

1                   **Observer-dependent variability of the**  
2 **thresholding step in the quantitative analysis**  
3 **of soil images and X-ray microtomography**  
4 **data:**

5  
6                   **Supplemental material**

7  
8                   by

9  
10           Philippe C. Baveye\*<sup>1</sup>, Magdeline Laba<sup>2</sup>, Wilfred Otten<sup>1</sup>, Liesbeth Bouckaert<sup>3</sup><sup>✉</sup>, Patricia  
11 Dello Sterpaio<sup>1</sup>, Rohit R. Goswami<sup>4</sup>, Dmitri Grinev<sup>1</sup>, Alasdair Houston<sup>1</sup>, Yaoping Hu<sup>5</sup>,  
12 Jianli Liu<sup>6</sup>, Sacha Mooney<sup>7</sup>, Radoslaw Pajor<sup>1</sup>, Steven Sleutel<sup>3</sup>, Ana Tarquis<sup>8</sup>, Wei Wang<sup>9</sup>,  
13 Qiao Wei<sup>5</sup>, Mehmet Sezgin<sup>10</sup>

14  
15  
16           <sup>1</sup> SIMBIOS Centre, Abertay University, Bell Street, Dundee, Scotland DD1 1HG, UK.

17           <sup>2</sup> School of Civil and Environmental Engineering, Hollister Hall, Cornell University,  
18           Ithaca, New York 14853, U.S.A.

19           <sup>3</sup> Department of Soil Management and Soil Care, Ghent University, 653 Coupure Links,  
20           9000 Gent, Belgium.

21           <sup>4</sup> Geosyntec consultants, 5901 Broken Sound Parkway N.W., Suite 300, Boca Raton,  
22           Florida 33487, U.S.A.

23           <sup>5</sup> Department of Electrical and Computer Engineering, Schulich School of Engineering,  
24           University of Calgary, 2500 University Drive, NW, Calgary, Alberta T2N 1N4, Canada

25           <sup>6</sup> Department of Soil Physics and Salinity, Institute of Soil Science, Chinese Academy of  
26           Sciences, 71 East Beijing Road, Nanjing 210008, China.

27           <sup>7</sup> Environmental Sciences, School of Biosciences, Biology Building, University of  
28           Nottingham, University Park, Nottingham NG7 2RD, U.K.

29           <sup>8</sup> Dept. of Applied Mathematics to Agriculture Engineering, Polytechnic University of  
30           Madrid (U.P.M.), Ciudad Universitaria s.n., 28040 Madrid, Spain.

31           <sup>9</sup> Department of Crop & Soil Sciences, Michigan State University, East Lansing,  
32           Michigan 48824, U.S.A.

33           <sup>10</sup> Tübitak Marmara Research Center, Information Technologies Research Institute,  
34           Gebze, Kocaeli, Turkey.

35  
36  
37  
\_\_\_\_\_  
\* Corresponding author: P.Baveye@Abertay.ac.uk

✉ The order of authors 4-16 is alphabetical.

1 ***Thresholding by experts***

2 *Expert 1*

3 This individual declined to work on the Pathway image, claiming a lack of experience with color  
4 image processing. The other two images were manipulated using the software ImageJ 1.41 (NIH,  
5 Bethesda, Maryland). Expert 1 considered that the greyscale histogram of image Thinsect exhibits a  
6 peak at large greyscale values, which was considered caused by the pores. The left arm of this peak is  
7 not straight but has a kink (inflection) at a greyscale value of 224, which was therefore chosen to  
8 threshold the image. In the case of the Nanotom image, contrast was first enhanced to get a larger  
9 difference of grey values between the resin and the solids. For this contrast enhancement, the  
10 minimum was set visually at 160, and the maximum at 245. Then the image was filtered, using a  
11 median filter with a radius of 4. This filter smoothed the image without blurring the edges. The  
12 resulting histogram exhibited 3 peaks. The first and second were thought to be associated with dark  
13 and light-grey solids, respectively, whereas the third, at lighter greyscale values, was clearly the resin.  
14 At that stage, two different approaches were adopted. In the first, based on De Gryze et al. (2005),  
15 four areas of light-grey values and 4 areas of resin were selected. From each area, the histogram was  
16 copied and pasted to a spreadsheet. In this spreadsheet, each count of the 256 grey values was  
17 divided by the sum of all the counts for that area. This resulted in 8 histograms with relative counts, 4  
18 for the resin and 4 for the light-grey solids out of which two average histograms were calculated. The  
19 means and standards deviation of the 2 histograms were used to calculate the two associated normal  
20 distributions. The threshold value was selected as the intersection of the two normal distributions,  
21 and was found to be 156. In the second method, the histogram of the whole image was used. The  
22 threshold value was chosen to be the grey value, 138, of the second peak. Even though method 1 was  
23 the standard method in Expert 1's group, this second approach yielded a "visually better" threshold.

