The estimation of porosity in Japan Trench plate boundary using low-resolution X-ray computed tomography (XCT) images and laboratory measurements

Hamed Amiri[1], Francesco Cappuccio[2], Mai-Linh Doan[1], Marianne Conin[3], Virginia Toy[2]

[1] Université Grenoble Alpes, Institut des Sciences de la Terre (ISTerre), Grenoble, 38058, France

[2] Institute of Geosciences, Johannes Gutenberg University Mainz, Mainz, 55128, Germany.

[3] University of Lorraine, CNRS, Nancy, 54000, France.

Correspondence to: Hamed Amiri (amiri.hamed@gmail.com)

Abstract. X-ray computed tomography (XCT) is an advanced imaging technique that has been increasingly used in the past years because it can provide valuable information on internal structures of a rock sample in a non-invasive manner. The maximum resolution of lab-based XCT facilities is ~ 0.5 which might be sufficient to capture macropores in some rocks (i.e., sandstone), but will result in underestimation of porosity in clay-rich sediments containing micro-and nano-scale pores. Furthermore, such high-resolution XCT facilities are quite expensive and not ubiquitous. In this study, we introduce a new methodology based on the K-means clustering algorithm to process of low-resolution XCT images, illustrating its capability through porosity analysis of drillcores obtained during Integrated Ocean Drilling Program (IODP) expedition 343. The cation exchange capacity (CEC) of the squeezed samples of the same cores was also measured and used to correct shipboard measurements of Moisture and Density (MAD) porosity for the effect of the water bound in the interlayer clay particles, thereby calculating interstitial porosity. The results indicate that the porosities estimated by our method are in agreement with these MAD_derived interstitial porosities in several cores acquired from the overthrusted sediments above the Japan trench plate boundary. Also, considering interstitial porosity as a realistic measurement of porosity, the results show that our semi-automatic method improves estimations compared with a manual thresholding segmentation, as the latter suffers from user subjectivity.

1 Introduction

In March 2011, the Mw ~9 Tohoku-oki earthquake, one of the largest seismic events ever recorded, occurred across a megathrust fault in the west of the Japan trench (Fig. 1). During the earthquake, the seismogenic region of the plate boundary between the Eurasian plate and subducting Pacific Plate was broken such that an unprecedented coseismic slip of about 50 to 60 m occurred over the shallow part of the megathrust (Fujiwara et al., 2011; Lay et al., 2011; Tanikawa et al., 2013; Chester et al., 2013; Sun; 2017). This devastating earthquake stressed the need for more detailed studies on the fault zone behavior and the main causes of this event. Consequently, expedition 343, also known as Japan Trench Fast Drilling (JFAST) project started at IODP site C0019 with the aim of acquiring and estimating the main physical properties of the fault zone.

Several holes were drilled during this expedition (Fig. 1). Logging while drilling (LWD) data were successfully recovered while drilling hole C0019B to the depth of 850 meters below seafloor (mbsf). The primary purpose for drilling hole C0019D up to 854.8 mbsf was to install observatory instruments to monitor wellbore temperature. Furthermore, drilling cores were acquired from hole C0019E, which was drilled until 844.5 mbsf, close to C0019B (about 10m south of this hole).

It is now well established that fluids can greatly affect fault strength and consequently trigger an earthquake. Porosity is a fundamental property of rocks that can help to understand tectonic processes and fault zone behavior as it can serve as an indicator of Coseismic volumetric strain and fluid pressure anomaly around the fault zone (Wang & Barbour, 2017). It is also a critical parameter for studying pore pressure, sediment compaction, and effective stresses, including their evolution due to thermal (fluid) pressurization (Sibson, 1973; Lachenbruch, 1980), which play an important role in rupture initiation and fault slip (Moore & Saffer, 2001; Saffer & Bekins, 2006). Therefore, it comes as no surprise that several methods have been developed to estimate porosity in the past decades, including laboratory measurements (e.g., AD and mercury intrusion porosimetry) and well-logging to imaging techniques.

X-ray computed tomography (XCT) is a non-destructive and now a mature tool frequently used in porous media to characterize the hydraulic properties of samples such as porosity (Van Geet et al., 2003; Taud et al., 2005; Nehler et al., 2019), pore geometry and structure (Mukunoki et al., 2016; Wildenschild & Sheppard, 2013), fracture network characterization, and permeability estimation (Okabe & Blunt, 2004; Ketcham et al., 2010; Mostaghimi et al., 2013; Peng et al., 2014). However, most of these studies were performed on sandstones or rocks with large grains rather than clays or clayey rocks having sub-micron feature size (i.e., grain or pore size) which is usually behind the resolution of most XCT instruments, and high-resolution imaging is both expensive and time-consuming (Kaufhold et al., 2016).

This paper proposes a new method to estimate the porosity of clayey rocks using low-resolution XCT images of the core samples acquired from hole C0019E during IODP expedition 343. Pixel size (or resolution) of images acquired in this expedition is 0.18 μm, which is far larger than the typical pore size range of clayey rocks of less than 2 nm (micropores) or in the range of 2-50 nm (mesopores) (Kuila & Prasad, 2013). Despite these limitations, it is shown that the proposed data-driven segmentation approach based on Otsu's method (Otsu, 1979), also known as the k-mean algorithm, can be used to quantify rock porosity. Our results are validated by an independent quantification of porosity using Cation Exchange Capacity (CEC) laboratory measurements on the squeezed samples. Furthermore, to examine the problem of user bias, the results of the proposed approach are compared to those obtained from manual thresholding segmentation by the commercial software Avizo® (FEI company).

2 Materials and methods

The overall structure of the region and location of hole C00019E can be seen in the seismic profile shown in Fig. 2. While drilling this hole, a rotary core barrel (RCB) system was used to obtain 21 cores, representing a total length of 55m, from the prism, décollement, and underlying Pacific plate. We focus on the décollement and the fault zone, so we only analyzed the last

ten cores (Cores 12R-21R). A medical CT scanner (GE Medical System LightSpeed Ultra16) was used in IODP expedition

65 343 to scan the cores. Dimensions of the images parallel to the planes perpendicular to the core axis were 512 × 512 pixels, and each pixel had a size of 0.188mm × 0.188mm.

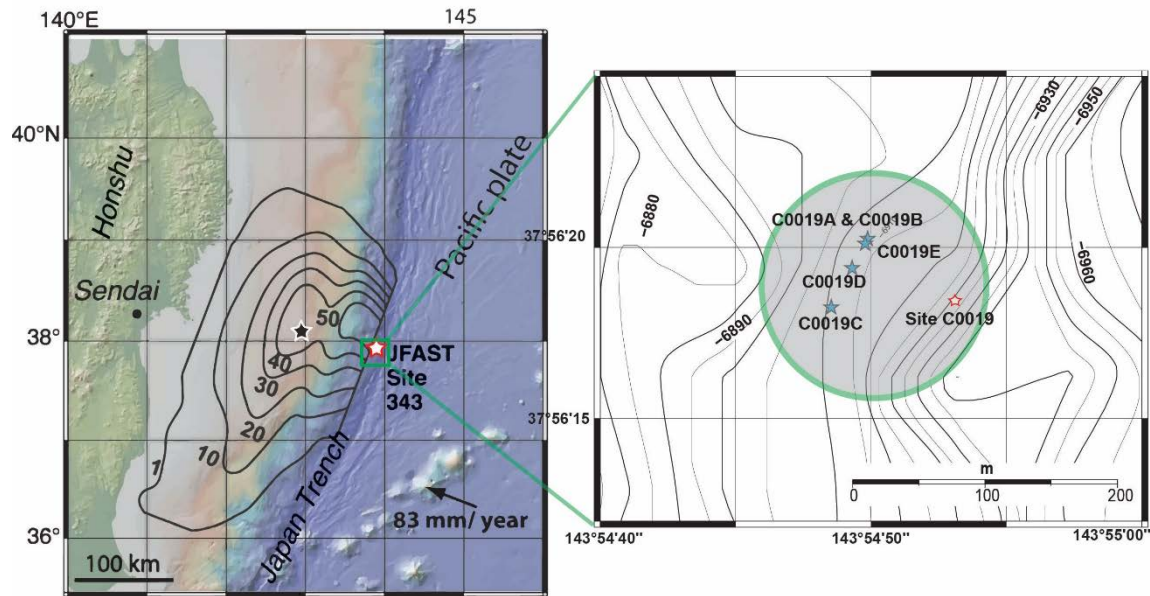



Figure 1: Location of JFAST site (IODP expedition 343); values between slip contours are in meters (modified after Chester et al.,
70 **2013; Kirkpatrick, 2015).**

