



CrossMark
click for updates

Research

Cite this article: Aydogan DB, Hyttinen J. 2014 Characterization of microstructures using contour tree connectivity for fluid flow analysis. *J. R. Soc. Interface* **11**: 20131042. <http://dx.doi.org/10.1098/rsif.2013.1042>

Received: 11 November 2013

Accepted: 5 March 2014

Subject Areas:

mathematical physics, biomaterials

Keywords:

three-dimensional characterization, microstructure, image analysis, structure–property relationship, connectivity

Author for correspondence:

Dogu Baran Aydogan

e-mail: baran.aydogan@tut.fi

Characterization of microstructures using contour tree connectivity for fluid flow analysis

Dogu Baran Aydogan^{1,2} and Jari Hyttinen^{1,2}

¹Department of Electronics and Communications Engineering, Tampere University of Technology, Tampere, Finland

²BioMediTech, Tampere, Finland

Quantifying the connectivity of material microstructures is important for a wide range of applications from filters to biomaterials. Currently, the most used measure of connectivity is the Euler number, which is a topological invariant. Topology alone, however, is not sufficient for most practical purposes. In this study, we use our recently introduced connectivity measure, called the contour tree connectivity (CTC), to study microstructures for flow analysis. CTC is a new structural connectivity measure that is based on contour trees and algebraic graph theory. To test CTC, we generated a dataset composed of 120 samples and six different types of artificial microstructures. We compared CTC against the Euler parameter (EP), the parameter for connected pairs, the nominal opening dimension (d_{nom}) and the permeabilities estimated using direct pore scale modelling. The results show that d_{nom} is highly correlated with permeability ($R^2 = 0.91$), but cannot separate the structural differences. The groups are best classified with feature combinations that include CTC. CTC provides new information with a different connectivity interpretation that can be used to analyse and design materials with complex microstructures.

1. Introduction

Several engineered and natural processes such as filtering, cell seeding into biomaterials and groundwater flow are governed by transport through porous media. Most relevant physical processes of flow take place at the micro or pore scale. This detailed morphological information can be obtained using high-resolution three-dimensional imaging devices such as microcomputed tomography (μ -CT) [1]. The capability of the imagers to reveal pore scale details has made three-dimensional microstructure imaging very attractive. As a result of the rapidly expanding number of devices, there is a growing demand for effective image processing techniques to process and analyse binary images [2].

Among the characterizations of porous medium, understanding the role of connectivity has become a very active research topic during the past decade. A paradigm shift towards studying connectivity has been pointed out earlier in the fields of hydrogeology, surface hydrology, geomorphology, landscape ecology and pore scale or soil physics [3]. Connectivity is an important property to study because it complements the information from local spatial variances and higher-order correlations that are traditionally used to model heterogeneity. One problem associated with connectivity is the lack of a unique mathematical definition. Being an intuitive notion, the quantification of connectivity is based on how it is interpreted. For example, it is common to refer to static, dynamic, local and global measures of connectivity [3]. Among the static methods, notable approaches include the Euler number [4], percolation theory [5] and connectivity function [6]. Dynamic connectivity metrics involve the physical processes of flow or transport. They are estimated experimentally or with modelling such as effective hydraulic conductivity [7].

Static and scalar measures of connectivity can be calculated from the binary images of microstructures. Static measures are not only faster, but also cheaper to obtain compared with dynamic connectivity measures. Additionally, scalar

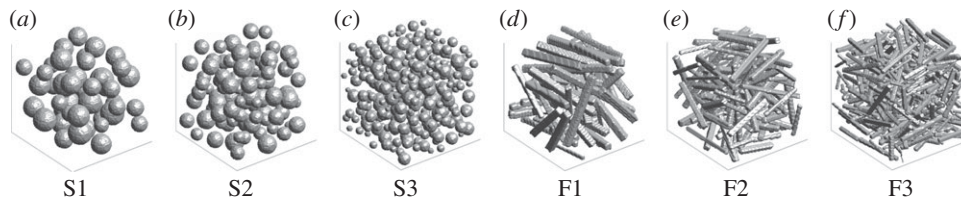


Figure 1. Sample visualizations of the six types of three-dimensional structures.

values are easier to understand and use. However, static and scalar connectivity measures are very limited in the literature. To the best of our knowledge, the proportion of connected pairs (Γ) from percolation theory and the Euler number (χ) are the only scalar measures described in the literature.

High-resolution μ -CT images of many microstructures have sharp boundaries, which makes it straightforward to segment permeable (pores) and impervious (material) regions using simple techniques such as thresholding [2]. The proportion of connected pairs is expressed as $\Gamma = (1/n^2) \sum_{i=1}^N n_i^2$ and shows the ratio of the pairs of cells that are connected to all the pairs of permeable cells. Here, n is the total number of permeable cells, N is the number of connected permeable regions and n_i is the number of permeable cells in the i^{th} region. One common case in which Γ does not provide information is when there is a medium with a single fully connected permeable region, for which $\Gamma = 1$.

The most widely accepted and commonly used connectivity measure for binary images is the Euler number that is described with the Euler–Poincaré formula as the alternating sum of Betti numbers, β_k , $\chi = \sum_k (-1)^k \beta_k$. For two-dimensional images, χ is simply the number of connected components (β_0) minus the number of holes (β_1). For three-dimensional, it is the number of connected components (β_0), minus the number of tunnels (β_1), plus the number of voids (β_2). Although the Euler number is an elegant topological invariant, topology alone is not sufficient to describe connectivity in most cases. For example, although the Euler number includes information regarding the number of connected components, information about the proximity of the connected components is not provided. In addition, the Euler number can give information about holes, but not cavities (holes are topologically not connected to the outside, cavities are).

To overcome the limitations of the available approaches, we use contour trees and develop an alternative static and scalar connectivity measure called the contour tree connectivity (CTC) [8]. Contour trees are topological abstractions that follow the evolution of level sets as they appear, join, split or disappear by means of graphs. They have mostly been used to simplify volumes and to visualize high dimensional scalar fields [9–14].

An application of contour trees for analysing microstructures, or binary images in general, is given in our earlier work [15]. In reference [15], a graph representation of microstructures is proposed that takes into account both the geometrical properties of foreground object(s) and how the object(s) is/are placed in the background. This representation of images transforms complex three-dimensional microstructures into simple graphs. In a follow-up study, we developed CTC that calculates the connectivity of binary images by calculating the connectivity of their contour tree representations [8]. The first application of CTC is given for the analysis of trabecular bone microstructure [16]. In reference [16], it is shown that CTC is

the third best morphometric parameter to predict ultimate bone strength among the 10 most commonly used measures.

