

Identification of pore spaces in 3D CT soil images using PFCM partitional clustering

B. Ojeda-Magaña, J. Quintanilla-Domínguez, R. Ruelas, A.M. Tarquis, L. Gómez-Barba, D. Andina

ABSTRACT

Recent advances in non-destructive imaging techniques, such as X-ray computed tomography (CT), make it possible to analyse pore space features from the direct visualisation from soil structures. A quantitative characterisation of the three-dimensional solid-pore architecture is important to understand soil mechanics, as they relate to the control of biological, chemical, and physical processes across scales. This analysis technique therefore offers an opportunity to better interpret soil strata, as new and relevant information can be obtained. In this work, we propose an approach to automatically identify the pore structure of a set of 200-2D images that represent slices of an original 3D CT image of a soil sample, which can be accomplished through non-linear enhancement of the pixel grey levels and an image segmentation based on a PFCM (Possibilistic Fuzzy C-Means) algorithm. Once the solids and pore spaces have been identified, the set of 200-2D images is then used to reconstruct an approximation of the soil sample by projecting only the pore spaces. This reconstruction shows the structure of the soil and its pores, which become more bounded, less bounded, or unbounded with changes in depth. If the soil sample image quality is sufficiently favourable in terms of contrast, noise and sharpness, the pore identification is less complicated, and the PFCM clustering algorithm can be used without additional processing; otherwise, images require pre-processing before using this algorithm. Promising results were obtained with four soil samples, the first of which was used to show the algorithm validity and the additional three were used to demonstrate the robustness of our proposal. The methodology we present here can better detect the solid soil and pore spaces on CT images, enabling the generation of better 2D–3D representations of pore structures from segmented 2D images.

1. Introduction

Soil systems play diverse and critical roles in the biosphere by regulating biogeochemical cycles, providing a habitat for microorganism, and a medium for plant roots, storing nutrients and minerals, and conducting fluids (such as water and gases). The latter is especially relevant for crop production and the environmental protection of gas emissions (Hamamoto et al., 2011). Thus, identifying the pore structure of a given soil segment is important, as it provides information on the soil mechanics and on how well the dirt can fulfil its function.

Solid-pore space is organised by its different scales: large bio-pores, inter-aggregate pores and small pores inside aggregates (Bronick and Lal, 2005). This complex arrangement is critical for soil function. In the past decade several studies have demonstrated that soil heterogeneity

and complexity at small scales determine its functionality and sustainability at larger scales (Crawford, 2010; De Bartolo et al., 2011; Jacobson et al., 2007). This observation makes it especially interesting to study, several soil characteristics, such as solid-pore structure, in greater detail. This information, obtained with the appropriate mathematical approaches and experimental systems, will be useful for tackling important environmental and ecological problems (Crawford, 2010; Perret et al., 2000).

With recent rapid advancements in digital cameras, computer processing information and data storage capacity and software, complete image analysis systems can be readily built for the quantitative analysis of soil morphology. A number of studies can thus be performed to characterise the pore geometry and topology of a soil sample, relating these factors with important characteristics, such as its effective water permeability (Unsal and Dane, 2006), hydraulic conductivity (Udawatta et al., 2008), and solute transport (Maximilian Köhne et al., 2011). The proposed approach therefore provides a new opportunity to develop techniques for quantifying the properties of soils based on the pores they contain.

Image analysis can be used to identify the components of the soil. Until now, several instruments have been used to obtain images of soil samples, including light microscopes, scanning electron microscopes (SEMs), and transmission electron microscopes (TEMs), as well as Computed Tomography (CT) and Nuclear Magnetic Resonance Imaging (MRI) systems (Chen et al., 2002).

Wang et al. (2011) have described the importance of an accurate grey scale to the quality of digital data characterisation of pore structures. The principal advantages of the CT technique are that it reduces the physical impact of the sampling, provides three-dimensional (3D) information, and allows for rapid scanning, such that the sample can be studied in nearly real-time (Rasiah and Aylmore, 1998). Several authors have recently dedicated their attention to solid-pore CT boundary identification, as this is a critical step in digital image processing on which the quality of the final results is highly dependent (Baveye et al., 2010; Tarquis et al., 2008, 2009, 2012). An overview of the steps and instrumental settings used to generate the X-ray CT greyscale soil images and the influence on the final quality can be seen in Houston et al. (2013a).

In general, image analysis involves many tasks, such as segmentation, classification, and interpretation. Segmentation involves identifying objects within images. The segmentation of soil images is very important for the subsequent measurement of the porosity, pore connectivity and pore size, as well as for detecting and recognising objects in the soil (Baveye et al., 2010; Hapca et al., 2013; Lehmann et al., 2006; Piñuela et al., 2010; Vogel et al., 2010). Classification assigns labels to individual pixels using prior information about the problem of interest. Interpretation involves extracting some meaning from the image as a whole.

Several methods have been used to segment soil images, including a simple binary threshold method (Perret et al., 2003), a multiple threshold method (Tarquis et al., 2009), and thresholds for typical and critical regions. Iassonov et al. (2009) have performed a wide review of different segmentation methods typically used in the field of geoscience. Other methods that appear to be promising for soil applications are the clustering and entropy-based methods (Cortina-Januchs et al., 2011; Iassonov and Tuller, 2010; Iassonov et al., 2009; Sezgin and Sankur, 2004), as well as the improvement that Houston made to the Indicator Kriging algorithm (Houston et al., 2013b) and the Schlüter method for determining a two-level intensity threshold (Schlüter et al., 2010). The fully automated segmentation method reported by Hapca et al. (2013) also shows promise.

A constant problem in the application of segmentation algorithms to images is how to evaluate the quality of the results when no ground-truth information about the soil sample is available. This problem requires the use of criteria that do not depend on ground-truth data to assess the accuracy of the segmentation method. In this work, the non-uniformity (NU) measure is used to determine segmentation accuracy, because it does not depend on the presence of ground-truth image (Zhang, 1996, 2001).

The proposal in this work is a first approach, which is based on the decomposition of 3D images into a series of consecutive 2D layers obtained by gamma-ray computerised tomography of soil samples that can be taken as slices. This set of images is treated with our segmentation method, slice by slice, such that the pore and solid spaces can be identified. This method allows for the generation of a good approximation of the 2D–3D visual representation of the pore structures from the segmented 2D images which are arranged consecutively, in turn enabling the identification of morphological properties of the sample, such as the pore distribution, connectivity, and dispersion/distribution between slices. This process also provides information about the characteristics of the soil, through the quantification of the total solid space in the sample.

Our proposal has two main goals: i) the first involves determining the robustness of the PFCM clustering algorithm to identify the region of interest (pore spaces), and ii) the second involves visualising pore space variation within a soil sample, sequentially aligning each of the 2D images to determine whether the pore spaces grow, shrink, connect

to each other or disappear. Consequently, this proposal is a first approach to generate a 2D–3D like image using the pore spaces identified from the 2D slices (consecutive 2D images). Later, we will apply a geometric interpolation algorithm that will allow us to connect the pore spaces of the 2D segmented images and finally generate a 3D image of the pore structures.

2. Materials and methods

2.1. Soil samples

One of the soil samples studied in this work was collected from the last horizon (the Bt/Bw horizon; see Table 1) of an argisol formed on the Tertiary Barreiras group of Pernambuco, Brazil, at the Itapirema Experimental Station. According to the Köppen classification, this area has a tropical monsoon climate. The physical and chemical characteristics of its soil have been broadly analysed (Melo and dos Santos, 1996).

Three soil samples were obtained at the Alameda del Obispo experimental farm (388 N, 58 W, altitude 110 m) in Cordoba, Spain. This soil is a loamy alluvial, Typic Xerofluvent (Soil Survey Staff, 2010), with a particle-size distribution in the upper (0–15 cm) soil layer as follows: sand, 350 g kg⁻¹; silt, 443 g kg⁻¹; and clay, 206 g kg⁻¹. Each sample corresponds to plots ploughed with different tillage tools: namely mouldboard, chisel and roller.

For this study, each soil sample was imaged using an EVS MSMicroCT scanner (now GE Medical, London, Canada). Though it was necessary to prune the sample in order to fit it into the 64 mm diameter of the imaging tubes, the field orientation was maintained. The imaging parameters for the sample were 155 keV, with 2205 projections and 25 µm. Proprietary software (GE Medical) was used to reconstruct the 3D image (8-bit) from the axial sequence views. One sub-volume measuring 256 × 256 × 256 units was extracted, corresponding to approximately 16.7 million voxels; the resulting voxel size was 45.1 µm. A high-pass copper filter was used, between the tube and the sample to minimise the hardening of artefacts. A 3D Gaussian filter was also run in Microview (GE Healthcare, 2006) to reduce noise and beam-hardening artefacts, which can occur in CT imaging (Tarquis et al., 2009). The last step was to generate a set of 200–2D images, each with a resolution of 256 × 256 pixels in 8-bit greyscale, from the 3D image of the soil sample. Fig. 1A shows three slices of the original images from the Brazilian soil sample, which are very dark. Under the dark condition it is difficult to identify the soil pores without image pre-processing.