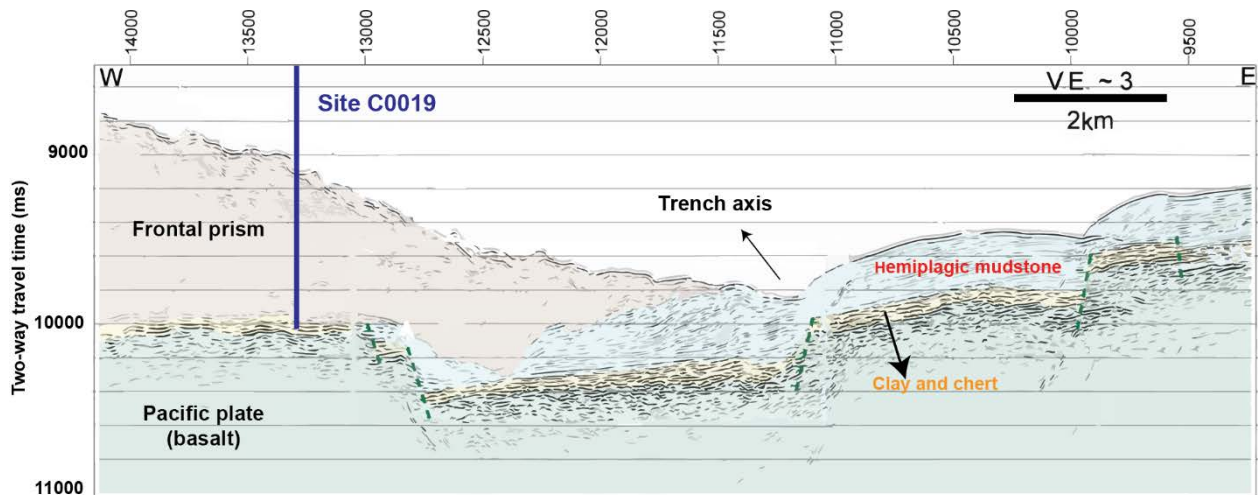



Figure 2: Seismic section showing the main structures in the Japan Trench plate boundary. Normal faults are illustrated with black arrows (modified after Nakamura et al., 2013).

75   Based on the previous studies during IODP Expedition 343, the frontal prism mainly consists of fairly uniform silty clays grading into claystone (Chester et al., 2013; Chester & Moore, 2018). Cores 2R-16R were acquired from the frontal prism down to 820 mbsf (Fig. 3). From the variations of gamma-ray and resistivity logs combined with bedding analysis from resistivity image logs, the plate boundary (décollement) was determined to lie at a depth of ~821 mbsf, consistent with the results of the seismic surveys (Nakamura et al., 2013; Boston et al., 2014). Core 17R, recovered from the décollement interval,

80   is composed of highly deformed clay with scaly fabric (Chester et al., 2013; Kirkpatrick et al., 2015), and high content of smectite (60-80 wt%) according to X-ray diffraction analysis (Kameda et al., 2015). Analysis of cores 18R-21R, recovered from the Pacific plate at the bottom of the hole (from 824 mbsf to 837 mbsf), showed that the Pacific plate is mostly composed of brown mudstone with silicoclastic grains (with more than 60% clay) and pelagic clay at the lower portion. Also, low gamma-ray and high resistivity may correspond to chert, which has also been observed at site DSPD 436 (Party, 1980). Fig. 3 displays

85   lithologies, locations of cores, and the physical logs acquired at the fault zone.

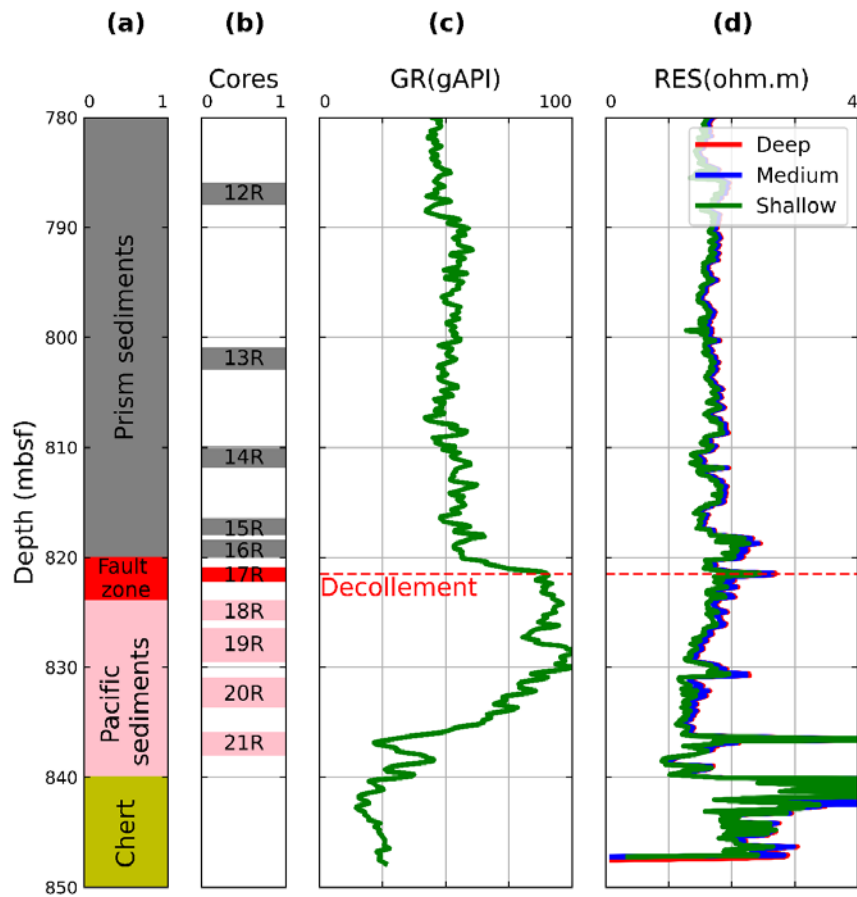



Figure 3 (a) stratigraphy and structural summary of IODP Expedition 343 Hole C00019E, (b) location of cores, (c) Gamma-ray, and (d) resistivity logs of this interval. Deep, medium, and shallow refer to the depths of investigation.

2.1 XCT image analysis

90   When an X-ray passes through an object, the ray might be scattered or absorbed, resulting in signal attenuation. This attenuation or intensity decrease is classically described by Beer's Law. For a monochromatic beam in a homogeneous material, this law is as follows:

$$I = I_0 \exp\left(-\mu l\right),\tag{1}$$

95   Where I is X-ray intensity, I_0 is initial intensity, μ is the Linear Attenuation Coefficient (LAC) of sample material, and l is the length of the X-ray path. LAC is dependent on the density and mineral composition of the material. CT number is a parameter used to evaluate the attenuation, defined as:

$$CT_n = \frac{\mu_t - \mu_w}{\mu_w} \times 1000,\tag{2}$$

Where μ_t and μ_w are linear attenuation coefficients of target material and water, respectively.

