zehner, 2021).

5 Conclusion

In our publication, we presented a methodology to quantitatively describe, compare and systematize 3D geometries, and proposed a set of regular standard geometries as benchmark models in geomodeling approaches. Demonstrating the quantification method on the standard geometries, their geometrical dis-/similarity is assessed. The combined evaluation of data distributions and a cluster analysis reproduces the main geometrical characteristics of input meshes and visualizes differences between various datasets. While distributions of combined horizontal extensional measurements provide insight into the anisotropy of datasets and the potential existence of overhangs, distributions of the vertical extent indicate the character of the top surface of structures and support or falsify the presence of overhangs. Distributions of gradient and curvature data (1) indicate the prevailing character of the slope of the lateral surfaces of structures, (2) further emphasize potentially present flat top surfaces and (3) give a general indicator on the sphericity of a structure. Cluster analysis of normalized, dimensionally reduced data groups and systematizes input structures based on the combined measured statistical parameters. In our application to synthetic datasets, clustering also serves to identify and exclude or merge benchmark models showing large geometrical similarity. Apart from cluster analysis and assessment of data distributions, comparison of parameter distributions is furthermore achieved using the Kullback-Leibler divergence.

As already indicated earlier, the first follow-up study aims at applying the method to a large database of structural geological models. Afterwards, the method will be applied to datasets of sparse, unmodeled input data instead of already precomputed meshes of structural models. At this point, the method might be coupled with spatial interpolation algorithms. Thereby, it would fit well into a study focusing on geomodeling based on progressively reduced datasets that is planned to be conducted later.

Code and data availability

Method development was carried out in Python. The method mostly relies on the capabilities of the libraries Shapely (https://shapely.readthedocs.io/en/stable/), PyVista (https://pyvista.org/) and Plotly (https://plotly.com/). The python code, the condensed database of standard geometries (as .vtk-files) and the datasets of raw extensional, gradient and curvature data are stored at https://doi.org/10.5281/zenodo.15795851, (Carl, 2025).

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

Author contribution

Friedrich Carl: writing–original draft; method development; data acquisition and analysis; conceptualization; visualization. Peter Achtziger-Zupančič: writing–review and editing; conceptualization; supervision; funding acquisition. Jian Yang: writing–review and editing. Marlise Colling Cassel: writing–review and editing. Florian Wellmann: writing–review and editing; supervision; funding acquisition.

Declarations

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Friedrich Carl, Jian Yang and Marlise Colling Cassel are funded by the German Federal Company for Radioactive Waste Disposal (BGE). Other authors declare that they have no conflict of interest.

Acknowledgements

This study was conducted as part of the GeoBlocks project. GeoBlocks is a collaboration of RWTH Aachen University, University of Aberdeen and the German Federal Institute for Geosciences and Natural Resources (BGR) within the research cluster URS (Uncertainty and Robustness with regard to the Safety of a repository for high-level radioactive waste), funded by the German Federal Company for Radioactive Waste Disposal (BGE), with the main objective being the creation of an open-source workflow for geological modeling that includes the quantification and visualization of uncertainties and aims at the optimization of sampling procedures. We thank for the financial support by BGE. The authors would also like to thank BGR for the data provided to the study.

References

BGR, LBGR, LLUR, and LUNG: 3D-Modell des geologischen Untergrundes des Norddeutschen Beckens (Projekt TUNB), 2022.

Bobenko, A. I., Schröder, P., Sullivan, J. M., & Ziegler, G. M. (2008). Discrete Differential Geometry. Birkhäuser Basel. http://publications.mfo.de/handle/mfo/521

Canul-Ku, M., Hasimoto-Beltran, R., Jimenez-Badillo, D., Ruiz-Correa, S., and Roman-Rangel, E.: Classification of 3D Archaeological Objects Using Multi-View Curvature Structure Signatures, IEEE Access, 7, 3298–3313, https://doi.org/10.1109/ACCESS.2018.2886791, 2019.

Çapar, A., Kurt, B., and Gökmen, M.: Gradient-based shape descriptors, Mach. Vis. Appl., 20, 365–378, https://doi.org/10.1007/s00138-008-0131-5, 2009.

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

555 Carl, F.: Code and datasets of the quantification algorithm and benchmark model collection, https://doi.org/10.5281/zenodo.15795851, 2025.