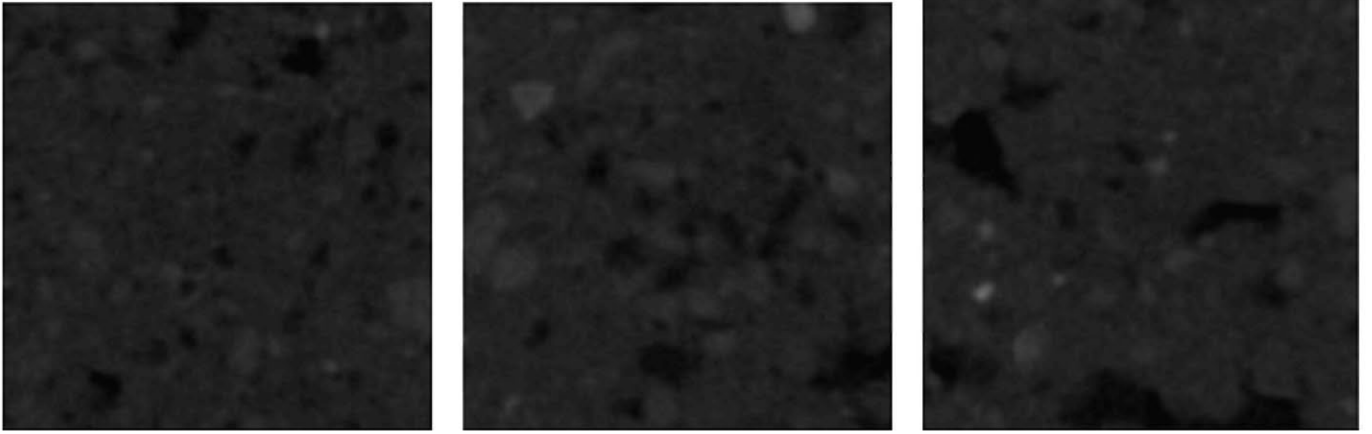
2.2. Image processing

The proposed methodology to obtain the solid-pore structure is based on three image processing steps, and a final alignment of the 2D–3D image with the solid regions and pores identified. As the 2D images are corrupted with noise, and there is usually limited contrast between the solids and the pores, the first step is to improve the images through a non-linear image enhancement technique. The second step

Table 1
Physical properties of argisol horizons, as per Melo and dos Santos (1996).

| Horizon | Depth (cm) | Consistency | | Bulk density (kg dm ⁻³) | Particle size distribution (%) | | | |
|---------|------------|---------------|--------------|-------------------------------------|--------------------------------|---------|------|------|
| | | Dry | Moist | | C. Sand | F. Sand | Silt | Clay |
| A2 | 10–35 | Slightly hard | very friable | 1.4 | 62 | 24 | 3 | 11 |
| AB | 35–57 | Slightly hard | friable | 1.5 | 26 | 53 | 4 | 17 |
| Bt2 | 98–152 | Slightly hard | friable | 1.4 | 21 | 40 | 4 | 35 |
| Bt/Bw | 150–190 | Slightly hard | friable | 1.3 | 18 | 37 | 10 | 35 |

A



B

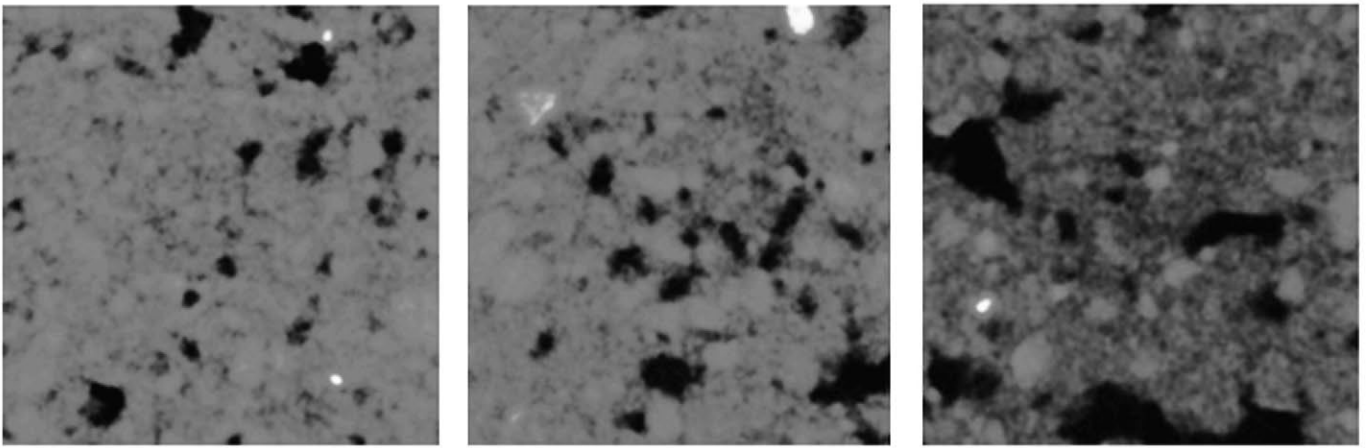


Fig. 1. Three examples of 2D images are shown in the first line (Img 001, Img 100, and Img 200) with 256×256 pixel resolution, and their respective transformed images with the $G(I)$ function are shown in the second line.

consists of dividing each image into solid regions and pores. A sub-segmentation method (Ojeda-Magaña et al., 2009a) based on a hybrid clustering algorithm was selected, such that the final images were divided into two groups: solid-pores and solid. To estimate the quality of the segmentation, a non-uniformity (NU) measure was applied, which evaluates the variance of the identified objects relative to the variance of the whole image and does not require ground-truth information. The fourth and final step consists of the consecutive alignment of the segmented 2D binary images to produce a visual representation of the pores in a 3D-like structure. These steps are described in more detail in the following sections.

The method used for the segmentation process of the soil images was coded in MATLAB v.7.7.0. The computer on which the algorithm was run had the following specifications: Intel Core i5-3 GHz processor, 12 GB RAM, and running the Windows 7 OS. The code runs the PFCM algorithm, performs the segmentation process with the matrix U and the matrix T (the latter of which carries out the sub-segmentation process), and calculates the NU value for each applied method. This code also shows all the windows and displays the resulting binary images with the pores for both matrices.

The required time for an algorithm with the above characteristics to produce its final results does not change linearly with the image size; for the package with 256×256 pixels, the resulting times were between 9 and 14 s. Other soil images with 512×512 pixels (not included in this study) resulted in processing times between 110 and 195 s.

2.3. Non-linear image enhancement

As the images contain noise (characterised by small differences in the grey values of the pixels) and the contrast between the solid regions and the pores is low, a preprocessing stage is recommended to both reduce the noise and simultaneously increase the values of the pixels that represent objects of interest. In this work, this process is implemented with a non-linear transformation, such that the grey values of the pixels are modified to be closer to the extreme values. Fig. 1A shows the unprocessed images, and Fig. 1B shows the processed images.

Applying these enhancement techniques to the images emphasises the important features while reducing the noise. Multiscale Adaptive Gain (Laine et al., 1994) suppresses pixel values with small amplitudes and enhances pixel values with large amplitudes bigger in amplitude of a threshold within each level of the transformed space. The selected function to provide this behaviour is

$$G(I) = \alpha[\text{sigm}(k(I-B)) - \text{sigm}(-k(I+B))] \quad (1)$$

and the α parameter is calculated according to the following equation.

$$\alpha = \frac{1}{\text{sigm}(k(1-B)) - \text{sigm}(-k(1+B))} \quad (2)$$

In Eq. (1), $I = I(i, j)$ represents the grey value of the pixel at coordinates (i, j) of the input image, and $\text{sigm}(x)$ is defined as

$$\text{sigm}(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

where B and k control the threshold and the rate of enhancement, respectively ($0 < B < 1$, $B \in \mathbb{R}$; $k \in \mathbb{N}$). The non-linear transformation function of Eq. (1) is represented in Fig. 2. As can be seen, $G(I)$ is a continuous and monotonically increasing function. Therefore, the enhancement does not introduce discontinuities into the transformed image.

2.4. Image segmentation using clustering techniques

Object identification in the soil images is very important, as it characterises the properties of the ground layer. The main problem with the CT soil images is the low contrast between the solid and pore spaces, with pore spaces represented by darker colours and the solid spaces by lighter colours. However, these objects can be correctly identified using a sub-segmentation method based on a hybrid clustering technique.

The main objective of using a clustering process to segment the images is to find pixels with similar grey-level intensities in order to integrate them into homogeneous groups. The similarity between pixels is evaluated according to a distance measure between each pixel and a prototype representing each object or region, such that each pixel is assigned to the group with the nearest or most similar prototype.