24 *Expert 2*

25 Image manipulation was carried out with the software package Image-Pro Plus version 5.0 (Media  
26 Cybernetics, Inc., Bethesda, Maryland). For the Pathway image, the hue channel was extracted in the

1 HSD (hue-saturation-darkness) color decomposition. The greyscale values of pixels deemed to be  
2 part of the background were replaced by a value close to the mean background intensity, according to  
3 the formula

$$4 \quad CI_{x,y} = I_{x,y} - BI_{x,y} + M \quad (1)$$

5 where  $CI_{x,y}$  is the new pixel value in the corrected image at location  $(x,y)$ ,  $I_{x,y}$  is a pixel value of the  
6 original image,  $BI_{x,y}$  is a pixel value of the background image, and  $M$  is the average pixel value of the  
7 background image. Following background correction, the image was thresholded using Image-Pro  
8 Plus's default, global, histogram-based algorithm. In the resulting image, some plant roots presented  
9 at the upper-left corner of the profile were mistakenly identified as preferential flow path, which  
10 requires visual inspection and further manual correction. Processing of the Thinsect image also  
11 involved the use of Image-Pro Plus's default histogram-based algorithm, without any additional  
12 manual fine-tuning. For the Nanotom image, manipulation involved a transformation from RGB to  
13 HSD color decomposition, followed by extraction of the hue channel and global thresholding as  
14 described earlier.

15 *Expert 3*

16 To verify the results of automated segmentation, this expert first requested information to clarify  
17 what one should consider to be foreground, transition, and background in the images. Each test  
18 image was sent back with one or more areas circled in different colors, and the expert asked for  
19 confirmation by the senior author that the proposed classification of these areas was correct. On the  
20 basis of that information, thresholding of the various test images was carried out. For the Pathway  
21 image, the initial RGB format was converted to YCrCb (luminance and chrominance) color  
22 decomposition (also often referred to as the YUV standard). The Cb (blue) chrominance channel was  
23 extracted. Denoising was carried out via anisotropic diffusion, using the noise distribution from the  
24 circled area associated with the background. Then, Otsu's (1979) algorithm was applied to the filtered  
25 image. In a second approach, Otsu's (1979) algorithm was applied after filtering the Cb chrominance  
26 channel using a 3x3 Wiener filter to remove noise (instead of using the noise distribution from the

1 circled background area, the noise was presumed to be Gaussian across the whole image). In the case  
2 of the Thinsect image, an identical approach to denoising and thresholding leads to two thresholded  
3 images, the second one again resulting from the use of a 3x3 Wiener filter. For the Nanotom image,  
4 the same general procedure was followed once again, except that in this case, four different noise  
5 reduction techniques were tried. Approach 1 consisted of applying Otsu's (1979) algorithm to the  
6 anisotropic-diffusion filtered image. In approach 2, the image resulting from approach 1 was filtered  
7 with a 3x3 median filter to reduce salt and pepper noise. Approach 3 was identical to approach 2  
8 except that a 5x5 median filter was used. In approach 4, the original image was filtered with a 3x3  
9 Wiener filter, and the resulting image was then thresholded with Otsu's (1979) algorithm.