100   These attenuations are then mapped onto consecutive 2D slices which are combined to give attenuation values in voxels (i.e., volumetric picture elements). Generally, the higher the CT number of a voxel, the brighter its greyshade the color in CT images. According to Eq. (2), a higher CT number implies higher attenuation in the material which can be due to either higher density or presence of elements with higher atomic numbers. On the other hand, darker colors are interpreted as pores and fractures where almost no attenuation is recorded compared with the solid phases (Ketcham and Carlson, 2001).

105   Although XCT imaging technique is a quick and non-invasive method, it has some limitations related to the presence of different artifacts, the partial volume effect, and the scanner resolution. Since a voxel volume might consist of several materials, the X-ray attenuation assigned to each voxel is, in fact, the average attenuation of several materials. In addition, a value assigned to a given voxel will be affected by the surrounding voxels so that sharp passages between phases with different attenuation coefficients will be smoothed to some extent. This phenomenon is known as the partial volume effect (Iassonov et

110   al., 2009; Ketcham & Carlson, 2001).

Porosity estimation or any other quantitative image analysis comprises three main steps: 1) image preprocessing (e.g., artifact removal and image filtering), 2) image segmentation, 3) object quantification. However, the analyzed core samples were partially damaged during the coring process and their transfer to the surface. Therefore, for this analysis, we first selected the portions of each core with no artificial damages.

115   ### 2.1.1 Image filtering

The first step in XCT image analysis is to apply an appropriate filter to reduce the noise level in the grey-scale images. This noise can significantly affect the quality of the image and the results of the subsequent steps.

To avoid over-smoothing the raw images and preserve the edges (i.e., the interface between two materials), a non-local mean filter (NLMean) was employed to eliminate the noise in the images. This filter calculates the similarity between a block of

120   interest (BOI) centered at a target pixel and all the neighborhood blocks (reference blocks) in a search area. A square Euclidean distance between pixel values in the BOI and the reference blocks is used as a measure of dissimilarity. Then, the target pixel is replaced by a weighted mean value of center pixels in the reference blocks with higher similarity (i.e., lower distance). The weight is calculated by applying a Gauss kernel on the similarity values. As a result, more details are preserved using this filter compared to the conventional median filter (Buades et al., 2005; Sarker et al., 2012; Matsuoka et al., 2017).

2.1.2 Image segmentation

125   The term segmentation can be defined as a process of classifying each voxel (or a set of voxels) as a material type called a phase (e.g., pores and minerals in geomaterials). If there is only one phase to be separated from the matrix, the process is called binary segmentation. In this process all voxels' values of interest are replaced with a binary value of 1 (e.g., pores in porosity evaluation), with the rest of voxels (e.g., solid phase) being assigned a value of 0. Segmentation is of vital importance as it

130   greatly affects all the subsequent interpretations, analyses, and modeling performed on segmented images (Kaestner et al., 2008; Iassonov et al., 2009).

Several methods currently exist for image segmentation. Global thresholding is one of the simplest and most frequently used methods in this regard. This method, which is based on the analysis of an intensity histogram, having defined a value as a threshold, pixels with values less and higher than the threshold are considered different phases.

135   Two approaches for image segmentation are employed in this work. First, the AVIZO® software is used, which provides an interactive thresholding panel with several tools for segmentation that can be combined to separate pores from solid. However, since the threshold is visually determined in this method, results can be highly biased by the user, leading to over/under-estimation of porosity (Chauhan et al., 2016; Iassonov et al., 2009; Kaestner et al., 2008). Fig. 4 shows the threshold value and segmented (binary) image obtained for one of the cores. To reduce the bias in the manual thresholding method, a second

140   approach based on the K-means clustering algorithm was adopted. The section below describes the different steps of this method.

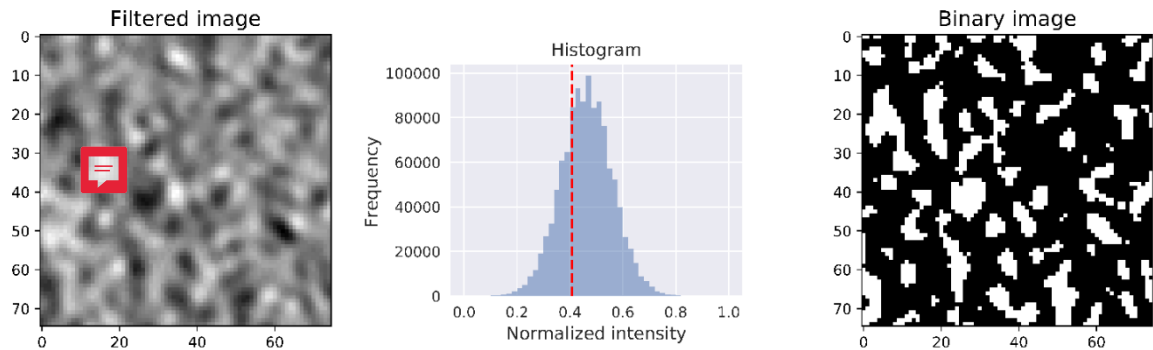



Figure 4 Segmentation using interactive thresholding method (Left: the grayscale image of a slice of core 14R2 after applying the non-local mean filter; Middle: histogram of pixel values in all slices of the same core; and Right: binary image in which white pixels

145   **indicate pores and black ones are solid.**

- **Image segmentation with the k-means algorithm**

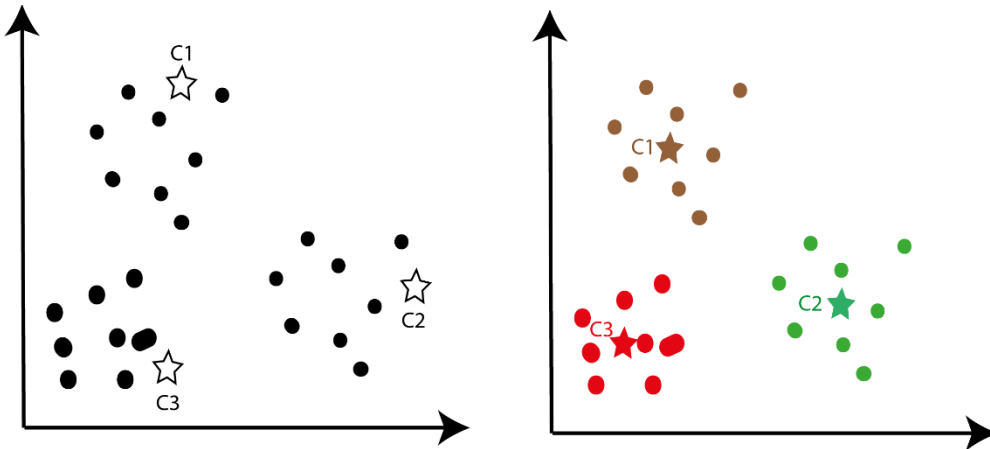
The amount of porosity obtained from CT images is largely dependent on the threshold applied, so that a slight change in the threshold might result in a significant error in the final porosity estimation. Hence, a clustering algorithm was used to identify the threshold in a semi-automatic manner. The main aim of clustering in segmentation is to group the pixel values into different

150 groups so that the pixels in the same group are as similar as possible and different from the pixel values in other groups (Ojeda-Magana et al., 2009; Cortina-Januchset al., 2011).