One of the most popular clustering algorithms is the Fuzzy c -Means (FCM) (Bezdek, 1981) due to its easy interpretation and implementation. However, this algorithm is very susceptible to outliers and, in some cases, can converge slowly. Other approaches have emerged to mitigate these problems. In particular, Krishnapuram and Keller have proposed an algorithm in which the probabilistic constraint is relaxed, called Possibilistic c -Means (PCM) (Krishnapuram and Keller, 1993). Nevertheless, this algorithm may have the problem of coincident clusters (Barni et al., 1996). The Possibilistic Fuzzy c -Means (PFCM)

clustering algorithm has therefore been proposed as a solution to the problems of both the FCM and the PCM algorithms (Pal et al., 2005). Using several experiments with a standard data set, Pal et al. have shown that the PFCM is a robust algorithm.

The PFCM algorithm has four adjustment parameters: a , b , η , and m . The a and b values represent the relative importance of membership and typicality in the computation of prototypes. If the value of a is greater than the value of b , the prototypes are more strongly influenced by the membership values; conversely, if b is greater than a , the typicality values have more influence, and the prototypes are expected to be less affected by noise. For the soil images used in this study, we proposed a value of a smaller than b to produce better pore identification. Parameter m has the same function as in the FCM algorithm, and parameter η is the same as in the PCM algorithm but for PFCM. The current literature suggests that these two values should be small and close to two. The same is true for the η parameter in the PCM algorithm. However, parameters a and b have a large influence on the effect of the outliers. In fact, Pal et al. (2005) recommend setting a greater than b to reduce the outlier effect.

The soil sample images consist of only two objects, the solids and the pore spaces. The segmentation of these images must therefore provide two regions, each representing one of the two objects. The sub-segmentation of the digital images, as proposed by Ojeda-Magaña et al. (2009a), divides the image into the minimum number of regions; in this case, into two regions containing the pore and the soil spaces (S_1 and S_2). However, the region S_1 represents both pore and solid spaces, while region S_2 represents only solid spaces. The sub-segmentation method can find atypical pixels inside each previously segmented region. Here, this allows for the identification of pore spaces in region S_1 as the atypical pixels, because this region contains both pore and solid spaces.

2.5. Image sub-segmentation with the PFCM algorithm

The sub-segmentation method is based on the concept of typicality (which allows the most typical or atypical data within a group to be found) and a hybrid partitional clustering algorithm, such as the PFCM

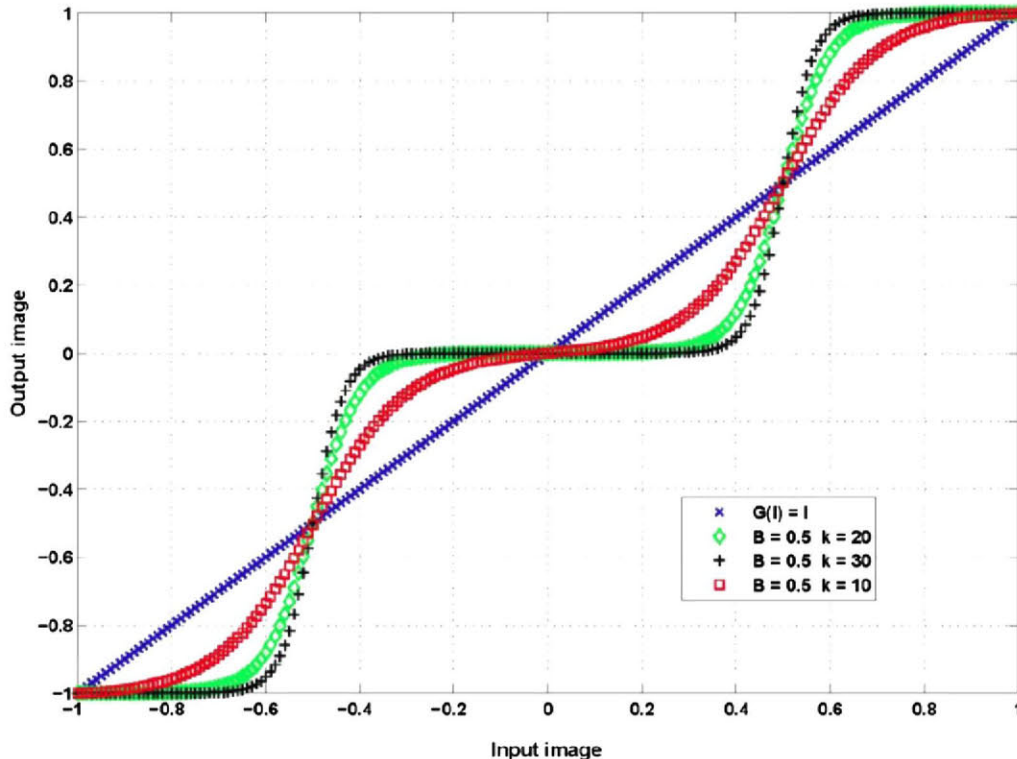


Fig. 2. The non-linear transformation function $G(I)$ used to improve the contrast of the original image. Parameters: $B = 0.5$, $k = 10, 20$, and 30 .

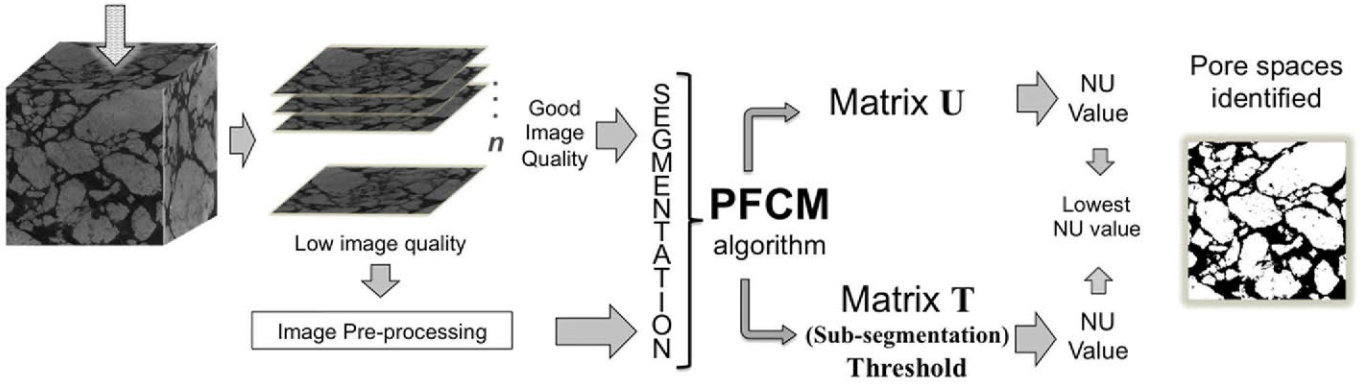


Fig. 3. Sub-segmentation method for pore and solid identification.

algorithm used here. An analogy (Ojeda-Magaña et al., 2009b) has been made between clustering algorithms, based on fuzzy and possibilistic concepts, and the *theory of prototypes* proposed by Rosch and Mervis (1975). In this case, the membership degrees $\{\mu_{ij}\}$ of the fuzzy sets can be interpreted as a relative typicality ("probabilistic constraint"). The typicality values $\{t_{ik}\}$ produced by the possibilistic part of the algorithm as an absolute typicality, can be used to differentiate those elements that are similar to a prototype, or the typical elements of a group, from those that belong to the same group but differ significantly from the prototype.

To put it differently, the relative typicality results from an external dissimilarity, as the membership degrees of an object are constrained to unity. In contrast the absolute typicality depends only on the elements of a particular group, because it is based on internal resemblance (Ojeda-Magaña et al., 2009c). This is an excellent alternative for finding more information about the particular application and is directly related to the pixels of an image. In this case, where typical or the most atypical data can be easily identified inside the regions provided by the clustering algorithm. See Fig. 3 for a visual representation of the proposed methodology.

Because typicality values fall within a set interval $[0, 1]$, a threshold α , within the interval $(0, 1)$ can be used to divide a region according to the value of typicality of each element. Specifically, the pixels of a region in an image can be considered either typical or atypical but not both. The pixels with values above α are defined as typical pixels, while the others are defined as atypical pixels. The threshold value must therefore be carefully selected, as it establishes the boundary between the typical and atypical pixels. Even if we try to find the atypical pixels in most applications, which requires an α value close to zero, it is also possible to find the most typical pixels, for which the α value would be close to one.