#### 10 *Expert 4*

11 The software analySIS (Soft Imaging System, GmbH) was used for all image manipulations. For the  
12 Pathway image, the first step was to use the automated optimize contrast tool to allow the stained  
13 area to be more easily identified. The image was then transformed from RGB to the HSI  
14 (Hue/Saturation/Intensity) colour space so the individual Intensity (I) channel image could be  
15 acquired. Previous research by expert 4 suggested that this approach is usually successful in  
16 highlighting the dye tracer, Brilliant Blue. Following this, because of the successful image  
17 enhancement obtained by the previous steps, no further processing was required and the image was  
18 thresholded manually (the auto tool under-estimated the dye stained area). This process involved the  
19 use of the "include pixel" and "exclude pixel" options in the software to clearly segment stained areas  
20 from non-stained areas. Using this method no areas of overlap, i.e., stained areas appearing as non-  
21 stained, or the converse, were identified. In the case of the Thinsect image, the automated 'enhance  
22 contrast' tool was applied, which promoted the separation of the pore and solid material observed  
23 from the greyscale histogram. A series of further filtering operations available as standard options  
24 within the software, including median, low-pass and edge enhancement, were evaluated. However  
25 these all appeared to decrease separation of the pore and solid space quality rather than enhance it.  
26 Therefore the original image with just one pass of the 'enhance contrast' tool was thresholded using

1 the 'auto' algorithm of analySIS to generate the binary image. The Nanotom image was first  
2 converted to greyscale and the contrast was also optimised using the automated tools in analySIS.  
3 Again, the standard enhancement filters were tested but did not enhance the segmentation results  
4 therefore the analySIS auto algorithm was again used for thresholding. However, it was considered  
5 (by comparison with the original image) that the thresholded binary image contained a significant  
6 portion of noise which was leading to an over-estimation of pore space. Therefore further binary  
7 morphology image processing was undertaken to remediate observed artefacts. An erosion filter set  
8 to a 1 pixel radius was applied on two occasions to the original thresholded image to arrive at the  
9 final result.

#### 10 *Expert 5*

11 This expert only worked on the Pathway image. To maximize the contrast between stained and  
12 background soil material, the storage format of the Pathway image was changed from RGB to cyan-  
13 yellow-magenta-black (CYMK), and the cyan channel was retained for further analysis. This channel,  
14 with no other pre-treatment, was thresholded separately in method 1, using the intermeans algorithm  
15 (Ridler and Calvard, 1978; Trussel, 1979), and in method 2, with the minimum-error algorithm (Kitler  
16 and Illingworth, 1986).

#### 17 *Expert 6*

18 This expert used the commercial software package ENVI 4.6.1 (ITT Visual Information Solutions,  
19 Boulder, Colorado) for all image manipulations. In the case of the Pathway image, a simple band  
20 ratio (red/green) was calculated, followed by isodata unsupervised classification with 5-10 classes,  
21 using ENVI's default settings for the parameters associated with this procedure (maximum  
22 iterations= 95, % change threshold = 5, min # of pixels in a class = 10, max class stdv = 1.000, min  
23 class distance = 5.00, max # merge pairs = 2). In the output map, the 5 classes identified were  
24 combined into 2 classes after visual comparison with the original image, to produce a binary  
25 representation. For the Thinsect image, ENVI's procedure for density slicing was used, with default  
26 settings. Two groupings of the resulting classes appeared equally plausible, after visual comparison

1 with the original image. One involved considering only the last density slice (with greyscale values  
2 between 217 and 248) as pores. The other involved the last two slices (186 to 216 and 217 to 248).  
3 For the Nanotom image, the isodata unsupervised classification was applied once again, using default  
4 settings. Among the 7 resulting classes, visual inspection of the original image suggested that classes  
5 1-3 should be grouped as solids, and classes 4 -7 should be merged to become the pore space.