The K-means clustering algorithm, which was primarily introduced by MacQueen (1967), is one of the simplest and the most conventional clustering method utilized by several researchers to solve segmentation problems (Taud et al., 2005; Cortina-Januchs et al., 2011; Chauhan et al., 2016; Al-Marzouqi, 2018; Guntoro et al., 2019). In this algorithm, first, k number of

155 cluster centroids are randomly selected in feature space (i.e., pixel intensity values) (e.g., k=3 in Fig. 5). Then, each data point (pixel intensity) is assigned to the nearest centroid to create a cluster. In the next step, each centroid center Cj iteratively moves to the mean value of the pixels within that cluster (μ_j), and the data are reassigned to the nearest centroids which can be different now. In each iteration, inertia or sum of square error within all clusters are calculated, and the algorithm continues updating centroid center until the inertia is minimized:

160 $$Inertia = \sum_{i=1}^{n}\sum_{j=1}^{k}\left\|x_i - C_j\right\|^2 , \qquad (3)$$

Where, n and k are the number of pixels and clusters, respectively. x_i is i[th] pixel value and C_j indicate j[th] cluster center.

165 **Figure 5 Illustration of K-means algorithm for three clusters. Left image shows the first step when the centroids (in this example three centroids) are randomly determined in a two dimension space. In the right image, each centroid has moved to the mean of the cluster after a number of iterations.**

The clustering results in the k-means algorithm highly depend on the initial positions of the centroids in the data space. Hence, the algorithm was run 20 times, and the best result (i.e., the one with the minimum inertia) was chosen.

The main drawback of the k-means algorithm is that the number of cluster k must be predefined. A fast and straightforward method is to determine k is to run the algorithm for a different number of clusters and calculate the cost function (inertia) each time (Ward Jr, 1963). Then, the optimum number of clusters will be the one that shows an elbow when plotting the inertia versus the number of clusters. In other words, inertia drops at a certain number of clusters and does not decrease much after that.

Figure 6 presents the inertia values versus the number of clusters between 2 and 10. The figure shows no distinct elbow due to the low resolution of the CT images. One of the distinguishing characteristics of low-resolution images is that the distribution of pixel intensities is unimodal with long tails. This unimodal distribution was observed in all the images. Moreover, a negative skewness in the histogram (Fig. 6) shows that the image resolution is not adequate to separate the pores from the matrix, i.e., the pores sizes are mostly below the image resolution (Nehler et al., 2019).

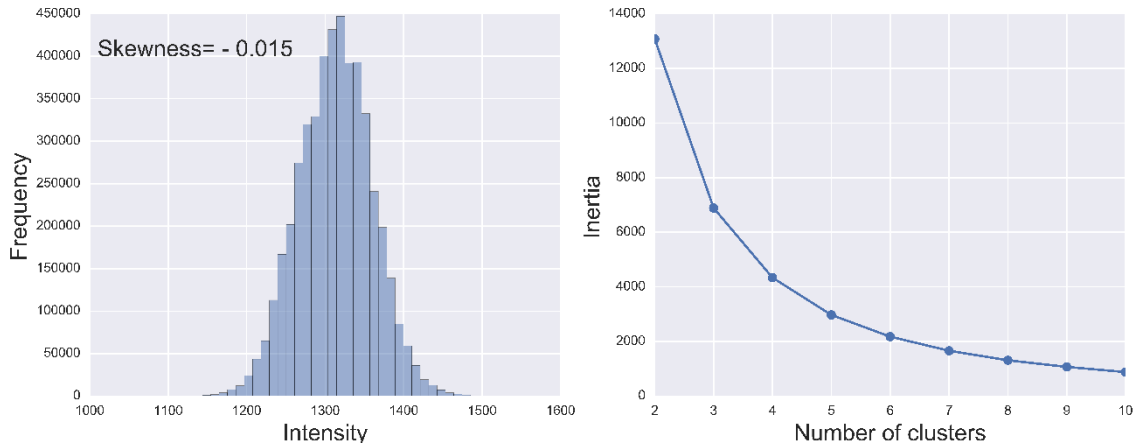



Figure 6 Histogram of voxels' intensity of core 19R3 indicates a negatively skewed distribution (left); in such a situation, there is no distinct elbow shape in the inertia plot (right).

This low resolution will increase the risk of underestimation (or overestimation) of porosity or other subsequent quantification on the binary images. For this reason, and also because the results of segmentation by visual inspection were unsatisfactory, the number of the clusters was further explored with another index known as the Davies-Bouldin index (Davies & Bouldin, 1979).

Generally, the idea behind the different clustering approaches is to identify different groups in a given dataset. The more separated are the clusters, the better is the clustering. Furthermore, the data points belonging to each cluster should be as close as possible in terms of intensity values (Halkidi et al., 2001).

Davies-Bouldin (DB) index is a measure of similarity between each cluster and the most similar one to it, C_j $(i, j = 1, ..., k)$. The index is defined as follows:

$$DB = \frac{1}{K}\sum_{i=1}^{k} \max R_{ij} \quad (i \neq j),$$ (4)

195 R_{ij} in this equation is calculated as:

$$R_{ij} = \frac{s_i+s_j}{d_{ij}},$$ (5)

Where s_i is cluster diameter, i.e., the average between the centroid of the cluster and its each datapoint i. d_{ij} is defined as the distance between two cluster centroids i and j.

Lower values for this index mean that the clusters are better separated from the other clusters, and the data points in each
200 cluster are less dispersed (Halkidi et al., 2001). In this study, both inertia and DB indices are used to explore the possibility of merging some clusters. Our approach can be summarized as follows:

1) For each core, k-means algorithm was run with a different number of clusters (between 2 and 10, and inertia and DB indices were calculated (Figs 7a, 7c, and 7e).

2) Unlike inertia, the DB index shows an elbow shape on the profile after a certain number of clusters K_{DB}. For
205 example, this index drops sharply after 4 clusters (K_{DB}= 4) in core 19R2 (Fig. 7a). Segmentation is then done by merging the different clusters after the K_{DB}.

3) The results of segmentation are converted to the threshold value in the intensity histogram of pixels. For example, if the data are grouped into 6 clusters (clusters are sorted in ascending order according to their centroids), and three of them are merged, the maximum pixel value in the third cluster indicates the threshold value (e.g., orange
210 dashed line in Fig. 7b). On the other hand, one can calculate a threshold value that corresponds to a given porosity by an iterative approach, i.e., by changing the threshold value, segmenting the image, and calculating the porosity based on that value, the threshold porosity is obtained. Therefore, the maximum threshold value is the one that corresponds to the total porosity (maximum porosity) measured using the MAD method on board in expedition 343 (Table 1).

215 4) The next step involves selecting close thresholds. This step can be thought of as another clustering step, but this time the threshold values are clustered instead of pixel values. For example, in Fig. 7d, one can see that the threshold corresponding to 4 clusters coincides with 9 clusters (when 3 of them are merged). In the same way, segmentation with 3 clusters gives the very close threshold to 10 clusters (4 merged) in the core 17R illustrated in Fig. 7d.

220 5) However, this method might provide more than one choice for the threshold (for example, in plots d and f in Fig. 7). In these cases, the final threshold is selected by visual inspection of the binary image.

6) Our approach can be seen as a semi-automatic segmentation in which the best candidates are identified using the k-means algorithm. Visual inspection is only a determining step if the situation in step 5 arises. This approach will reduce the effect of user subjectivity in the segmentation. All the calculations related to clustering are done
225 using the Scikit-learn python package (Pedregosa et al., 2011).