In this study, we assess and compare the capability of various geometrical characteristics including CTC in relating structural properties with the permeability of microstructures. To achieve this goal, we generate 120 samples that represent six different types of microstructures with equal porosities using Matlab. We correlate the nominal opening dimension (d_{nom}), CTC, the EP, the parameter for connected pairs (PCPs), and the permeabilities estimated using direct pore scale modelling. We show that the connectivity interpretation and information provided by CTC is different from the other measures and can be used to quantify connectivity in order to identify the differences and changes in microstructures.

2. Description of materials

We prepared random porous structures that resemble segmented microstructures obtained from a μ -CT device with an isotropic voxel spacing of 7.8125 μm . All samples were designed to have physical dimensions of $1 \times 1 \times 1 \text{ mm}$ which correspond to $128 \times 128 \times 128$ voxels in image domain. In order to quantify connectivity independently of porosity, all samples were designed to have approximately 85% porosity. Porosity was calculated as the ratio of the number of permeable voxels to the number of all voxels.

We generated six groups of data in which each group contains 20 samples. Three of the groups contain samples made of spheres and the rest of the three contain samples of fibres. We abbreviate the sphere samples as S1, S2, S3 and the fibre samples as F1, F2, F3. S1 and F1 have 45 big disconnected objects, S2 and F2 have 100 medium-sized disconnected objects and S3 and F3 have 300 small disconnected objects.

For S1, S2 and S3, the diameters of the spheres vary between 15–26, 9–20 and 3–14 voxels, respectively. All fibres in F1, F2 and F3 are square prisms with a varying ratio of width to height between 10 and 20. The widths of fibres in F1, F2 and F3 vary between 6–8, 4–6 and 2–4 voxels, respectively. Figure 1 shows sample images from each group.

We designed and generated the images using in-house developed Matlab software. Starting with an empty image, our code iteratively places a sphere or a fibre of desired dimensions in a random fashion wherever there is no collision. The iteration stops when the necessary number of objects are placed in the image.

3. Characterization of microstructures using image processing

We calculated four geometrical characteristics of microstructures. These were (i) nominal opening dimension, (ii) CTC, (iii) EP and (iv) PCPs.

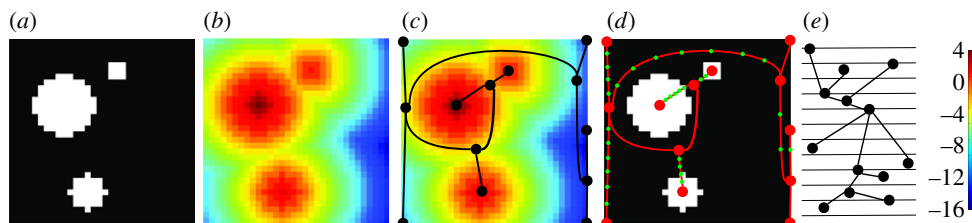


Figure 2. (a) Input image; (b) Euclidean distance transform (EDT) of input; (c) contour tree drawn on top of the EDT image; and (d) supplemented contour tree (SCT) with red nodes from the contour tree and green nodes from supplementation. (e) Contour tree with level lines and the corresponding colour bar for EDT images shown in b and c.

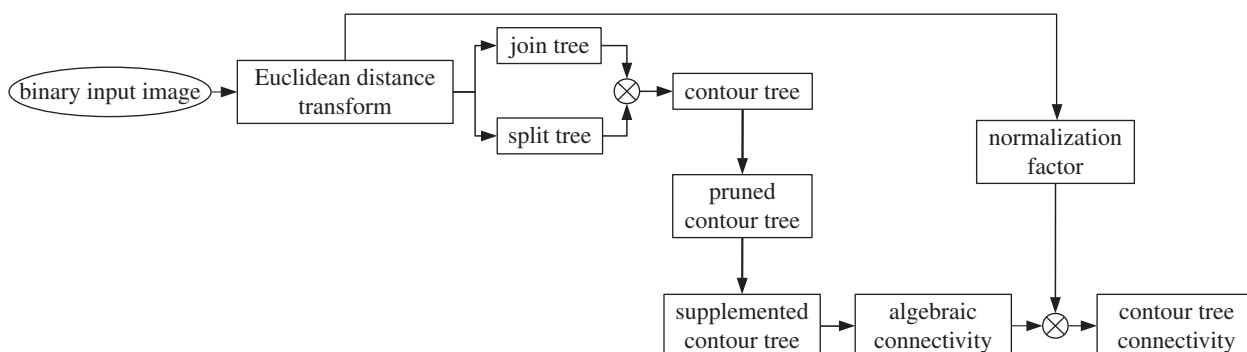


Figure 3. Flowchart for CTC feature extraction. From a binary input image, firstly EDT, then SCT and finally the CTC is calculated as the normalized algebraic connectivity of SCT.

3.1. Nominal opening dimension

d_{nom} is a morphological feature which is also used as an estimate for the characteristic length scale [17]. d_{nom} is computed based on the opening operator which eliminates features smaller than the structure element. By iteratively using the opening operator with increasing sizes of spheres as structure elements, the opening size distribution is obtained [4].

Let $\mathcal{O}_d(\cdot)$ denote the opening operator with a spherical structure element with diameter d . Then, the cumulative opening size distribution of image I is a function of d that is defined as

$$O(d) = \sum \text{xor}(I, \mathcal{O}_d(I)), \quad (3.1)$$

where xor denotes the *exclusive or* operator. The opening size distribution is $o(d) = O'(d)$. Using $o(d)$, d_{nom} is calculated by the following expected value

$$d_{\text{nom}}(I) = \frac{\sum o(d) \cdot d}{\sum o(d)}. \quad (3.2)$$

3.2. Contour tree connectivity

The main idea of CTC is to compute the structural connectivity of a binary image by computing the connectivity of its contour tree representation. Figure 2 shows an example of a contour tree representation on a small two-dimensional image, and figure 3 shows the flowchart for extracting CTC.

The connectivity of a contour tree is calculated using algebraic graph theory. The formal expression for CTC is given as follows [8]

$$\text{CTC} = \frac{\lambda_1(\mathcal{G})}{2 \left[1 - \cos \left(\frac{\pi}{\lfloor l_{\text{max}}^{\mathcal{G}} \rfloor - \lfloor l_{\text{min}}^{\mathcal{G}} \rfloor + 1} \right) \right]} \quad (3.3)$$

where \mathcal{G} is the supplemented contour tree (SCT). λ_1 is the algebraic connectivity of \mathcal{G} . $l_{\text{max}}^{\mathcal{G}}$ and $l_{\text{min}}^{\mathcal{G}}$ are the global max

and min of Euclidean distance transform (EDT) image and $\lfloor \cdot \rfloor$ is the rounding operator.