The sub-segmentation method is based on the PFCM and an iterative optimisation process that uses the prototypes, consisting of the following steps:

- I. Obtain the image and identifying features to be used in the algorithm. In this case we only use the grey value of the pixels as a feature.
- II. Assign a value to the parameters (a, b, m, η) of the PFCM algorithm, and establish the number of clusters c or number of regions $S_i, i = 1, \dots, c$ to partition the space of the image features and select a value for the threshold α .
- III. Run the PFCM algorithm (Pal et al., 2005) to obtain the following:
 - a. The membership matrix U .
 - b. The typicality matrix T .
- IV. Label each pixel $z_k, k = 1, \dots, N$, with the name of the Fuzzy Region (FR) SF_i according to its maximum membership value in $U = [\mu_{ik}], i = 1, \dots, c$. That is,

$$z_{FR, LABELi} = \max_i [\mu_{ik}], i = 1, \dots, c. \quad (4)$$

Therefore each pixel z_k can only belong to one region SF_i .

- V. Label each pixel $z_k, k = 1, \dots, N$, with the name of the Possibilistic Region (PR) SP_i according to the maximum typicality value in $T = [t_{ik}], i = 1, \dots, c$, as follows:

$$z_{PR, LABELi} = \max_i [t_{ik}], i = 1, \dots, c. \quad (5)$$

Therefore each pixel z_k can only belong to one region SP_i .

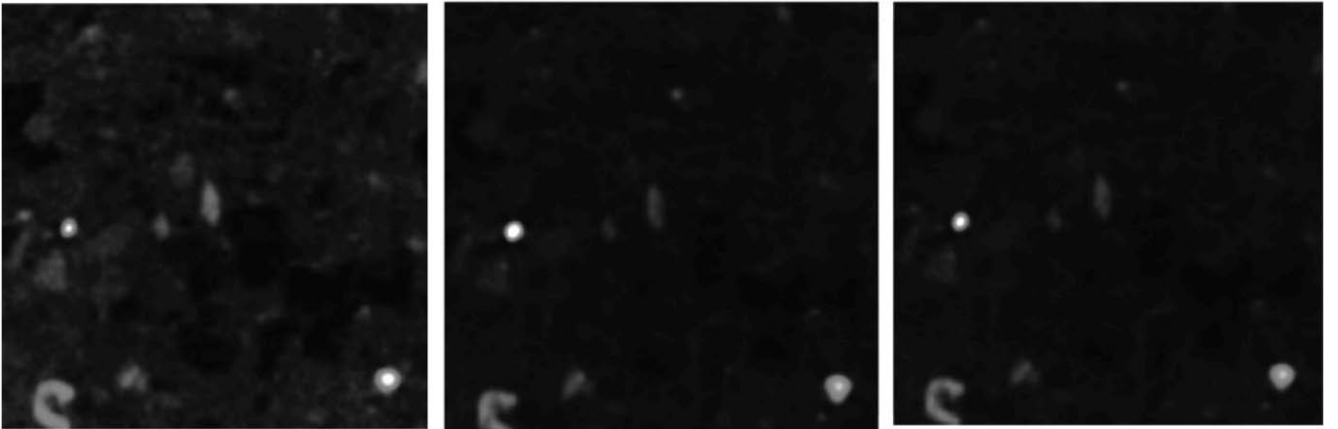
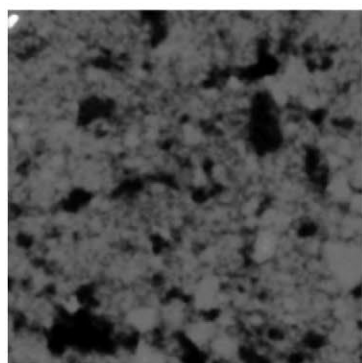
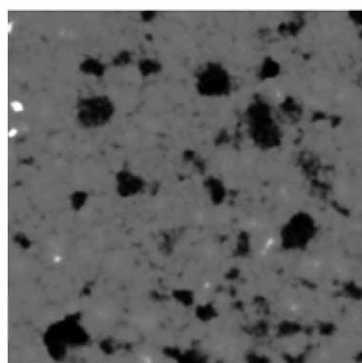
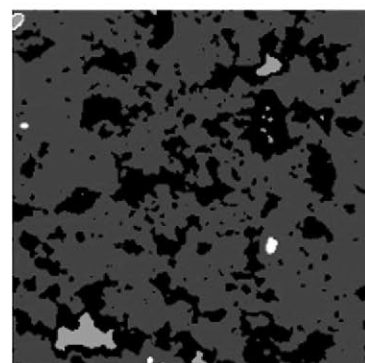


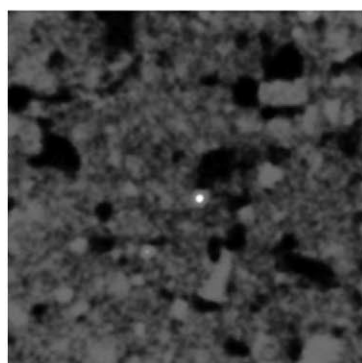
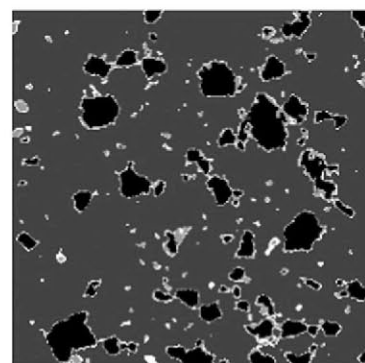
Fig. 4. Examples of CT soil images with poor contrast (Img 087-left, Img088middle, and Img 089-right).



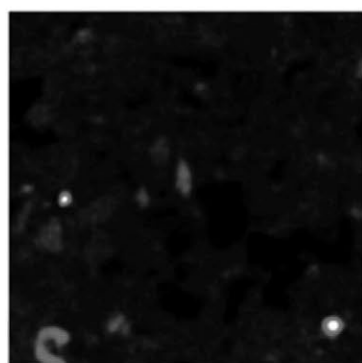
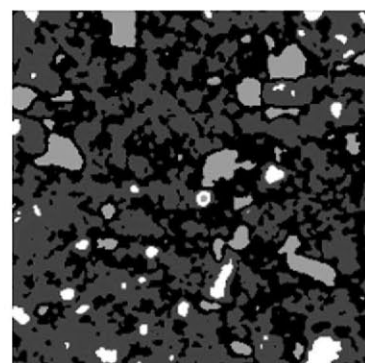
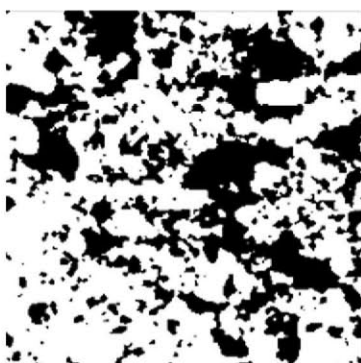
Img 007



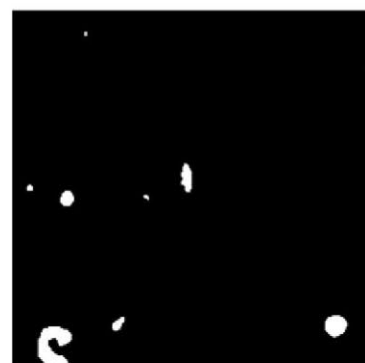
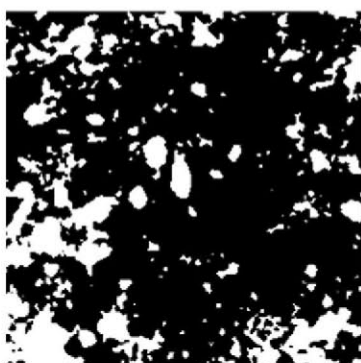
Img 012



Img 036



Img 087



A)

B)

C)

Fig. 5. Processed images (Img 007, Img 012, Img 036, and Img 087): a) transformed images, b) segmented images through the U matrix, and c) pores and solid spaces at the end of the sub-segmentation process.

VI. Group the pixels in the T_{\max} matrix according to their labels.

$$T_{\max} = z_{pk, LABEL_i}, i = 1, \dots, c. \quad (6)$$

VII. Separate T_{\max} with α in two sub-matrices (T_1, T_2), such that

$$T_1 = \{t_{ij}/t_{ij} \in T_{\max}, t_{ij} \geq \alpha\} \quad (7)$$

and

$$T_2 = \{t_{ij}/t_{ij} \in T_{\max}, t_{ij} < \alpha\}. \quad (8)$$

In this case, T_1 contains the typical pixels, T_2 contains the atypical pixels, and their elements define the two PR sub-regions. We then select the sub-matrix T_1 or T_2 of interest for the corresponding analysis; for the soil images, the region of interest for finding the pore spaces is T_2 which contains the atypical pixels.

VIII. If necessary, the typical or atypical z_k pixels can be found by their regions. In such a case,

$$T_{1, LABEL_i} = T_1, LABEL_i, \dots, LABEL_c \quad (9)$$

or

$$T_{2, LABEL_i} = T_2, LABEL_i, \dots, LABEL_c. \quad (10)$$

Fig. 3 shows the sub-segmentation process, for which the input is a 2D soil image (the grey level is the only feature). In this process, we run the PFCM clustering algorithm to divide the pixels of the feature space into two regions or clusters, as matrix U (FR) and matrix T (PR). We then reconstruct the segmented image using the values of matrix U. Finally, we divide the regions of matrix T into two regions under a threshold value to obtain the typical and atypical pixels within each region.