Carl, F., de los Angeles Gonzalez de Lucio, G., Yang, J., Achtziger-Zupančič, P., Kukla, P. A., Bense, F., and Wellmann, J. F.: Host rock analysis for the German nuclear waste disposal site-selection: review of subsurface geometries and input data for geological modelling, RING, 1, 2023.

560 Celenk, M.: Three-dimensional object recognition using cross-sections, in: Proceedings of the Twenty-Seventh Southeastern Symposium on System Theory, Twenty-Seventh Southeastern Symposium on System Theory, Starkville, MS, USA, 92–96, https://doi.org/10.1109/SSST.1995.390611, 1995.

Demicco, R. V., & Klir, G. J. (2003). Fuzzy Logic in Geology. Elsevier.

Drevin, G. R. (2006). Computational Methods for the Determination of Roundness of Sedimentary Particles. Mathematical
565 Geology, 38(7), 871–890. https://doi.org/10.1007/s11004-006-9051-y

Dumitru R. G. and Gorgan, D.: 3D Object Recognition using Enhanced Slicing, in: 2023 IEEE 19th International Conference on Intelligent Computer Communication and Processing (ICCP), 2023 IEEE 19th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 163–170, https://doi.org/10.1109/ICCP60212.2023.10398710, 2023.

570 Dumitru, R. G., Antonio Toma, S., and Gorgan, D.: 3D Object Recognition Method Using CNNs and Slicing, in: 2022 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR), 2022 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR), Cluj-Napoca, Romania, 1–6, https://doi.org/10.1109/AQTR55203.2022.9802063, 2022.

Ettl, S., Kaminski, J., Olesch, E., Strauss, H., and Häusler, G.: Fast and robust 3D shape reconstruction from gradient data,
575 2007.

Goh, W.-B. and Chan, K.-Y.: Shape Description Using Gradient Vector Field Histograms, in: Scale Space Methods in Computer Vision, vol. 2695, edited by: Griffin, L. D. and Lillholm, M., Springer Berlin Heidelberg, Berlin, Heidelberg, 713–728, https://doi.org/10.1007/3-540-44935-3_50, 2003.

Hillier, M., Wellmann, J. F., Brodaric, B., de Kemp, E., & Schetselaar, E. (2021). Three-Dimensional Structural Geological
580 Modeling Using Graph Neural Networks. Mathematical Geosciences, 53(8), 1725–1749. https://doi.org/10.1007/s11004-021-09945-x

Hudec, M. R. and Jackson, M. P. A.: Growth of allochthonous salt sheets in passive margins and orogens., AAPG Bull., 1535–1564, 2006.

Hudec, M. R. and Jackson, M. P. A.: Terra infirma: Understanding salt tectonics, Earth-Sci. Rev., 82, 1–28,
585 https://doi.org/10.1016/j.earscirev.2007.01.001, 2007.

Hudec, M. R., Jackson, M. P. A., Cottington, N., Vendeville, B. C., Schultz-Ela, D. D., and Dooley, T. P.: The salt mine: a digital atlas of salt tectonics, Bureau of Economic Geology, Jackson School of Geosciences, University of Texas at Austin; American Association of Petroleum Geologists, Austin, Tex.: Tulsa, Okla, 305 pp., 2011.

Jackson, M. P. A. and Talbot, C. J.: A Glossary of Salt Tectonics, 1991.

Jolliffe, I. T.: Principal Component Analysis, Springer-Verlag, New York, https://doi.org/10.1007/b98835, 2002.

Kindratenko, V. V., Van Espen, P. J. M., Treiger, B. A., & Van Grieken, R. E. (1996). Characterisation of the Shape of Microparticles via Fractal and Fourier Analyses of Scanning Electron Microscope Images. In D. Benoit, J.-F. Bresse, L. Van't dack, H. Werner, & J. Wernisch (Hrsg.), Microbeam and Nanobeam Analysis (S. 355–361). Springer. https://doi.org/10.1007/978-3-7091-6555-3_28

Kullback, S. and Leibler, R. A.: On Information and Sufficiency, Ann. Math. Stat., 22, 79–86, https://doi.org/10.1214/aoms/1177729694, 1951.

Laga, H., Guo, Y., Tabia, H., Fisher, R. B., and Bennamoun, M.: 3D Shape Analysis: Fundamentals, Theory, and Applications, John Wiley & Sons, 368 pp., 2019.