SCT is a concept similar to the augmented contour tree [9]. However, unlike augmented contour tree, the levels of the vertices in SCT do not necessarily exist in the EDT image. Because the denominator of equation (3.3) is the algebraic connectivity of the path from global max to global min of the EDT image, it sets the upper bound for any λ_1 with the same global extrema. Therefore, $0 < \text{CTC} \leq 1$.

The formal definition of SCT and other technical details to extract CTC were explained in reference [8]. The only difference compared with reference [8] is that in this work, the pruning step is fully automated by setting the level threshold to 0.1 and volume threshold to 64. The complete flowchart used to extract CTC is shown in figure 3.

3.3. Euler parameter

The Euler number (χ) is the most commonly used connectivity feature in the literature. χ is calculated using graph counts as follows [18]

$$\chi = v - e + f - s, \quad (3.4)$$

where v is the number of vertices, e is the number of edges, f is the number of faces and s is the number of solids calculated in three-dimensional with six-connectivity.

Note that, for each group of test samples, permeable regions have a constant Euler number. For example, permeable regions of all the test samples in group S1 and F1 have $\chi = 46$, because there is a single connected permeable region with 45 holes. In order to include information regarding the local variations in structures, the Euler number is generally calculated on excursion sets obtained by thresholding scalar fields [4,19,20]. The scalar fields for our test samples are obtained using EDT. A series of thresholded images along with corresponding χ -values are shown in figure 4 for a sample from group S1.

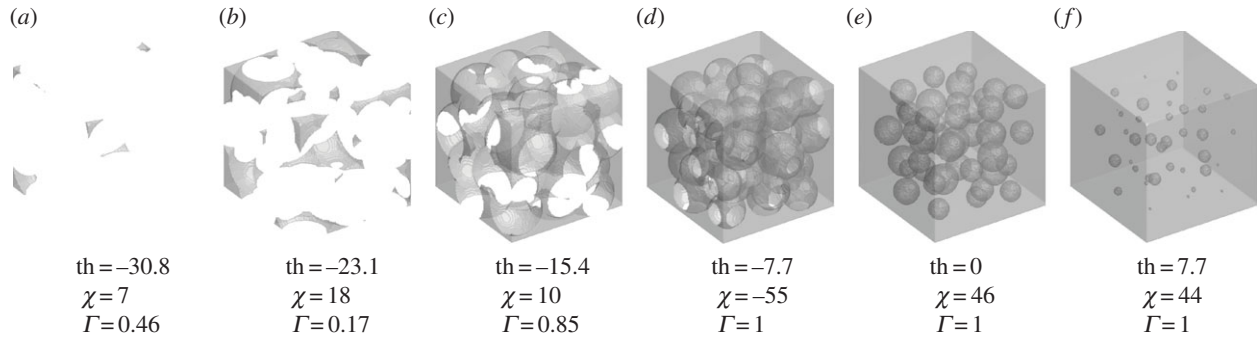


Figure 4. Euler number (χ) and proportion of connected pairs (Γ) corresponding to the excursion sets obtained by thresholding EDT of a sample from group S1. From left to right the value of the threshold (th) is increased. At $th = 0$, the resultant image is the image of the permeable region.

By repeating the experiments for 400 threshold levels that are linearly separated between the minimum and maximum values of each scalar field, functions for χ were obtained. An example is shown in figure 5a that is applied on the sample shown in figure 4. As the EP, we used the following simple expression $EP = 1/N \sum_{i=1}^N |\chi(i)|$, where $N = 400$.

3.4. Parameter for connected pairs

The proportion of connected pairs (Γ) is a commonly used feature to characterize structural connectivity as explained in §1. However, for our test images, because all permeable regions are made of a single connected component, $\Gamma = 1$ for all samples. In order to include the effects of local variability, we apply the same approach done in the computation of EP. Similar to χ , this iterative way of computing Γ was also done earlier, for example recently in reference [3]. We calculate Γ using the same excursion sets and thresholds used for the EP, as shown in the example of figure 4. The resulting Γ profile is shown in figure 5b. As the parameter of connected pairs, we used the following expression $PCP = 1/N \sum_{i=1}^N \Gamma(i)$, where $N = 400$.

We calculate Γ values using a short Matlab script which is based on the regionprops command. This command calculates the number of permeable cells in each connected component when called with the area parameter.

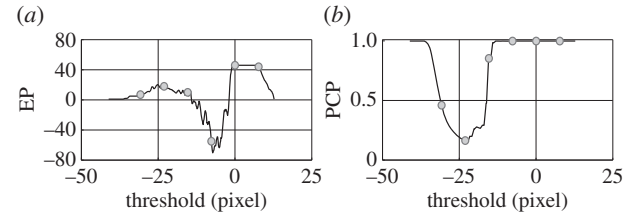


Figure 5. (a) Euler number (χ) and (b) proportion of connected pairs (Γ) as a function of threshold. The functions are obtained using the sample in figure 4. Dots on the graphs indicate the location of the χ - and Γ -values calculated in figure 4.

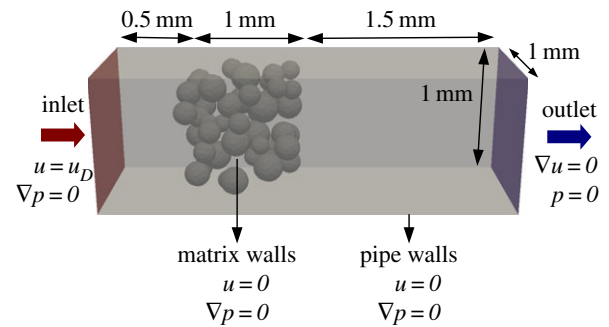


Figure 6. The set-up used in the models. The figure shows the initial and boundary conditions for velocity and pressure.

4. Estimation of permeability using direct pore scale modelling

Direct pore scale modelling or pore-level numerical simulation (DPLS) has become a popular method to estimate permeability during the past few decades owing to its speed and low cost compared with conducting experiments [21]. The common approach used in DPLS is based on the set-up shown in figure 6 and Darcy's law, which states that

$$\nabla p = -\frac{\mu}{K} u_D, \quad (4.1)$$

here ∇ is the *del* operator, p is the pressure (N m^{-2}), μ is the dynamic viscosity ($\text{kg m}^{-1} \text{s}^{-1}$) of the fluid, K is the permeability (m^2) of the porous media and u_D is the Darcian velocity (m s^{-1}) of the fluid.