As with other segmentation methods, this method requires the establishment of a threshold value that defines the boundary between regions (in this application we want to distinguish between the solid and the pore space region). In the particular case of the partitioning clustering algorithms, this boundary is automatically determined and depends on the number of clusters c and the type of membership function used. However, c is unknown and can be estimated through an iterative process that, in some cases, can be performed automatically. However c is determined, it is always a discrete value, and the regions produced by

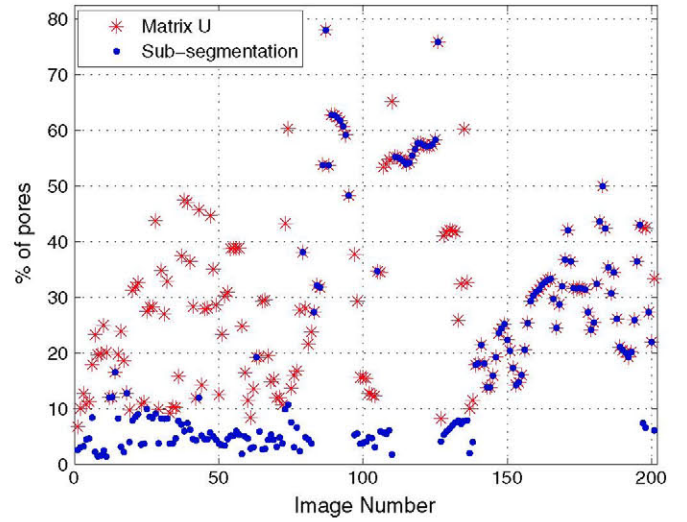


Fig. 7. Porosity in each of the 200-2D images.

a partitioning clustering algorithm are a discrete approximation of the space characteristics. This, along with the low computational cost, is a great advantage of the sub-segmentation method, as the regions can be continuously approximated by varying α in the interval $(0, 1)$, and the regions of interest can be better identified.

2.6. Non-uniformity measures

The measure of non-uniformity has been developed to quantitatively assess the segmentation quality of a region, even when the ground-truth information is not available (Zhang, 1996, 2001). This measure is defined as follows:

$$NU = \frac{P}{T} \cdot \frac{\sigma_P^2}{\sigma^2} \quad (11)$$

where P and T represent the number of pixels of the pores and the pixels of the whole image, respectively, in the segmented image. Here σ_P^2 and σ^2 represent the variances of the grey-scale values in the pore space, and the total variance in the simulated greyscale image respectively. We have used NU to study the quality of the method presented here.

2.7. Detection of pore spaces in CT soil images

The detection of pores largely depends on the grey scale of each image, as small variations in these values make it a very difficult task to identify the pixels that correspond to the pores. Their identification is simpler when there are large changes in the grey values. As there are only two main objects in a soil image, it must be segmented into two regions, representing the solid spaces and the pore spaces. The pixels of the solid spaces are those with a lighter colour, while the pixels of the pores are those with a darker colour.

The problem now is to find the boundary that divides the pixel sets of the image. For this purpose, we apply the PFCM clustering algorithm because it provides the membership degrees and typicality values that can better differentiate between the two sets. Of note, 200 images of a sample of soil are used, each corresponding to a different depth stratum in which the pores should have a rough correspondence; otherwise, large variations in consecutive images may be indicative of poor image quality. We also used three different soil samples (chisel, roller and mouldboard, see Fig. 8 upper, middle and lower, respectively) to test the robustness of the proposed algorithm.

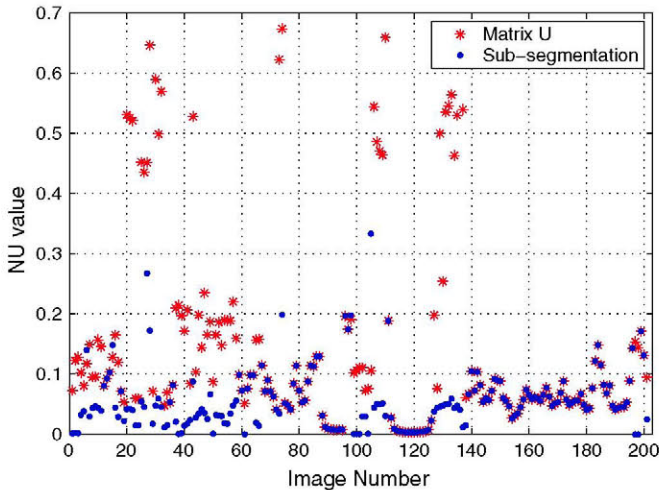


Fig. 6. NU value calculated for each of the 200-2D images of a soil sample.

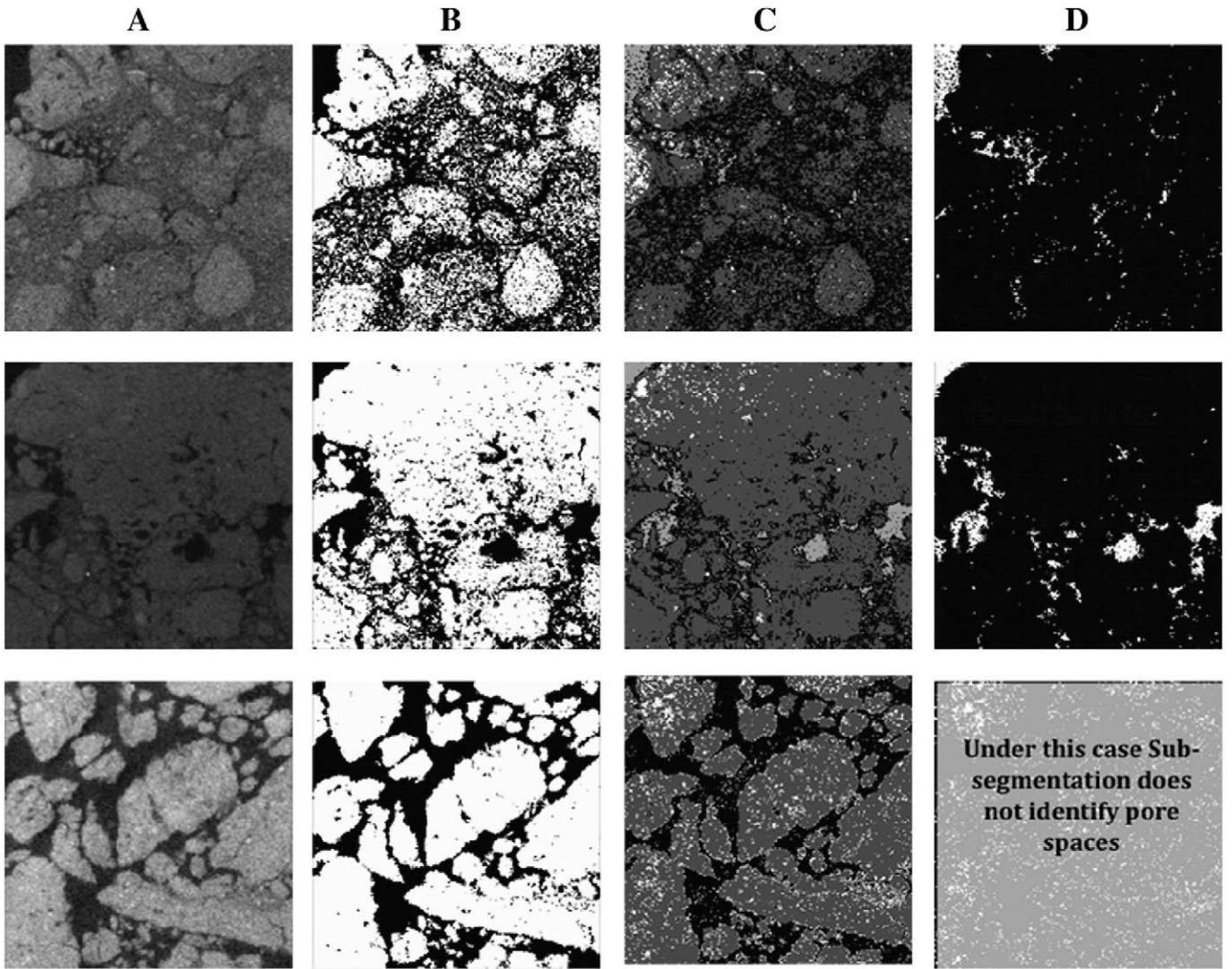


Fig. 8. Chisel soil sample (top image), roller soil sample (middle image), mouldboard soil sample (bottom image), A) original B) PFCM (matrix U) C) sub-segmentation D) identified pores.