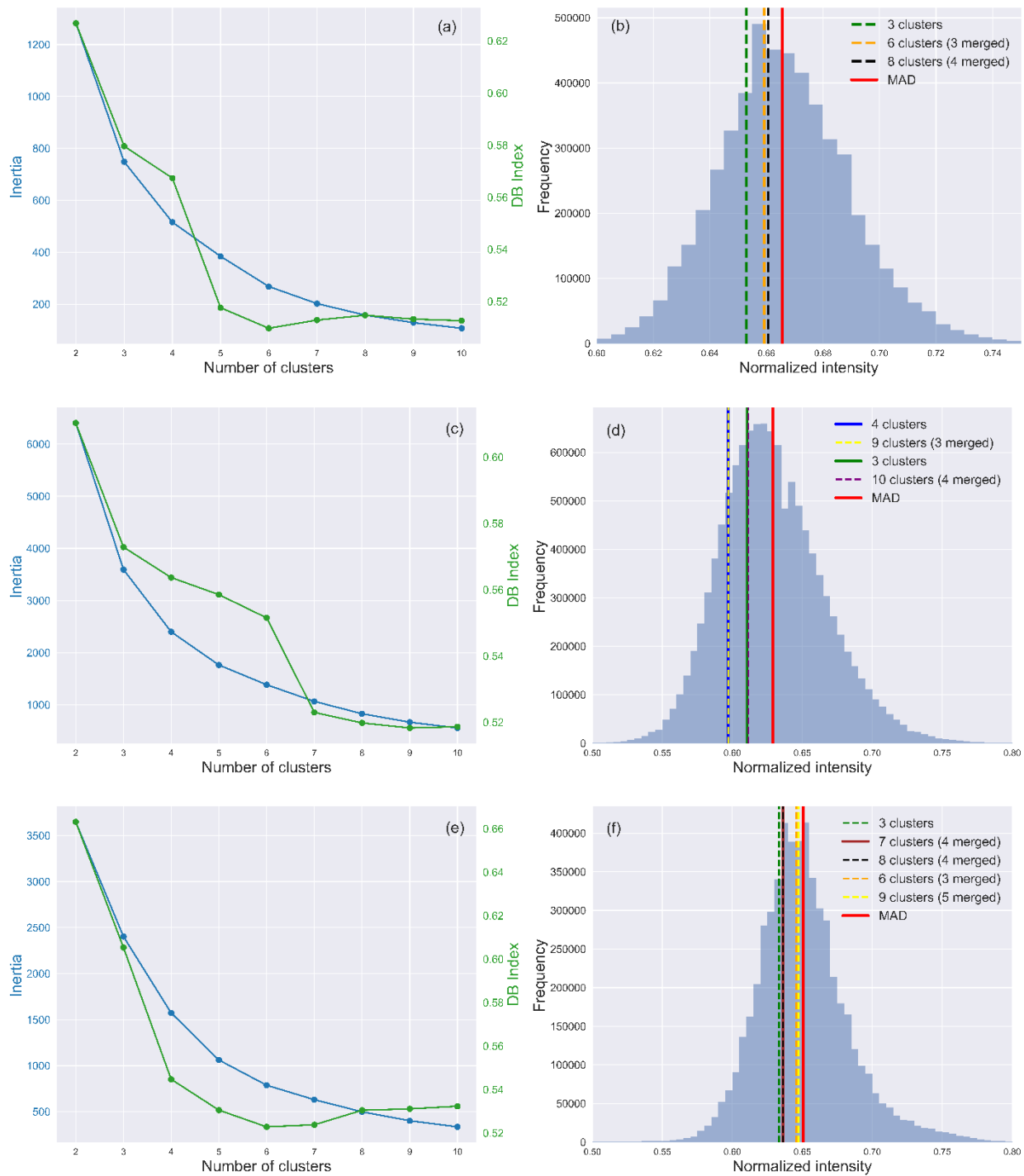



Figure 7 Determining thresholds for segmentation by merging the clusters with the lowest DB index. Results are shown for core 19R2 (a and b), 17R (c and d), and 15R-1 (e and f).

2.2 Interstitial porosity from CEC

As described earlier, porosity estimated with the MAD method refers to the total porosity of the sample. However, a fraction of this porosity relates to the water present in clay minerals (within the clays and adsorbed at their surface. Therefore, cation capacity exchange (CEC) was employed to correct the total porosity and obtain the interstitial porosity corresponding to the water contained within the pores. The resulting porosity can also be compared with that measured on CT scans. However, to estimate water-bound porosity, other imaging techniques such as transmission electron microscopy (TEM) and focused ion beam combined with scanning electron microscopy (FIB-SEM) are required to obtain images at submicron resolution. Unfortunately, such scanning resolutions were not available during IODP expedition 343.

The water present in the minerals is calculated using CEC measurements on the core samples, and total porosity is corrected using the following equation:

$$\phi_i = \phi_t - n\frac{m_w}{\rho_w}.CEC.\rho_g(1 - \phi_t) \tag{6}$$

Where ϕ_t and ϕ_i are total (i.e., MAD porosity) and interstitial porosity, respectively, m_w is water molar mass (0.018 kg mol^{-1}), ρ_g and ρ_w are grain and water density, respectively, which are considered to be 2650 and 1024 kg m^{-3}, CEC is the Cation Exchange Capacity, expressed in moles per kilogram of the sample in dry condition, and n is the average number of water molecules per cation charge. In theory, each cation charge present in a hydrated smectite interlayer is associated with 15 water molecules (Ransom & Helgeson, 1994).

CEC is defined as the total amount of the ion extracted after mixing a given mass of sample with a specific volume of a solution. Several methods exist for measuring CEC, all of which involve the complete saturation of the sample with a mono-ionic solution. CEC is then calculated by measuring the difference between the ion concentration before and after the experiment (Aran et al., 2008).

In the present study, CEC values were quantified by exchanging the cations with cobaltyhexammine chloride (Conin et al., 2011; Henry & Bourlange, 2004). Measurements were performed at the Institute National de Recherche Agronomique (INRA) laboratory in Arras (France).

3 Results

Figure 7 shows the results of clustering analysis to determine the best threshold for the segmentation of core images. As described earlier, when two candidates exist for the segmentation, the best one is selected by visual inspection. For example, Figure 8 presents the segmentation results on a cropped slice of core 17R using the two thresholding candidates shown in Figure 7d. In this case, 10 clusters (of which four merged) are selected as the best for thresholding.

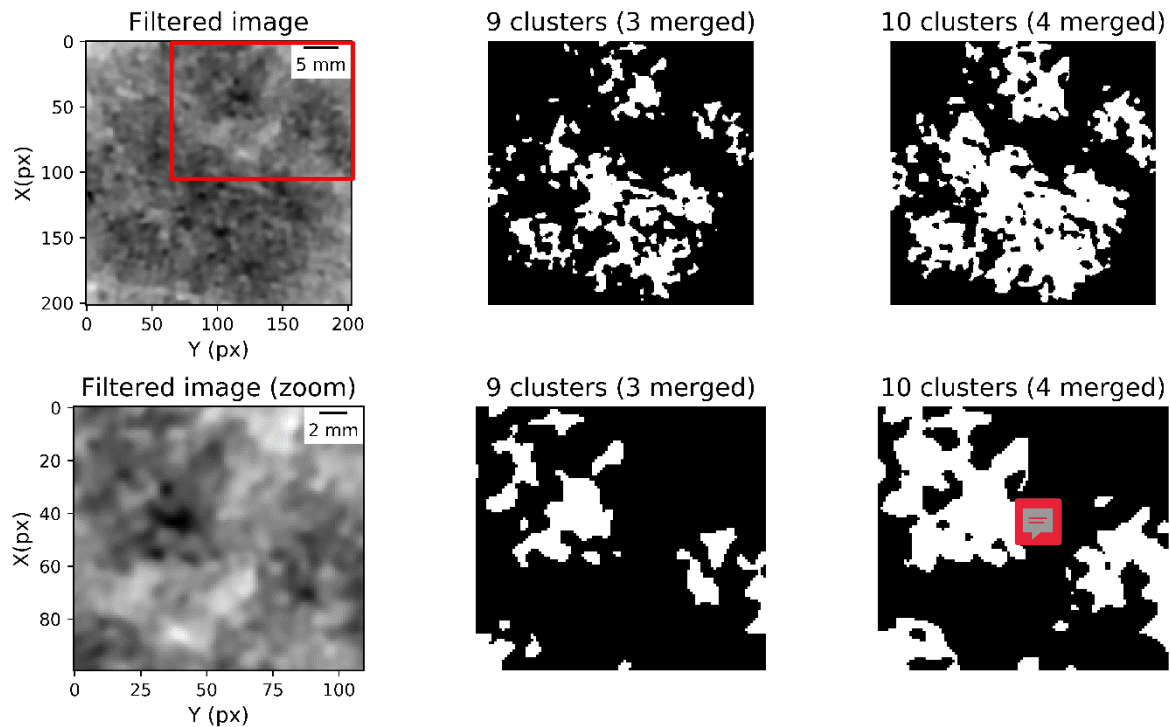



Figure 8. Results of segmentation in core 17R. Lower plots illustrate zoom within the red box. Comparing the binary images suggests
260 **that segmentation with 10 clusters (4 merged) gives the best results.**