The main goal of the numerical simulation is to calculate the pressure distribution inside the porous medium for a given velocity. Therefore, the model requires an input u_D , in addition to the fluid parameters such as μ . The value for u_D is determined based on an assumption on the Reynold's number (Re). Because

equation (4.1) is valid only for low fluid velocities, turbulence is neglected and Reynold's number is assumed to be less than 2. Followed by this assumption, u_D can be estimated from the modified expression for the Reynolds number:

$$Re = -\rho \frac{u_D d_{\text{nom}}}{\epsilon \mu} \Rightarrow u_D = -\frac{\epsilon \mu Re}{d_{\text{nom}} \rho} \quad (4.2)$$

where ρ is the density of the fluid (kg m^{-3}), ϵ is the porosity and d_{nom} is the nominal opening dimension (m) of the medium.

Lastly, by combining equations (4.1) and (4.2), K can be expressed as

$$K = \frac{\epsilon \mu^2 Re}{d_{\text{nom}} \rho \nabla p} \quad (4.3)$$

here the only unknown is the pressure gradient, which can be estimated by a Navier–Stokes solver. The governing equations are

$$\begin{aligned} \nabla \cdot \mathbf{v} &= 0 \\ \rho \left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) - \mu \nabla^2 \mathbf{v} &= -\nabla p \end{aligned} \quad (4.4)$$

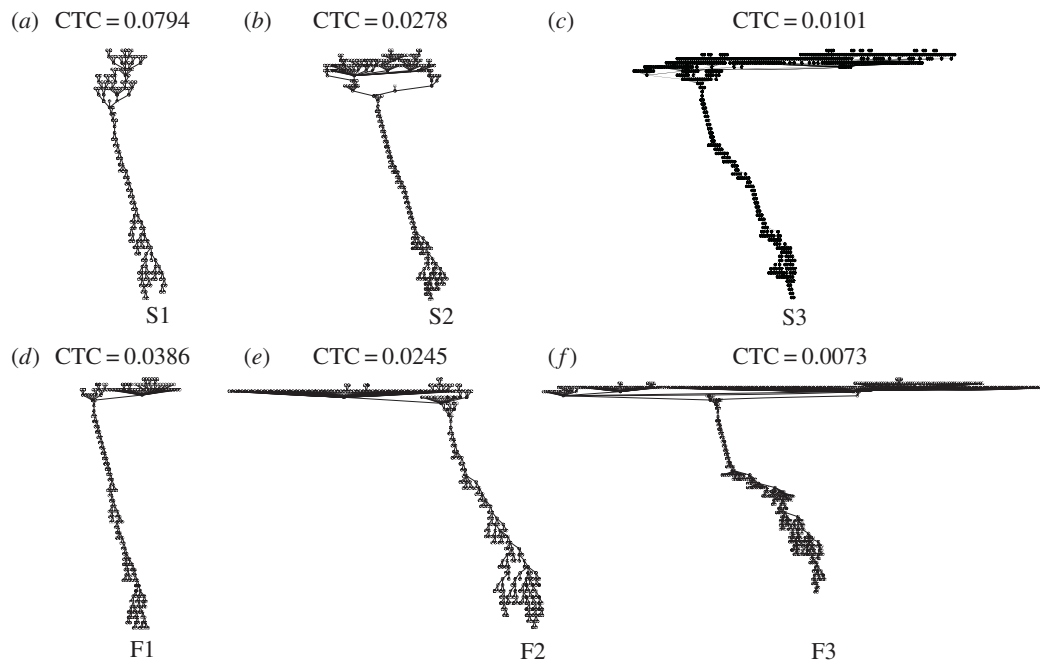


Figure 7. Example contour trees and CTC values. The trees belong to the samples shown in figure 1. Increasing the number of objects, as in S1 to S3 or F1 to F3, introduces more details in the graph which decreases CTC.

where \mathbf{v} is the velocity vector. Equation (4.4) couples velocity and pressure and can be solved by the semi-implicit method for pressure linked equation algorithm (SIMPLE), which is a popular approach applied to obtain a steady-state solution for incompressible fluid flow [22].

As a Newtonian, incompressible fluid, we use water in the models. Specifically, we use the fluid properties at 20°C with a density of $\rho = 998.2 \text{ kg m}^{-3}$ and a dynamic viscosity of $\mu = 0.001 \text{ kg m}^{-1} \text{ s}^{-1}$ [23]. In the models, water is flowed through a rectangular pipe of dimensions $1 \times 1 \times 3 \text{ mm}$. This set-up is shown in figure 4. The meshing of the geometry is done using the enGRID software [24]. As boundary conditions, 0 velocity on the sides of the tube and 0 pressure on the outlet is used. For the walls of the porous medium, no slip condition is assumed (figure 6). Turbulence is neglected and a laminar flow with $Re = 0.12$ is modelled. The initial condition for u_D at the inlet is calculated using equation (4.2). Time step for the simulation is determined by setting the Courant number to 0.1. OPENFOAM and SIMPLE algorithms are used for the solution of the pressure distribution.

5. Results

Here, starting with example contour trees and CTC values, all the values calculated for the parameters are given. In the last part of the results, we check the correlation between the parameters and perform a classification test to identify which of the connectivity features are useful to discriminate microstructures.

5.1. Contour trees and contour tree connectivity values

Figure 7 shows example contour trees and the corresponding CTC values for each of the six microstructure types shown in figure 1.

Samples with a fewer objects (S1 and F1) have fewer vertices in their contour trees owing to a smaller number of local maxima and minima in their EDT image (not shown). With the increase

in the number of objects, the number of local maxima and minima also increase. As a result, samples with more objects contain more vertices in their contour trees (S3 and F3). Increasing the number of pendant vertices decreases algebraic connectivity. Therefore, in general, increasing the number of objects in a random manner decreases CTC.

5.2. Permeability, nominal opening dimension, contour tree connectivity, Euler parameter, parameter for connected pair

Figures 8 and 9 show the results of the FE analysis. Figure 8a shows an example of the velocity distribution through the material, and the resulting pressure distribution is shown in figure 8b. A negative pressure gradient between 0.5 and 1.5 mm is shown in figure 8c where the material is placed. This trend is observed for all the samples. By using equation (4.3) and the obtained pressure gradients, permeabilities are estimated. Example values are shown in figure 8c.