Li, J., Lu, L., & Lai, M. O. (2000). Quantitative analysis of the irregularity of graphite nodules in cast iron. Materials Characterization, 45(2), 83–88. https://doi.org/10.1016/S1044-5803(00)00052-8

Lindsay, M. D., Jessell, M. W., Ailleres, L., Perrouty, S., de Kemp, E., and Betts, P. G.: Geodiversity: Exploration of 3D geological model space, Tectonophysics, 594, 27–37, https://doi.org/10.1016/j.tecto.2013.03.013, 2013.

Liu, M., Vashisth, D., Grana, D., & Mukerji, T. (2023). Joint Inversion of Geophysical Data for Geologic Carbon Sequestration Monitoring: A Differentiable Physics-Informed Neural Network Model. Journal of Geophysical Research: Solid Earth, 128(3), e2022JB025372. https://doi.org/10.1029/2022JB025372

Markl, G.: Minerale und Gesteine: Mineralogie – Petrologie – Geochemie, Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-662-44628-7, 2015.

Meyer-Baese, A. and Schmid, V. J.: Pattern Recognition and Signal Analysis in Medical Imaging, Elsevier, 467 pp., 2014.

Mousa, M. H.: Matching 3D objects using principle curvatures descriptors, in: Proceedings of 2011 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, 2011 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PacRim), Victoria, BC, Canada, 447–452, https://doi.org/10.1109/PACRIM.2011.6032935, 2011.

Muzahid, A. A. M., Wan, W., Sohel, F., Wu, L., and Hou, L.: CurveNet: Curvature-Based Multi-task Learning Deep Networks for 3D Object Recognition, IEEECAA J. Autom. Sin., 8, 1177–1187, https://doi.org/10.1109/JAS.2020.1003324, 2021.

Nichols, G.: Sedimentology and Stratigraphy, 2nd Edition., 2009.

Novotni, M. and Klein, R.: A geometric approach to 3D object comparison, Proc. Int. Conf. Shape Model. Appl., 2001.

Philpotts, A. and Ague, J.: Principles of Igneous and Metamorphic Petrology, 2nd ed., Cambridge University Press, Cambridge, https://doi.org/10.1017/CBO9780511813429, 2009.

Picard, M. D.: Classification of Fine-grained Sedimentary Rocks, SEPM J. Sediment. Res., Vol. 41, https://doi.org/10.1306/74D7221B-2B21-11D7-8648000102C1865D, 1971.

Rockafellar, R. T.: Convex analysis, Princet. Math. Ser., 28, 1970.

https://doi.org/10.5194/egusphere-2025-3203
Preprint. Discussion started: 25 August 2025

Schweizer, D., Blum, P., and Butscher, C.: Uncertainty assessment in 3-D geological models of increasing complexity, Solid Earth, 8, 515–530, https://doi.org/10.5194/se-8-515-2017, 2017.

Selley, R.: Applied Sedimentology, Academic Press, 523 pp., 2000.

Shamir, A.: A survey on Mesh Segmentation Techniques, 2008.

Stephenson, P.: The Shoelace Formula & Pick's Theorem, Math. Sch., 47, 35–38, 2018.

Sun, J., & Li, Y. (2015). Multidomain petrophysically constrained inversion and geology differentiation using guided fuzzy c-means clustering. Geophysics, 80(4), ID1–ID18. https://doi.org/10.1190/geo2014-0049.1

Suzuki, S., Caumon, G., and Caers, J.: Dynamic data integration for structural modeling: model screening approach using a distance-based model parameterization, Comput. Geosci., 12, 105–119, https://doi.org/10.1007/s10596-007-9063-9, 2008.

Thiele, S. T., Jessell, M. W., Lindsay, M., Ogarko, V., Wellmann, J. F., and Pakyuz-Charrier, E.: The topology of geology 1: Topological analysis, J. Struct. Geol., 91, 27–38, https://doi.org/10.1016/j.jsg.2016.08.009, 2016.

Winter, J.: Principles of Igneous and Metamorphic Petrology, 2013.

Winter, J. (2013). Principles of Igneous and Metamorphic Petrology. https://elibrary.pearson.de/book/99.150005/9781292034768

Wohlkinger, W. and Vincze, M.: Shape distributions on voxel surfaces for 3D object classification from depth images, in: 2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), 115–120, https://doi.org/10.1109/ICSIPA.2011.6144153, 2011.

Zehner, B.: On the visualization of 3D geological models and their uncertainty, Z. Dtsch. Ges. Für Geowiss., 83–98, https://doi.org/10.1127/zdgg/2020/0251, 2021.