Table 1 presents the results of CEC measurements and interstitial porosity estimated by Eq. 6. The CEC values range from 0.19 to 0.47 meq/g, with the maximum value for core 17R1 taken from the decollement. The average difference between total porosity (MAD) and interstitial porosity is 8%, while the maximum difference is 14.8% for the decollement, which was expected because of the high content of smectite (see Kameda et al., 2017).

265 To obtain a continuous estimation of porosity and compare the results with other methods, interstitial and MAD porosities were linearly interpolated. Figure 9 shows the results of porosity estimation with different methods (third column). Figure 10 compares the results of interstitial porosity measurements with estimations using global thresholding (Figure 10a) and the proposed method in this study (Figure 10b). The correlation coefficient, R-squared value, and root mean square error (RMSE) indicate a clear improvement in porosity estimation using the clustering approach proposed in this study.

270 **4 Discussion**

The initial objective of this study was to propose a method for the estimation of the porosity of clayey rocks using low-resolution XCT images. These rocks have small pores (i.e., micropores and mesopores), and this necessitates using high-resolution imaging techniques to capture these small pores. However, high-resolution imaging instruments are expensive and not always available in many research projects.

Table 1. Results of CEC measurements and comparison of the total (MAD) and interstitial porosity

Core ID	Depth (mbsf)	CEC (meq/g)	Interstitial porosity (%)	MAD (%)
12R1	786.1	0.19	35	42.9
12R2	786.9	0.19	34.6	42.6
13R1	801.5	0.23	37.5	46.4
13R2	801.9	0.23	38.7	47.4
14R1	810.3	0.24	36.9	46.2
14R2	811.2	0.24	38.3	47.5
15R1	816.9	0.25	44.1	52.7
16R1	818.6	0.314	29.2	42.4
17R1	822.5	0.467	30.5	48.2
18R1	825.5	0.347	34.3	47.5
19R1	826.9	0.289	22.4	35.9
19R2	828	0.245	37.9	47.3
19R3	828.7	0.213	38.1	46.4
20R1	831.1	0.210	45.3	52.5
20R2	832.4	0.210	42.6	50.2
21R1	836.8	0.210	25	35

275

The overall results presented in Figure 9 indicate that, considering interstitial porosity calculated by CEC measurements to be the ground truth (red markers), merging clusters in Otsu's method provides more consistent estimations of porosity (black markers) than MAD-derived porosity measurements (green markers) made during expedition 343 (Chester et al., 2013). Furthermore, it can be seen that the results of the three methods follow a similar trend: lower porosity in the decollement with
280   an increase towards the prism and Pacific sediments. This relatively low porosity in the decollement may be explained by the shear-enhanced compaction in the fault core which has formed a distinct scaly fabric observed in core in 17R (Chester et al., 2013). It can also be observed that MAD porosity is higher than our estimations in all the cores as the MAD gives total porosity including water absorbed on the surface of clay minerals. Furthermore, we showed that our semi-automatic approach could significantly reduce errors due to user bias, which are common in manual thresholding approaches for the segmentation of
285   XCT images (Figure 10). This is important because a bias in the segmentation of images will lead to an error in the subsequent analysis of pore network modeling.

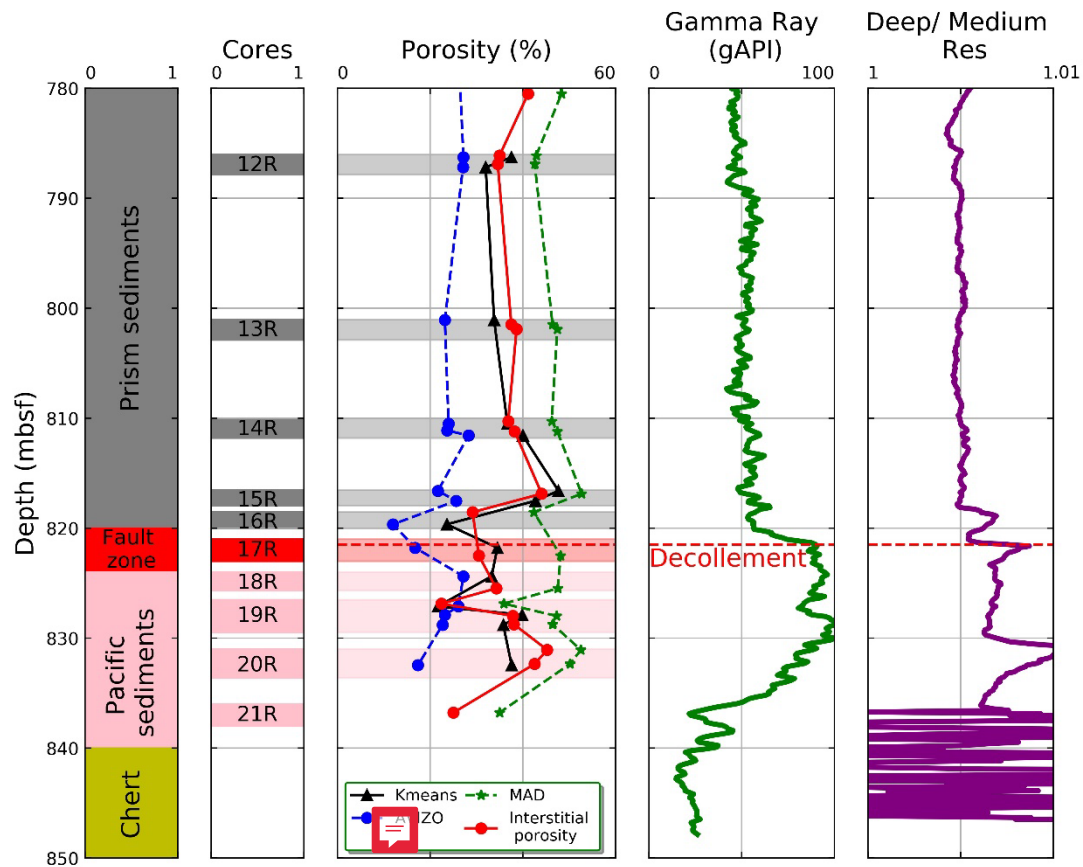



Figure 9. Porosity estimations using different methods; the first and second columns indicate the lithology and the location of core samples. The third column presents porosity calculated by different methods. The fourth and fifth columns show the gamma-ray and the ratio of deep over medium resistivity, respectively.