The calculated permeability, d_{nom} , CTC, EP and PCP values are plotted for each test group in figure 9. Table 1 shows the means and variances of the parameters for each group.

From table 1, it is observed that among the test groups, S1 has the highest mean permeability and is followed by S2 and F1. The lowest mean permeability value belongs to group F3. F2 and S3 also have low permeability values. A similar trend is observed with d_{nom} and EP (inversely). However, CTC and PCP do not show this trend.

5.3. Correlation, clustering analysis and classification results

Figure 10 shows scatter plots between permeability, d_{nom} and CTC, EP, PCP. Table 2 shows the values for the coefficients of determination (R^2). The first column shows the R^2 -values calculated using all the test samples, and the rest of the columns show values corresponding to each group separately.

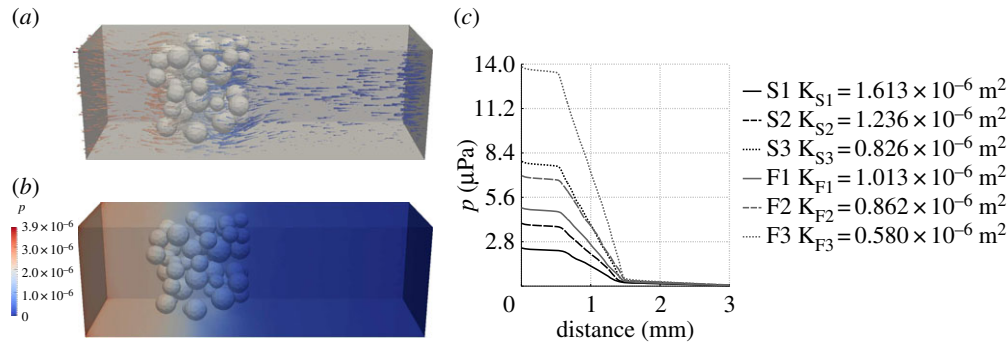


Figure 8. (a) Velocity distribution and (b) pressure distribution computed for the S1 sample shown in figure 1a. (c) Example pressure gradients and permeability values for samples belonging to different groups.

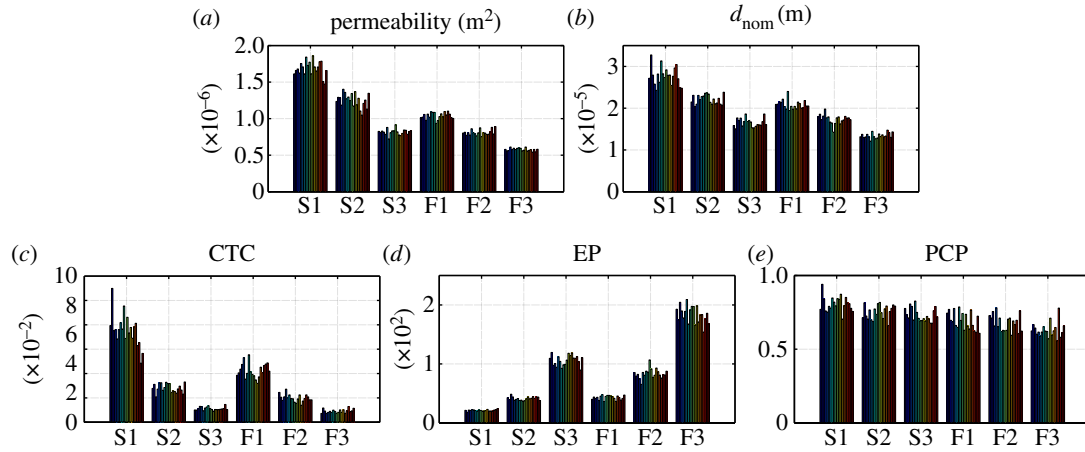


Figure 9. FE and image analyses results for all the test samples (a) permeability (b) d_{nom} (c) CTC (d) EP (e) PCP. For each group different colours represent different samples.

Table 1. Mean (μ) and variances (σ^2) of the parameters calculated for each group.

| | | S1 | S2 | S3 | F1 | F2 | F3 |
|-------------------------------|------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| permeability (m^2) | μ | 1.69×10^{-6} | 1.25×10^{-6} | 0.82×10^{-6} | 1.04×10^{-6} | 0.82×10^{-6} | 0.58×10^{-6} |
| | σ^2 | 9.90×10^{-15} | 8.69×10^{-15} | 1.67×10^{-15} | 2.08×10^{-15} | 1.49×10^{-15} | 0.39×10^{-15} |
| d_{nom} (m) | μ | 2.77×10^{-5} | 2.21×10^{-5} | 1.65×10^{-5} | 2.08×10^{-5} | 1.74×10^{-5} | 1.34×10^{-5} |
| | σ^2 | 4.88×10^{-12} | 1.25×10^{-12} | 1.05×10^{-12} | 1.11×10^{-12} | 1.22×10^{-12} | 0.41×10^{-12} |
| CTC | μ | 7.54×10^{-2} | 2.81×10^{-2} | 1.13×10^{-2} | 4.24×10^{-2} | 1.99×10^{-2} | 0.92×10^{-2} |
| | σ^2 | 1.73×10^{-4} | 0.13×10^{-4} | 0.02×10^{-4} | 0.37×10^{-4} | 0.09×10^{-4} | 0.02×10^{-4} |
| EP | μ | 21.24 | 41.28 | 106.13 | 42.98 | 83.62 | 184.06 |
| | σ^2 | 1.82 | 9.02 | 78.48 | 12.60 | 68.52 | 205.39 |
| PCP | μ | 0.81 | 0.75 | 0.73 | 0.69 | 0.68 | 0.63 |
| | σ^2 | 2.70×10^{-3} | 2.14×10^{-3} | 2.03×10^{-3} | 3.45×10^{-3} | 3.06×10^{-3} | 2.48×10^{-3} |

From figure 10a–d and table 2, it is observed that d_{nom} has the highest linear correlation with permeability ($R^2 = 0.91$). CTC, EP and PCP follows d_{nom} with R^2 -values of 0.78, 0.7 and 0.45, respectively.