6 *Expert 7*

7 The commercial software Adobe Photoshop 3.0 (Adobe Systems, Inc., San Jose, California) was used  
8 to process all three images. For the Pathway image, the red channel (RGB color decomposition) was  
9 selected. Its contrast was enhanced by stretching the range of its greyscale from 0-197 to 0-255. The  
10 image was then divided into two equal parts along its width. Both bottom and top parts were  
11 separately “equalized” [Adobe’s Equalize Command finds the brightest and darkest values in the  
12 image and averages all the brightness values so that the darkest value represents black (or as close to  
13 it as possible) and the brightest value represents white. Photoshop then attempts to equalize the  
14 brightness, that is, to distribute the intermediate pixel values evenly throughout the greyscale.] Then  
15 the two sub-images were thresholded manually, by visual comparison with the original image. For the  
16 Thinsect image, the “Equalize” command was used once again, followed by application of a Gaussian  
17 blur filter with a radius of 1 pixel. The resulting image was then thresholded manually. In the case of  
18 the Nanotom image, the same procedure was repeated, except for the fact that a radius of 3 pixels  
19 was selected for the Gaussian blur filter.

20 *Expert 8*

21 The commercial software Adobe Photoshop CS3 Extended version 10.0.1 (Adobe Systems, Inc., San  
22 Jose, California) was used to process all three images. The cyan channel of the Pathway image, and  
23 the single greyscale channel of the other two images, were all thresholded manually by trial and error,  
24 and visual comparison with the original images, until what appeared to be a satisfactory binary image  
25 was obtained.

26 *Expert 9*

1 This individual declined to work on the Pathway image, claiming a lack of experience with color  
2 image processing. The other two images were manipulated using the software ImageJ 1.41 (NIH,  
3 Bethesda, Maryland). A median filter with a radius (kernel size) of 3 pixels was used to filter out the  
4 noise. The resulting image was thresholded with ImageJ's built-in algorithm. This algorithm requires  
5 a starting value, which was generated by the "Auto function", an iterative procedure based on the  
6 isodata algorithm (Ridler and Calvard, 1978). The threshold value recommended by this approach  
7 was manually adjusted to ensure the presence and connectedness of small and thin valley pores in soil  
8 aggregates. Finally, the "Despeckle" procedure was applied to remove isolated groups of pixels,  
9 which correspond to irregularities in the resin material.

10 *Expert 10*

11 This individual declined to work on the Pathway image, claiming a lack of experience with color  
12 image processing. The other two images were manipulated using the software ImageJ 1.41 (NIH,  
13 Bethesda, Maryland). The "wand tracing tool" was used, which when centered on a given point in an  
14 image allows one to identify the greyscale value at that point. The wand tool was directed to areas of  
15 pore space and solid to ascertain the greyscale value ranges of each. Values were obtained inside  
16 homogeneous areas, then at transition points where air meets solid. Then the "threshold histogram"  
17 procedure was carried out, using the default seed value supplied by ImageJ. This value was within  
18 the ranges previously determined with the wand, making it unnecessary to perform "cleanup" using  
19 the "Despeckle" procedure. The expert then compared the thresholded and original images in a  
20 number of areas of the images, and adjusted the threshold manually until a visually satisfactory  
21 outcome was obtained. For the particularly noisy Nanotom image, this approach was adopted with  
22 and without initial clean up (or denoising). These two options are labelled as Method 1 and 2,  
23 respectively. The process without denoising was thought by the expert to eventually lead to a better  
24 image, characterized by noise but without significant loss of the solid phase.

25 *Expert 11*

1 This individual declined to work on the Pathway image, claiming a lack of experience with color  
2 image processing. The other two images were manipulated using the software ImageJ 1.41 (NIH,  
3 Bethesda, Maryland). Threshold values were chosen by visual comparison of original and thresholded  
4 images. In the Nanotom image, a median filter (with a 2.5 pixel radius) was applied, followed by  
5 despeckling.