290

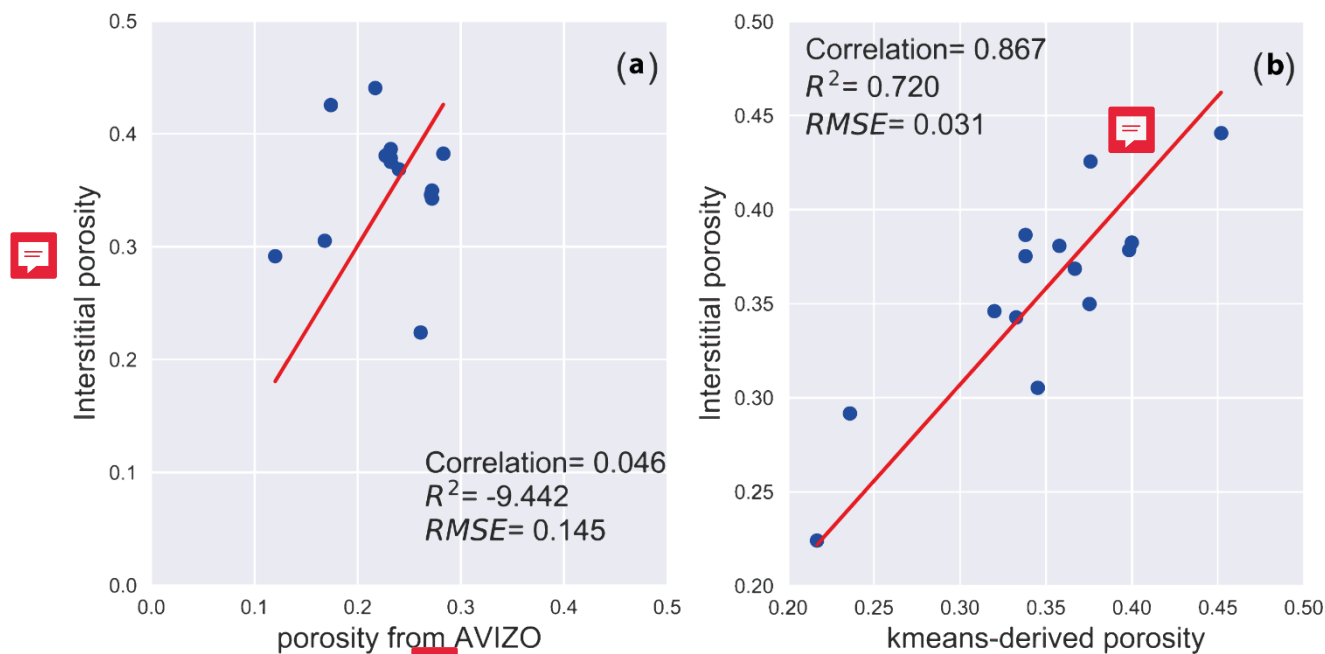



Figure 10. Comparison between the results of manual thresholding (a) and the proposed method in this study (b). The amount of porosity estimated using these methods is compared to interstitial porosity. The correlation coefficient, r-squared, and RMSE shows an enhanced estimate of porosity using the k-means method.

5 Conclusions

The main aim of this study was to propose a new approach for the segmentation of low-resolution XCT images. K-means is one of the simplest and fastest clustering algorithms which can be used for automatic segmentation of XCT images. However, its application is limited to high-resolution images and samples containing minerals with distinct attenuation coefficients. Our investigation suggests that merging clusters in k-means algorithm using both inertia and DB index can result in more accurate and unbiased segmentation when compared to independent experimental measurements. Although this study is only based on the XCT images of core samples from IODP expedition 343, our findings give new insights into dealing with low-resolution images. This will be of interest to researchers who do not have access to a high-resolution XCT scanner. However, to evaluate the generalizability of the proposed method, further research should be carried out on rock samples of different compositions. A reasonable approach could be to apply our method to XCT images of the same sample obtained at different resolutions.

Author contributions

MC performed CEC measurements and obtained interstitial porosity. HA and FC processed XCT images. HA wrote the manuscript draft. MC, MD, FC, and VT reviewed the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

Support for this work was provided by Centre National de la recherche scientifique (CNRS). This research was conducted on samples and data publicly provided by the Integrated Ocean Drilling Program (IODP) Expedition 343. We would like to thank Institut National de Recherche Agronomique (INRA) for CEC measurements.

315 **References**

Al-Marzouqi, H. (2018). Digital rock physics: Using CT scans to compute rock properties. IEEE Signal Processing Magazine, 35(2), 121–131. https://doi.org/10.1109/MSP.2017.2784459

Aran, D., Maul, A., & Masfaraud, J.-F. F. (2008). A spectrophotometric measurement of soil cation exchange capacity based on cobaltihexamine chloride absorbance. *Comptes Rendus Geoscience*, *340*(12), 865–871. 320 https://doi.org/10.1016/j.crte.2008.07.015

Boston, B., Moore, G. F., Nakamura, Y., & Kodaira, S. (2014). Outer-rise normal fault development and influence on near-trench décollement propagation along the Japan Trench, off Tohoku. *Earth, Planets and Space*, *66*(1), 135.

Buades, A., Coll, B., & Morel, J.-M. (2005). A non-local algorithm for image denoising. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, *2*, 60–65.

325 Chauhan, S., Rühaak, W., Anbergen, H., Kabdenov, A., Freise, M., Wille, T., & Sass, I. (2016). Phase segmentation of X-ray computer tomography rock images using machine learning techniques: an accuracy and performance study. *Solid Earth*, *7*(4), 1125–1139. https://doi.org/10.5194/se-7-1125-2016

Chester, F. M., Mori, J. J., Eguchi, N., Toczko, S., Kido, Y., Saito, S., Sanada, Y., Anderson, L., Behrmann, J. H., Bose, S., Conin, M., Cook, B., Fulton, P., Hirose, T., Ikari, M., Ishikmawa, T., Jeppson, T., Kameda, J., Kirkpatrick, J., … Tao, Y. 330 (2013). Expedition 343/343T summary. *Proceedings of the Integrated Ocean Drilling Program; Japan Trench Fast Drilling Project (JFAST); Expedition 343/343T of the Riser Drilling Platform, Shimizu, Japan, to Sendai, Japan; Site C0019; 1 April-24 May 2012; and Minami-Ise, Japan, to Hachinohe, J*, *343*, 28. http://hdl.handle.net/10.2204/iodp.proc.343343T.101.2013

Conin, M., Henry, P., Bourlange, S., Raimbourg, H., & Reuschlé, T. (2011). Interpretation of porosity and LWD resistivity from the Nankai accretionary wedge in light of clay physicochemical properties: Evidence for erosion and local overpressuring. 335 *Geochemistry, Geophysics, Geosystems*, *12*(3), 1–17. https://doi.org/10.1029/2010GC003381

Cortina-Januchs, M. G., Quintanilla-Dominguez, J., Vega-Corona, A., Tarquis, A. M., & Andina, D. (2011). Detection of pore space in CT soil images using artificial neural networks. *Biogeosciences*, *8*(2), 279. https://doi.org/10.5194/bg-8-279-2011

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-1*(2), 224–227. https://doi.org/10.1109/TPAMI.1979.4766909

Fujiwara, T., Kodaira, S., No, T., Kaiho, Y., Takahashi, N., & Kaneda, Y. (2011). The 2011 Tohoku-Oki earthquake: Displacement reaching the trench axis. *Science*, *334*(6060), 1240. https://doi.org/10.1126/science.1211554

Guntoro, P. I., Ghorbani, Y., Koch, P.-H., & Rosenkranz, J. (2019). X-ray Microcomputed Tomography (µCT) for Mineral Characterization: A Review of Data Analysis Methods. *Minerals*, *9*(3), 183. https://doi.org/10.3390/min9030183

Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, *17*(2–3), 107–145. https://doi.org/10.1023/A:1012801612483

Henry, P., & Bourlange, S. (2004). Smectite and fluid budget at Nankai ODP sites derived from cation exchange capacity. *Earth and Planetary Science Letters*, *219*(1–2), 129–145. https://doi.org/10.1016/S0012-821X(03)00694-0

Iassonov, P., Gebrenegus, T., & Tuller, M. (2009). Segmentation of X-ray computed tomography images of porous materials: A crucial step for characterization and quantitative analysis of pore structures. *Water Resources Research*, *45*(9), 1–12. https://doi.org/10.1029/2009WR008087

Kaestner, A., Lehmann, E., & Stampanoni, M. (2008). Imaging and image processing in porous media research. *Advances in Water Resources*, *31*(9), 1174–1187. https://doi.org/10.1016/j.advwatres.2008.01.022

Kameda, J., Inoue, S., Tanikawa, W., Yamaguchi, A., Hamada, Y., Hashimoto, Y., & Kimura, G. (2017). Alteration and dehydration of subducting oceanic crust within subduction zones: implications for décollement step-down and plate-boundary seismogenesis. *Earth, Planets and Space*, *69*(1), 52.