The following is a detailed description of the proposed methodology to automatically detect the pore spaces:

- Select a soil sample and its corresponding images (200 CT images in 2D, which have been extracted from the 3D object).
- Improve the contrast between the pore and solid spaces in the CT images using a multiscale contrast enhancement.
- Identify, through sub-segmentation based on the PFCM, the pore and the solid spaces. The recommended parameters for the PFCM algorithm are as follows:

$$a = 1, b = 4, m = 2, \eta = 2.$$

- With the PFCM algorithm generates matrices U and T.
 - Once the T matrix is available, sub-segment the image into two PR: S_1 (containing pore and solid spaces) and S_2 (containing only solid spaces).
 - Next, find the atypical pixels of each region using a threshold value of $\alpha = 0.08$. As S_1 contains the pore spaces that are represented by the lowest value of typicality, the atypical pixels correspond to these elements.
- d) Following the identification of pore spaces, apply analysis to estimate the quality of the results applying the NU measure. The better segmented images are those whose corresponding NU value is nearer to zero.

- Calculate the percentage of pore spaces in each 2D image.
- Consecutively align the 200-2D segmented images to visually identify the structure of the pore spaces of the soil sample as a 2D-3D like image.

3. Results and discussion

3.1. Image processing

As the original CT soil images had a poor contrast, and were very dark, a preprocessing stage was necessary. This stage was performed through a continuous non-linear function and, particularly, a multiscale adaptive gain to enhance the image contrast. Encouraging results were obtained in most images, although it was very difficult to enhance the contrast in some images, as seen in the cases shown in Fig. 4. It is important to note that the three samples used to prove the robustness of our proposal are of good quality, and there is consequently no need for image pre-processing (Fig. 8A).

3.2. Image sub-segmentation

In traditional approaches to image segmentation with clustering algorithms, such as the *k*-Means or the FCM methods, the number of regions or clusters to identify is unknown in advance. Therefore, this

Table 2
Experiment with the three soil samples.

| Soil sample | Pre-processing of the soil image | Pores identification PFCM | | Threshold value |
|-------------|----------------------------------|---------------------------|------------------|-----------------|
| | | Matrix U | Sub-segmentation | |
| Chisel | Without preprocessing | Yes | Yes | 0.05 |
| Roller | Without preprocessing | Yes | Yes | 0.05 |
| Mouldboard | Without preprocessing | Yes | No | 0.08 |

number must be established at its lowest value. Then, the corresponding regions should be identified and increased iteratively until a region provides a good approximation of the pore spaces (Cortina-Januchs et al., 2011).

On the other hand, the sub-segmentation method (Ojeda-Magaña et al., 2009a), based on the PFCM clustering algorithm, is advantageous because it only needs to identify two regions or clusters (S₁ and S₂), a region containing the whole image (the pore, and the solid spaces), and another region containing only the solid space. Then, a fine tuning of the threshold, can divide both regions into typical and atypical pixels, without the need to execute the algorithm again. The atypical pixels correspond to pore spaces in a region that contains both pore and solid elements. The best results were obtained with a threshold value $\alpha \approx 0.08$ for the set of 200-2D images. Thus, the sub-segmentation approach allows for a very fine adjustment of the regions of interest, and has the additional advantage of a low computational cost.

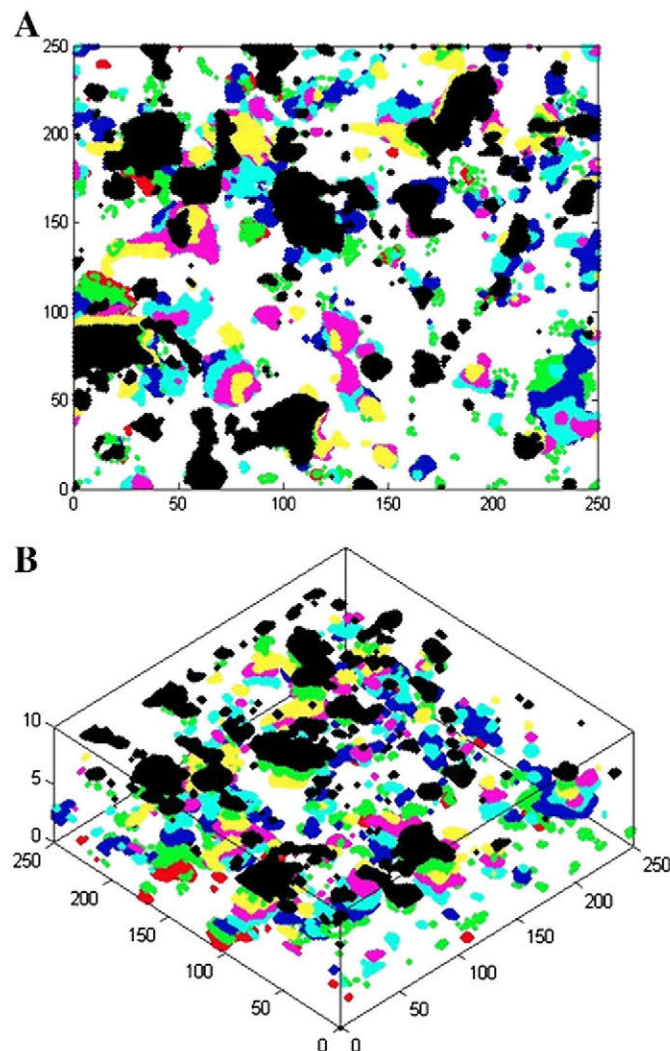


Fig. 9. View of the pores for the first 2D stacked images: A) aerial view, B) 3D view.

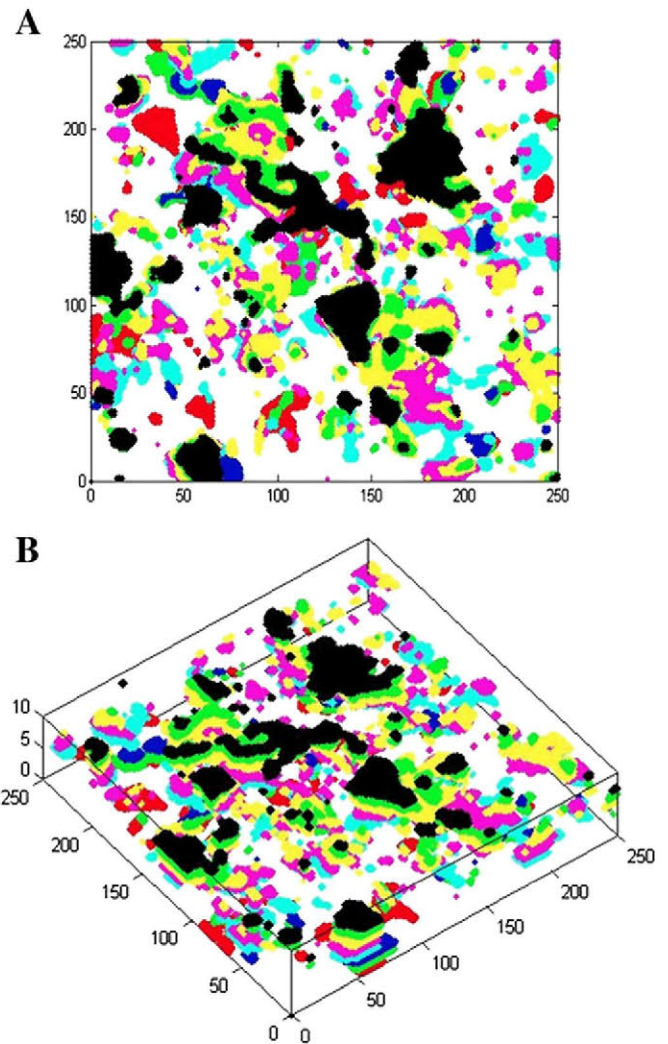


Fig. 10. View of the pores for the medium 2D stacked images: A) aerial view, B) 3D view.

Fig. 5 shows four interesting examples from the sub-segmentation of the 200 images of a soil sample. The intermediate results are organised as follows: the first column contains the images with the enhanced contrast, the second column shows the fuzzy regions (FR), the third column illustrates the two elements of a soil image (the pores and the solids), using different colours and identified through a fine tuning of a threshold in the possibilistic regions (PR). In the last column, the black and dark grey colours represent the pores and solids, respectively, in the PR, whereas the light grey and white colours represent the same elements but in the sub-segmented PR.

Close analysis of the results of Fig. 5 demonstrates that it is very easy to identify the pores and the solids when the images have good contrast between both elements. Such is the case in Img 007, where these elements are clearly identified from the FR, and the sub-segmentation of the PR only confirms the homogeneity of the regions. Img 012 has a higher contrast than Img 007 and has more homogeneous regions. In this case, the FR and the PR are the same. However, the sub-region of pores in the PR only corresponds to the pore contours, which could be considered the ideal case, as the pores are very well identified.