In order to quantify how well the groups are separated in figure 10, we calculated the distances between the groups using the Bhattacharyya distance. The Bhattacharyya distance measures the overlap of two statistical populations as explained in appendix A. A larger Bhattacharyya distance indicates better separation of groups. Figure 11 shows the logarithm of the distances between groups. It is observed that although d_{nom} is the best predictor for permeability, d_{nom} –permeability combination is not able to separate S2 from F1 and S3 from F2. These groups are also not discriminated well with EP–permeability and

PCP–permeability combinations. CTC–permeability combination gives the largest distance between the similar groups. Using d_{nom} instead of permeability as the second feature decreases almost all the distances between groups. However, CTC– d_{nom} combination still provides a good contrast between groups. EP– d_{nom} combination is better at discriminating S3 and F2 compared with EP–permeability, however, S2 and F1 are still not separated well. PCP– d_{nom} combination does not provide useful information to separate different groups.

Lastly, we performed a classification test using the parameter combinations shown in figure 10. We used a supervised classification approach based on linear discriminant analysis (LDA). Half of the samples from each group were used to train a classifier, and the rest of the samples were

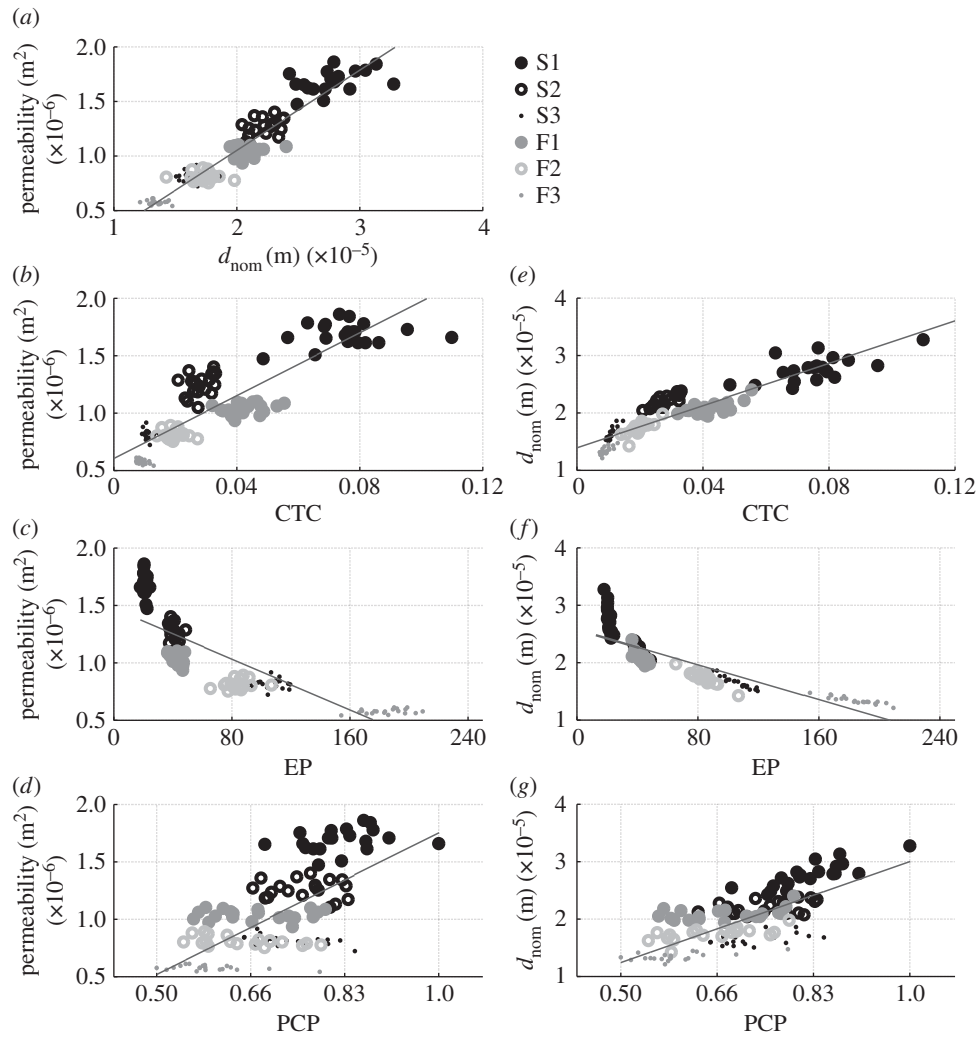


Figure 10. Scatter plots of different parameter combinations. Straight lines show the best linear fits calculated using all the samples. R^2 -values corresponding to the fits are given in table 2.

used for the tests. We repeated the experiments 100 times with different training and test groups. Figure 12 shows the average results for the classification tests and accuracies using each parameter combination. Rows in figure 12 show the true classes and columns show the predicted ones.

6. Discussions

The main goal of our study is to introduce CTC as an alternative quantitative measure for connectivity that can be used to relate structural properties with permeability. By restricting our designs to certain numbers of non-overlapping objects and porosity, we aimed to establish an intuition towards understanding CTC. In general, the use of complicated structures as in references [25,26] indeed creates a more realistic set-up. With such an arrangement however, it is difficult to explain why and how the value of the CTC varies within and among groups which is the goal of our work. CTC is mainly designed to complement the missing geometrical information of the Euler number. By using simpler structural designs, we successfully introduced local variations within groups that alter the permeability while keeping the Euler number (χ) and the proportion of connected pairs (I') the same and the porosity almost constant.

Structural properties, especially porosity and connectivity, have been previously reported to be well correlated with

permeability [27]. Because we aimed to study the effect of connectivity on flow that is independent of porosity, we designed six different types of microstructures with the same porosities (approx. 85%). To the best of our knowledge, our work is the first in which connectivity is analysed for different microstructures with similar porosities. We aimed to work with a statistically meaningful amount of data, and therefore we used 20 different samples for each group, 120 samples in total.

An AMD Phenom II X4 955 processor with 12 GB system memory was used to test the samples in this study. The average computation times for the nominal opening dimension (d_{nom}), CTC, the EP and the PCPs were approximately 27 min, approximately 6 min, approximately 21 s and approximately 39 s per sample, respectively. For CTC, the computation speed can be increased significantly and can be done in under a minute using optimized C++ code. Finite-element (FE) modelling, on the other hand, comprises several steps, some of which have to be done manually. The total time spent for the preparations, running and post processing took approximately 6 h per sample for the FE analysis, which is significantly longer than the time used by any image processing method.