6 *Expert 12*

7 This individual declined as well to work on the Pathway image, claiming a lack of experience with  
8 colour image processing. The other two images were manipulated using the software MATLAB and  
9 3DMA-Rock (Oh and Lindquist, 1999). The four methods used were the iterative method (Riddler  
10 et al., 1978), Otsu's (1979) method, the entropy-based method (Sahoo et al., 1997) and the  
11 indicator kriging method (Oh and Lindquist, 1999). Moreover, a 60% majority filter was applied to  
12 remove noise in the resulting images for each segmentation method. For the Thinsect image, user-  
13 specified thresholds (130, 170) were selected for the segmentation using the indicator kriging method.  
14 For the Nanotom image, because of the unimodal pattern of the grey scale distribution, the entropy  
15 and indicator kriging methods were not applied.

16 *Expert 13*

17 The images were processed using the image analysis toolbox in MATLAB® (The MathWorks, Inc.,  
18 Natick, Massachusetts). For the Pathway image, the YCbCr color space was used to analyze the  
19 input image using the built-in MATLAB function `rgb2ycbcr`. The YCbCr format stores luminance  
20 information in the Y component whereas the chrominance information is stored as two color-  
21 difference in the Cb and Cr components. Histogram analysis of all the channels in the image showed  
22 that the Cr component (difference between red channel and a reference value) provided the  
23 maximum contrast (sensitivity) and was therefore selected for thresholding the image data. The data  
24 range for Cr channel is in the range [16, 240]. Simple thresholding was then conducted visually to  
25 obtain a threshold value range of [120, 133] and produce a binary image. The Thinsect image was  
26 thresholded using a similar process by testing the sensitivity of the processed binary image to

1 histogram plots by visual inspection. The binary image was produced using a threshold value of 218,  
2 just below the inflexion point in the histogram. The Nanotom image was analyzed by first converting  
3 it from three channels (in RGB colour decomposition) to a single greyscale channel using the built-in  
4 `rgb2grey` function in MATLAB. This function uses the following formula to convert from three  
5 channels to one:

$$6 \qquad 0.2989 \cdot C_1 + 0.5870 \cdot C_2 + 0.1140 \cdot C_3 \qquad (2)$$

7 where  $C_1$ ,  $C_2$  and  $C_3$  are the first, second and third channels respectively. Manual thresholding, based  
8 on visual inspection of the histogram, was used again and yielded a threshold value of 160.

## 9 **References**

- 10 Boast, C., Baveye, P. 2006. Alleviation of an indeterminacy problem affecting two classical iterative  
11 image thresholding algorithms. *International Journal of Pattern Recognition and Artificial*  
12 *Intelligence*, 20, 1-14.
- 13 De Gryze, S., Jassogne, L., Six, J., Bossuyt, H., Wevers, M., Merckx, R. 2005. Pore structure changes  
14 during decomposition of fresh residue: X-ray tomography analyses. *Geoderma*, 134, 82-92.
- 15 Kittler, J., Illingworth, J. 1986. Minimum error thresholding. *Pattern Recognition*, 19, 41-47.
- 16 Oh, W., Lindquist, B. 1999. Image thresholding by indicator kriging. *IEEE Transactions on Pattern*  
17 *Analysis and Machine Intelligence*, 21, 590-602.
- 18 Otsu, N. 1979. Threshold selection method from grey-level histograms. *IEEE Transactions on*  
19 *Systems Man and Cybernetics*, 9, 62-66.
- 20 Ridler, T.W., Calvard, S. 1978. Picture thresholding using an iterative selection method. *IEEE*  
21 *Transactions on Systems Man and Cybernetics*, 8, 630-632.
- 22 Sahoo, P., Wilkins, C., Yeager, J. 1997. Threshold selection using Renyi's entropy. *Pattern*  
23 *Recognition*, 1(1), 71-84
- 24 Trussell, H.J. 1979. Picture thresholding using an iterative selection method - comments. *IEEE*  
25 *Transactions on Systems Man and Cybernetics*, 9, 311-311.

26