In Img 036, the transition between the pores and the solid regions is smoother, so it is more difficult to establish the boundaries between them. As can be seen in Fig. 5, the FR are not correctly identified because of the solid space that appears to be pore space. This error could be a consequence of the heterogeneity of the pore spaces. In this example the sub-segmentation of the PR provides much better results, as the pores and solids are more closely matched.

In Img. 087, the last example of Fig. 5, there is poor contrast between the grey values of the pores and the solids that limits the quality of the results and, in this case, both the FR and the PR give poorer results than in the previous examples. However, the NU value can be used to determine the best results in each of the previous cases.

To verify the robustness of the PFCM algorithm used for the segmentation process on the soil images we looked at three additional soil samples, chisel, roller and mouldboard (Fig. 8A). Each soil sample has 200 images. These images have a resolution of 256×256 and different threshold values (Table 2).

From the new experiments, we found that, if images are of good quality as in the three additional soil samples (chisel, roller, mouldboard), there is no need for image pre-processing and the PFCM segmentation can be applied directly to the unprocessed images. We also find that if there is a high contrast between the pore space and the soil space, we can identify the pore spaces in matrix U without the need for sub-segmentation (see Fig. 8B for the mouldboard sample, and Fig. 8C for the chisel and roller with sub-segmentation). Finally, we have found that the optimal threshold value for the sub-segmentation process can vary according the image quality. During the experimentation, we have used threshold values from 0.05 to 0.08, which allow for good pore identification, (Fig. 8D).

3.3. Comparative analysis of segmentation results

In the previous section, we analysed four particular cases among a set of 200 images of a soil sample. The selection was made according to a visual inspection and to show some characteristics of the images. In this section, however, we apply the NU value as a way to estimate the quality of the segmentation results of the pores for the U matrix and the sub-segmentation. As previously stated, the lower the value the better the results, keeping in mind that the segmentation of poor quality images usually produces NU values that are near zero. Thus, if the image is of poor quality even after enhancement, the pores cannot be identified, and because of this the proposed method cannot be automated. However, a visual inspection is still necessary since, in some images, the NU value is very low; nevertheless even though when visually inspecting the images we find that the pores are not well identified.

The NU value was calculated for the 200 images, and the corresponding values are plotted (Fig. 6). We can compare the NU values from the U matrix to those resulting from the sub-segmentation. As shown, the values are very similar for Img 007 (Fig. 5a), although the sub-segmentation provides a lower value due to the greater homogeneity of the pores.

The NU value is very close to zero for the second image, Img 012 (Fig. 5b), and it is almost the same for both methods. This scenario is indicative of high quality segmentation results, and, as seen in the corresponding figure, the contrast between the pores and the solid spaces can be very well differentiated. An important fact here is that approximately 50% of the images present similar characteristics, and the corresponding NU value is lower than 0.1 for these images.

Unlike the similarity of the NU values in Img 007 and Img 012, these values exhibit significant differences for Img 036, with a lower value for the sub-segmented regions; approximately 40% of the 200 images have similar characteristics. In Img 087, the fourth and last case, the values of NU are very close to zero. This result is observed because the darkness of the image, and the low contrast variation result in homogeneous regions and, consequently, low NU values. Images such as this have poor quality, and approximately 9% of the 200 images produce similar results. These images can be easily identified in Fig. 6, as they are consecutive images with a NU value close to zero.

For the segmentation of the three extra soil samples, the conclusions are as follows: the threshold value is used with the sub-segmentation, but if the contrast between the solids and pores is high there is no need to apply the sub-segmentation. In this case, it is sufficient the use of matrix U to identify the pores, as demonstrated for the mouldboard sample. As stated above, it is not always necessary to apply a pre-processing to the images; if images are of good quality, the PFCM can be applied directly, without sub-segmentation.

3.4. Porosity

Once the pore spaces were identified, we proceeded to quantify the percentage of porosity in each image (Fig. 7). As shown, there is a large difference between the results obtained from the U matrix and from the sub-segmentation. For the first images, the U matrix suggested a great

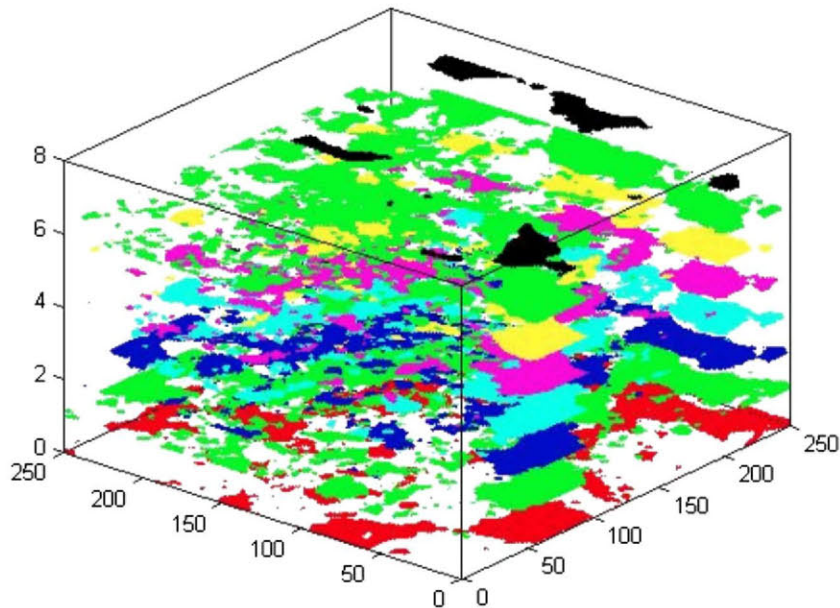


Fig. 11. Aerial view of the pores for the last 2D stacked images.

dispersion and a rather high percentage of porosity, whereas the sub-segmentation suggested an average porosity percentage below 10%.

The percentage of pores in the images of poor quality is very high and constant but this value is approximately 25% for the last images.

Qualitatively, as the soil depth increases, the percentage of porosity also increases. Poor quality images present atypical values, as shown. An alternative method to estimate more real values is to consider the smooth variations in the size of pores such that polynomial interpolation can be used for this approximation.

3.5. Reconstruction of the soil sample including the pores identified

Once the pores have been identified in a set of images, they are put into place for the discrete reconstruction (200-2D images) of the 3D earth stratum. Some analysis can therefore be carried out to evaluate the pore variation according depth in terms of pore growth, connection, vanishing, etc. To show these characteristics, eight binary images were randomly selected from the top, middle and lower levels. Fig. 9A and B shows the 2D and approximated 3D views of the upper stratum of earth. In the overhead view of Fig. 9A, each colour represents a slice of the 3D sample, and the identified pores can be appreciated. These pores vary in size from slice to slice, and in few cases, they vanish. Fig. 9B shows a clearer view, as it contains a discrete 3D reconstruction that makes it easier to appreciate the size of the pores and their relations.

The selected images from the middle section of the earth stratum are plotted in Fig. 10A and B. Compared to the previous figures, there is a greater number of pores with more important size, and the larger pores are found in relatively the same positions in the middle images. These pores can be observed until the lower layer of the soil, in which the percentage of pores decrease significantly. The corresponding images for this last layer are shown in Fig. 11.

4. Conclusion

This work proposes an alternative method to detect pore spaces in CT soil images through image processing, data clustering, and sub-segmentation. The only feature used here was the grey level, and a pre-processing step was sometimes necessary to improve the contrast, when the original images had only small differences between the grey levels of the pore spaces and solid spaces.

Grey levels are represented with 8 bits, such that there are 256 different levels, from 0 to 255. The values closest to zero are related to the pores, and the goal is identifying the threshold such that the pores are well identified. In this work, we have taken the pixels with lowest values, which are still related to the higher value and brighter pixels. This step can be performed with the sub-segmentation method, as it allows for the fine-tuning of the threshold, thus enhancing the separation between both classes. In this case, the value of the threshold was the same for the 200-2D images of the same soil sample.

The calculation of the NU measure, which represents an estimation of the quality of the sub-segmentation results, provides an indication of the variance of the pores with respect to the variance of the whole image. As the variance of the pores diminishes, the value of the NU measure approaches zero, which is indicative of a good identification. Nevertheless, the NU measure must be handled with care, as the values calculated with images of poor quality are also near zero, even though the images contain a high porosity percentage. This problem appears when there are comparatively low percentages of pores in the upper and lower 2D images with respect to the intermediate images.

The results show that the proposed method is robust and can generate promising results. It can be therefore be used as an alternative method to determine the characteristics and properties of a soil, such that better-informed decisions can be made in terms of whether a particular soil could fulfil a specific function. In some cases the sub-segmentation is not needed (as for the mouldboard soil sample). Setting the threshold value for the sub-segmentation process remains a challenge, as this

number is subjective and depends largely on the image quality. We have also observed that the NU method has some deficiencies, especially with poor quality images. In forthcoming papers we will perform a deeper analysis of the threshold for the sub-segmentation, and we will try to optimise this parameter by taking into account the NU value or other more robust methods.