The Euler number has been the most commonly used connectivity measure in the literature and has been reported to be closely linked to flow through porous media [19,28]. Despite being an elegant topological invariant, the Euler

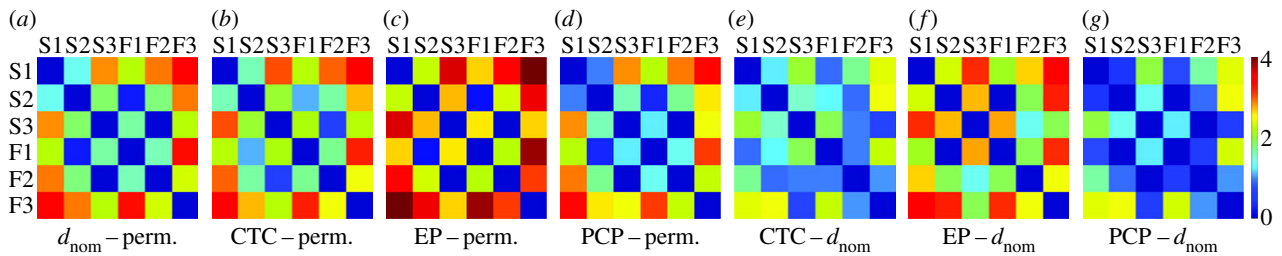


Figure 11. Logarithm of the Bhattacharyya distances between groups using different parameter combinations. Largest values of separation are obtained using EP. However, both EP–permeability and EP– d_{nom} feature combinations fail to separate S2 and F1. CTC provides a good contrast between groups that are difficult to separate by other parameter combinations.

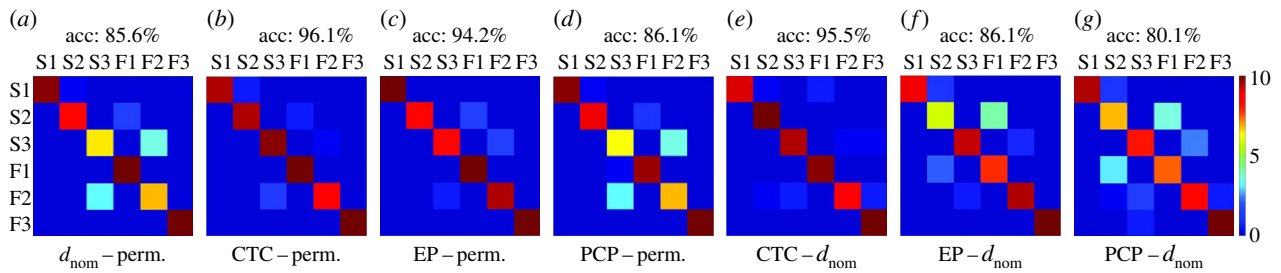


Figure 12. Results of the classification tests. Rows of the confusion matrices show the true classes and columns show the predicted ones. Best classification accuracy is obtained using CTC–permeability combination. EP also performs well when used together with permeability. In the absence of permeability information, the best classification result is obtained using CTC– d_{nom} combination.

Table 2. The coefficients of determination (R^2) obtained using linear correlation analysis. The first column shows the R^2 -values calculated using all the test samples. Other columns show the correlation between the parameter pairs for each group separately.

| | all | S1 | S2 | S3 | F1 | F2 | F3 |
|--|------|------|------|------|------|------|------|
| $d_{\text{nom}} - \text{permeability}$ | 0.91 | 0.17 | 0.22 | 0.00 | 0.02 | 0.00 | 0.04 |
| CTC – permeability | 0.78 | 0.08 | 0.14 | 0.03 | 0.08 | 0.01 | 0.25 |
| EP – permeability | 0.70 | 0.07 | 0.20 | 0.00 | 0.16 | 0.06 | 0.13 |
| PCP – permeability | 0.45 | 0.07 | 0.03 | 0.29 | 0.00 | 0.24 | 0.18 |
| CTC – d_{nom} | 0.84 | 0.40 | 0.70 | 0.66 | 0.47 | 0.61 | 0.51 |
| EP – d_{nom} | 0.78 | 0.65 | 0.82 | 0.92 | 0.60 | 0.88 | 0.85 |
| PCP – d_{nom} | 0.54 | 0.69 | 0.16 | 0.20 | 0.21 | 0.42 | 0.31 |

number has two major limitations when quantifying connectivity: (i) the distances between the connected components do not affect χ , and (ii) the holes in the structures change χ but cavities do not. Owing to these limitations, χ is constant for all the samples in each test group, and there is no variability within the groups. Similarly, there is no variance in the proportion of connected pairs, because for all the test samples there is only one connected component for the permeable region and $\Gamma = 1$. It is worth mentioning that the results from percolation theory and the Euler number were earlier found to yield similar information [29].

Because the Euler number and the proportion of connected pairs fail to explain the variation in permeability as static and scalar parameters, we computed χ and Γ iteratively on excursion sets. The resultant curves shown in figure 5 for χ and Γ were used earlier in references [4,19,20]. However, CTC and d_{nom} are scalars, and therefore in order to make comparisons, we used the average values of χ and Γ as the EP and the PCPs, respectively. Because we could not find any study that shows how the sign of χ relates to flow and permeability, we used the absolute value for computing the

EP. Another approach could have been to use the excursion sets and the curves obtained for the impervious region together with the permeable region as proposed in reference [3]. Although it is an important research question whether or not it is possible to extract better measures from these curves, this question is beyond the scope of this study.

The value of CTC varies owing to several factors that affect geometry and topology such as the number of connected components, their proximity to each other and the cavities or holes they have. The example contour trees shown in figure 7 demonstrate that, in general, compact graphs give larger CTC values and details decrease CTC. Although the information CTC provides is important in quantifying structural properties, we are also interested in correlating this information with permeability.

Table 2 and the scatter plots show the correlation between parameters. d_{nom} is the most correlated parameter with permeability ($R^2 = 0.91$). The scatter plots in figure 10e–g show the relationship between d_{nom} , CTC, the EP and the PCP. This demonstrates the case where costly FE modelling or experiments are not possible, and permeability values are

not known. Because d_{nom} is highly correlated with permeability, we investigated its relation with other parameters. The relationships between CTC, the EP and the PCP are not shown because they yielded very little information about flow and microstructures.

In order to visually explain how each parameter separates different groups, we plotted the Bhattacharyya distances in figure 11. It is observed that although d_{nom} is the most correlated parameter with permeability, it does not provide as useful information as CTC in terms of separating structural differences. For example, S3 and F2 have similar permeability and d_{nom} but with different structures. On the other hand, CTC is both correlated with permeability ($R^2 = 0.78$) and can separate differences in structures. Both the EP and the PCP have lower correlations with permeability. Additionally, they both fail to separate the structural differences within groups.