Kameda, J., Shimizu, M., Ujiie, K., Hirose, T., Ikari, M., Mori, J. J., Oohashi, K., & Kimura, G. (2015). Pelagic smectite as an important factor in tsunamigenic slip along the Japan Trench. *Geology (Boulder)*, *43*(2), 155. http://hdl.handle.net/10.1130/G35948.1

Kaufhold, A., Halisch, M., Zacher, G., & Kaufhold, S. (2016). X-ray computed tomography investigation of structures in Opalinus Clay from large-scale to small-scale after mechanical testing. *Solid Earth*, *7*(4), 1171–1183.

Ketcham, R. A., & Carlson, W. D. (2001). Acquisition, optimization and interpretation of x-ray computed tomographic imagery: Applications to the geosciences. *Computers & Geosciences*, *27*(4), 381–400. https://doi.org/10.1016/S0098-3004(00)00116-3

Ketcham, R. A., Slottke, D. T., Sharp, J. M., & Sharp Jr, J. M. (2010). Three-dimensional measurement of fractures in heterogeneous materials using high-resolution X-ray computed tomography. *Geosphere*, *6*(5), 499–514. https://doi.org/10.1130/GES00552.1

Kirkpatrick, J. D., Rowe, C. D., Ujiie, K., Moore, J. C., Regalla, C., Remitti, F., Toy, V., Wolfson-Schwehr, M., Kameda, J., Bose, S., Chester, F. M., Wolfson-Schwehr, M., Kameda, J., & Bose, S. (2015). Structure and lithology of the Japan Trench subduction plate boundary fault. *Tectonics*, *34*(1), 53–69. https://doi.org/10.1002/2014TC003695

Kuila, U., & Prasad, M. (2013). Specific surface area and pore-size distribution in clays and shales. *Geophysical Prospecting*, *61*(Rock Physics for Reservoir Exploration, Characterisation and Monitoring), 341–362.

Lachenbruch, A. H. (1980). Frictional heating, fluid pressure, and the resistance to fault motion. *Journal of Geophysical Research: Solid Earth*, *85*(B11), 6097–6112.

Lay, T., Ammon, C. J., Kanamori, H., Xue, L., & Kim, M. J. (2011). Possible large near-trench slip during the 2011 M w 9.0 off the Pacific coast of Tohoku Earthquake. *Earth, Planets and Space*, *63*(7), 32. https://doi.org/10.5047/eps.2011.05.033

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, *1*(14), 281–297.

Moore, J. C., & Saffer, D. (2001). Updip limit of the seismogenic zone beneath the accretionary prism of southwest Japan: An effect of diagenetic to low-grade metamorphic processes and increasing effective stress. *Geology*, *29*(2), 183–186.

Mostaghimi, P., Blunt, M. J., & Bijeljic, B. (2013). Computations of absolute permeability on micro-CT images. *Mathematical Geosciences*, *45*(1), 103–125. https://doi.org/10.1007/s11004-012-9431-4

Mukunoki, T., Miyata, Y., Mikami, K., & Shiota, E. (2016). X-ray CT analysis of pore structure in sand. *Solid Earth*, *7*(3), 929–942. https://doi.org/10.5194/se-7-929-2016

Nakamura, Y., Kodaira, S., Miura, S., Regalla, C., & Takahashi, N. (2013). High-resolution seismic imaging in the Japan Trench axis area off Miyagi, northeastern Japan. *Geophysical Research Letters*, *40*(9), 1713. http://hdl.handle.net/10.1002/grl.50364

Nehler, M., Stoeckhert, F., Oelker, A., Renner, J., & Saenger, E. (2019). *Evaluating porosity estimates for sandstones based on X-ray micro-tomographic images*.

Ojeda-Magana, B., Quintanilla-Domínguez, J., Ruelas, R., & Andina, D. (2009). Images sub-segmentation with the PFCM clustering algorithm. *2009 7th IEEE International Conference on Industrial Informatics*, 499–503. https://doi.org/10.1109/INDIN.2009.5195854

Okabe, H., & Blunt, M. J. (2004). Prediction of permeability for porous media reconstructed using multiple-point statistics. *Physical Review E*, *70*(6), 66135. https://doi.org/10.1103/PhysRevE.70.066135

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66.

Party, S. S. (1980). Site 436: Japan trench outer rise, Leg 56. *Initial Reports of the Deep Sea Drilling Project*, *56*(October), 399–446.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.

Peng, S., Marone, F., & Dultz, S. (2014). Resolution effect in X-ray microcomputed tomography imaging and small pore's contribution to permeability for a Berea sandstone. *Journal of Hydrology*, *510*, 403–411. https://doi.org/10.1016/j.jhydrol.2013.12.028

Ransom, B., & Helgeson, H. C. (1994). A chemical and thermodynamic model of aluminous dioctahedral 2: 1 layer clay minerals in diagenetic processes; regular solution representation of interlayer dehydration in smectite. *American Journal of Science*, *294*(4), 449–484.

Saffer, D. M., & Bekins, B. A. (2006). An evaluation of factors influencing pore pressure in accretionary complexes: Implications for taper angle and wedge mechanics. *Journal of Geophysical Research: Solid Earth*, *111*(B4).

Sibson, R. H. (1973). Interactions between temperature and pore-fluid pressure during earthquake faulting and a mechanism for partial or total stress relief. *Nature Physical Science*, *243*(126), 66–68.

Tanikawa, W., Hirose, T., Mukoyoshi, H., Tadai, O., & Lin, W. (2013). Fluid transport properties in sediments and their role in large slip near the surface of the plate boundary fault in the Japan Trench. *Earth and Planetary Science Letters*, *382*, 150–160. https://doi.org/10.1016/j.epsl.2013.08.052

Taud, H., Martinez-Angeles, R., Parrot, J. F., & Hernandez-Escobedo, L. (2005). Porosity estimation method by X-ray computed tomography. *Journal of Petroleum Science and Engineering*, *47*(3–4), 209–217. https://doi.org/10.1016/j.petrol.2005.03.009

Van Geet, M., Lagrou, D., & Swennen, R. (2003). Porosity measurements of sedimentary rocks by means of microfocus X-ray computed tomography (µCT). *Geological Society Special Publication*, *215*(1), 51–60. https://doi.org/10.1144/GSL.SP.2003.215.01.05

Wang, C.-Y., & Barbour, A. J. (2017). Influence of pore pressure change on coseismic volumetric strain. *Earth and Planetary Science Letters*, *475*, 152–159.

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*(301), 236–244.

Wildenschild, D., & Sheppard, A. P. (2013). X-ray imaging and analysis techniques for quantifying pore-scale structure and processes in subsurface porous medium systems. *Advances in Water Resources*, *51*, 217–246.