Acknowledgements

The authors wish to thank to The National Council for Science and Technology (CONACyT) of Mexico. Funding provided by the Spanish Ministry of Science and Innovation (MICINN) through project no. AGL2010-21501/AGR is greatly appreciated. Also we would like to thank reviewers and Dr. Benhard Angele for their valuable comments to improve this article.

References

- Barni, M., Cappellini, V., Mecocci, A., 1996. A possibilistic approach to clustering. *IEEE Trans. Fuzzy Syst.* 4 (3), 393–396.
- Baveye, P.C., Laba, M., Otten, W., Grinev, D., Bouckaert, L., DelloStarpao, P., Goswami, R.R., Hu, Y., Liu, J., Mooney, S., Pajor, R., Sleutel, S., Tarquis, A., Wang, W., Wei, Q., Sezgin, M., 2010. Observer-dependent variability of the thresholding step in the quantitative analysis of soil images and X-ray microtomography data. *Geoderma* 157, 51–63.
- Bezdek, J.C., 1981. In: Press, P. (Ed.), *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York.
- Bronick, C.J., Lal, R., 2005. Soil structure and management: a review. *J. Geoderma* 124, 3–22.
- Chen, Q., Wolfgang, K., Sascha, O., 2002. Nuclear magnetic resonance imaging for studies of flow and transport in porous media. *J. Environ. Qual.* 31 (2), 477–486.
- 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, 279–288.
- Crawford, J.W., 2010. Can complex be simple? *Geoderma* 160 (1), 1–2.
- De Bartolo, S., Otten, W., Cheng, Q., Tarquis, A.M., 2011. Modeling soil system: complexity under your feet. *Biogeosciences* 8, 3139–3142.
- Hamamoto, S., Moldrup, P., Kawamoto, K., Wickramarachchi, P.N., Nagamori, M., Komatsu, T., 2011. Extreme compaction effects on gas transport parameters and estimated climate gas exchange for a landfill final cover soil. *J. Geotech. Geoenviron. Eng.* 137, 653–663.
- Hapca, S.M., Houston, A.N., Otten, W., Baveye, P.C., 2013. New local thresholding method for soil images by minimizing grayscale intra-class variance. *Vadose Zone J.* 12 (3).
- Houston, A.N., Schmidt, S., Tarquis, A.M., Otten, W., Baveye, P.C., Hapca, S.M., 2013a. Effect of scanning and image reconstruction settings in X-ray computed tomography on soil image quality and segmentation performance. *Geoderma* 207–208, 154–165.
- Houston, A.N., Otten, W., Baveye, P.C., Hapca, S., 2013b. Adaptive-window indicator kriging: a thresholding method for computed tomography images of porous media. *Comput. Geosci.* 54, 239–248.
- Iassonov, P., Tuller, M., 2010. Application of image segmentation for correction of intensity bias in X-ray CT images. *Vadose Zone J.* 9, 1–5.
- 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 Resour. Res.* 45, W09415.
- Jacobson, A.R., Dousset, S., Andreux, F., Baveye, P.C., 2007. Electron microprobe and synchrotron X-ray fluorescence mapping of the heterogeneous distribution of copper in high-copper vineyard soils. *Environ. Sci. Technol.* 41 (18), 6343–6349.
- Krishnapuram, R., Keller, J.M., 1993. A possibilistic approach to clustering. *IEEE Trans. Fuzzy Syst.* 1 (2), 98–110.
- Laine, A., Schuler, S., Fan, J., Huda, W., 1994. Mammographic feature enhancement by multiscale analysis. *IEEE Trans. Med. Imaging* 13, 725–740.
- Lehmann, P., Wyss, P., Flisch, A., Lehmann, E., Vontobel, P., Krafczyk, M., Kaestner, A., Beckmann, F., Gygi, A., Fluhler, H., 2006. Tomographical imaging and mathematical description of porous media used for the prediction of fluid distribution. *Vadose Zone J.* 5, 80–97.
- Maximilian Köhne, J., Schlüter, S., Vogel, H.J., 2011. Predicting solute transport in structured soil using pore network models. *Vadose Zone J.* 10 (3), 1082–1096.
- Melo, F.J.R., dos Santos, M.C., 1996. Micromorfologia e mineralogia de dois solos de Tabuleirocoiteiro de Pernambuco. *R. Bras. Ci. Solo* 20, 99–108.
- Ojeda-Magaña, B., Ruelas, R., Corona-Nakamura, M.A., Andina, D., 2009a. Better interpretation of numerical data sets by relative and absolute typicality of fuzzy clustering algorithms. *Spec. Issue Adv. Pattern Recognit.* 44, 157–166.
- Ojeda-Magaña, B., Ruelas, R., Buendía-Buendía, F., Andina, D., 2009b. A greater knowledge extraction coded as fuzzy rules and based on the fuzzy and typicality degrees of the GKPCM clustering algorithm. *Intell. Autom. Soft Comput.* 15 (4), 555–5571.
- Ojeda-Magaña, B., Quintanilla-Dominguez, J., Ruelas, R., Andina, D., 2009c. Images sub-segmentation with the PFCM clustering algorithm. 7th IEEE International Conference on Industrial Informatics. INDIN 2009, pp. 499–503.
- Pal, N., Pal, S., Keller, J., Bezdek, J., 2005. A possibilistic fuzzy c-means clustering algorithm. *IEEE Trans. Fuzzy Syst.* 13 (4), 517–530.
- Perret, J., Prasher, S.O., Kantzas, A., Langford, C., 2000. A two-domain approach using CAT scanning to model solute transport in soil. *J. Environ. Qual.* 29 (3), 995–1010.

- Perret, J.S., Prasher, S.O., Kacimov, A.R., 2003. Mass fractal dimension of soils macropores using computed tomography: from the box counting to the cube-counting algorithm. *J. Hydrol.* 267, 285–297.
- Piñuela, J., Alvarez, A., Andina, D., Heck, R.J., Tarquis, A.M., 2010. Quantifying a soil pore distribution from 3D images: multifractal spectrum through wavelet approach. *Geoderma* 155, 203–210.
- Rasiah, V., Aylmore, L.A.G., 1998. Characterizing the changes in soil porosity by computed tomography and fractal dimension. *Soil Sci.* 163, 203–211.
- Rosch, E., Mervis, C., 1975. Family resemblance: studies in the internal structure of categories. *Cogn. Psychol.* 7 (4), 573–605.
- Schlüter, S., Weller, U., Vogel, H.J., 2010. Segmentation of X-ray microtomography images of soil using gradient masks. *Comput. Geosci.* 36, 1246–1251.
- Sezgin, M., Sankur, B., 2004. Survey over image thresholding techniques and quantitative performance evaluation. *J. Electron. Imaging* 13 (1), 146–165.
- Soil Survey Staff, 2010. *Keys to Soil Taxonomy*, 11th ed. United State Department of Agriculture-Natural Resources Conservation Service, Washington, DC, USA.
- Tarquis, A.M., Heck, R.J., Grau, J.B., Fabregat, J., Sanchez, M.E., Antón, J.M., 2008. Influence of thresholding in mass and entropy dimension of 3-D soil images. *Nonlin. Processes Geophys.* 15, 881–891.
- Tarquis, A.M., Heck, R.J., Andina, A., Antón, J.M., 2009. Pore network complexity and thresholding of 3-D soil images. *Ecol. Complex.* 6, 230–239.
- Tarquis, A.M., Sanchez, M.E., Antón, J.M., Jimenez, Juan, Saa-Requejo, A., Andina, D., Crawford, J.W., 2012. Variation in spectral and mass dimension on three-dimensional soil image processing. *Soil Sci.* 177 (2), 88–97.
- Udawatta, Ranjith P., Anderson, Stephen H., Gantzer, Clark J., Garrett, Harold E., 2008. Influence of prairie restoration on CT-measured soil pore characteristics. *J. Environ. Qual.* 37 (1), 219–228.
- Unsal, E., Dane, J.H., 2006. Equivalent soil pore geometry to determine effective water permeability. *Vadose Zone J.* 5 (4), 1278–1280.
- Vogel, H.-J., Weller, U., Schlüter, S., 2010. Quantification of soil structure based on Minkowski functions. *Comput. Geosci.* 36, 1236–1245.
- Wang, W., Kravchenko, A.N., Smucker, A.J.M., Rivers, M.L., 2011. Comparison of image segmentation methods in simulated 2D and 3D microtomographic images of soil aggregates. *Geoderma* 162, 231–241.
- Zhang, Y.J., 1996. A survey on evaluation methods for image segmentation. *Pattern Recogn.* 29, 1335–1346.
- Zhang, Y.J., 2001. A review of recent evaluation methods for image segmentation. *International Symposium on Signal Processing and Its Applications*, Kuala Lumpur, Malaysia, 13–16 August, 2001, pp. 148–151.