We used the Bhattacharyya distance, because it is commonly used as a class separability measure for feature selection [30–32]. In order to further verify that CTC information is useful in characterizing microstructures, we performed a different experiment that was not based on the Bhattacharyya distance. In this second experiment, a classification test was carried out using LDA that is a standard classification method [33]. The results of the classification tests show that structural properties are best separated using the CTC–permeability combination with an accuracy of 96.1%. Although this is not significantly larger than the EP–permeability combination (94.2% accuracy), the gap between CTC– d_{nom} and EP– d_{nom} becomes 9.4% which is the case where costly experimentation or modelling results are not available. It is worth mentioning that the results gained by using the Bhattacharyya distance and classification tests were consistent, and both showed that the feature combinations that use the EP are not as good as the ones that use CTC to separate S2 from F1 and S3 from F2.

7. Conclusion

CTC is a new structural measure that quantifies connectivity with a single real number that is between 0 and 1. In this work, we characterized different microstructures with CTC and showed that the connectivity information provided by

CTC is different from the commonly used measures based on the Euler number and the proportion of connected pairs.

The results of our analysis show that although mean opening dimension (d_{nom}) is highly correlated with permeability, it is not sufficient in discriminating the structural differences. The differences between groups are best explained with a feature combination that includes CTC. The Euler number-based parameter (EP) performed close to CTC where permeability information was available. Whereas in the case where permeability is not known, CTC performed significantly better compared with the EP and the PCPs.

Compared with experimental methods to estimate permeability or FE modelling, image processing approaches can be applied fully automatically and in very short times with little cost. With this work, we have shown that CTC can be used as a structural connectivity measure with some advantages over other available approaches. With its new connectivity interpretation, CTC is a promising connectivity measure that can be used to characterize microstructures, analyse materials and aid the design of new structures.

Acknowledgements. This work is conducted in the Department of Electronics and Communications Engineering, Tampere University of Technology, Tampere, Finland.

Funding statement. This work is supported by the Jenny and Antti Wihuri Foundation.

Appendix A. The Bhattacharyya distance

The Bhattacharyya distance calculates the similarity of classes and has been used as a measure for feature selection in pattern recognition to separate classes from each other [32]. For two multivariate normal distributions, the Bhattacharyya distance is defined as follows [34]

$$b = \frac{1}{8}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) + \frac{1}{2} \ln \left(\frac{\det \boldsymbol{\Sigma}}{\det \boldsymbol{\Sigma}_1 \det \boldsymbol{\Sigma}_2} \right), \quad (\text{A } 1)$$

where $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are the means and covariances of the i^{th} distribution and $\boldsymbol{\Sigma} = \frac{\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2}{2}$. The normality of the parameters was verified using the Lilliefors test [35].

References

1. Stock SR. 2008 Recent advances in X-ray microtomography applied to materials. *Int. Mater. Rev.* **53**, 129–181. (doi:10.1179/174328008X277803)
2. Schladitz K. 2011 Quantitative micro-CT. *J. Microsc.* **243**, 111–117. (doi:10.1111/j.1365-2818.2011.03513.x)
3. Renard P, Allard D. 2013 Connectivity metrics for subsurface flow and transport. *Adv. Water Resour.* **51**, 168–196. (doi:10.1016/j.advwatres.2011.12.001)
4. Vogel HJ. 1997 Morphological determination of pore connectivity as a function of pore size using serial sections. *Eur. J. Soil Sci.* **48**, 365–377. (doi:10.1046/j.1365-2389.1997.00096.x)
5. Hovadik JM, Larue DK. 2007 Static characterizations of reservoirs: refining the concepts of connectivity and continuity. *Pet. Geosci.* **13**, 195–211. (doi:10.1144/1354-079305-697)
6. Allard D. 1993 On the connectivity of two random set models: the truncated Gaussian and the boolean. In *Geostatistics Tróia '92*, volume 5 of *Quantitative Geology and Geostatistics*, pp. 467–478. Springer, The Netherlands.
7. Knudby C, Carrera J. 2005 On the relationship between indicators of geostatistical, flow and transport connectivity. *Adv. Water Resour.* **28**, 405–421. (doi:10.1016/j.advwatres.2004.09.001)
8. Aydogan DB, Hyttinen J. 2013 Contour tree connectivity of binary images from algebraic graph theory. In *20th IEEE Int. Conf. on Image Processing (ICIP)*, pp. 3054–3058. 15–18 Sept. 2013, Melbourne, Australia.
9. Carr H, Snoeyink J, Axen U. 2003 Computing contour trees in all dimensions. *Comput. Geom. Theory Appl.* **24**, 75–94. (doi:10.1016/S0925-7772(02)00093-7)
10. Pascucci V, Cole-McLaughlin K. 2002 Efficient computation of the topology of level sets. In *IEEE Visualization, 2002. VIS 2002*, pp. 187–194. 27 Oct.–1 Nov. 2002, Boston, MA, USA.
11. Weber GH, Dillard SE, Carr H, Pascucci V, Hamann B. 2007 Topology-controlled volume rendering. *IEEE Trans. Visual Comput. Graph.* **13**, 330–341. (doi:10.1109/TVCG.2007.47)
12. Carr H, Snoeyink J, Van De Panne M. 2010 Flexible isosurfaces: simplifying and displaying scalar topology using the contour tree. *Comput. Geom. Theory Appl.* **43**, 42–58. (doi:10.1016/j.comgeo.2006.05.009)
13. Oesterling P, Heine C, Janicke H, Scheuermann G. 2010 Visual analysis of high dimensional point clouds using topological landscapes. In *2010 IEEE*

- Pacific Visualization Symposium (PacificVis)*, pp. 113–120. 2–5 Mar. 2010, Taipei, Taiwan.
14. Duke D, Carr H, Knoll A, Schunck N, Nam HA, Staszczak A. 2012 Visualizing nuclear scission through a multifield extension of topological analysis. *IEEE Trans. Visual Comput. Graph.* **18**, 2033–2040. (doi:10.1109/TVCG.2012.287)
 15. Aydogan DB, Hyttinen J. 2012 Binary image representation by contour trees. In *Proc. SPIE 8214, Medical Imaging 2012: Image Processing*, 83142X. 23 Feb. 2012, San Diego, CA, USA.
 16. Aydogan DB, Moritz N, Aro HT, Hyttinen J. 2013 Analysis of trabecular bone microstructure using contour tree connectivity. In *Medical image computing and computer-assisted intervention MICCAI 2013* (eds K Mori, I Sakuma, Y Sato, C Barillot, N Navab), volume 8150 of *Lecture Notes in Computer Science*, pp. 428–435. Berlin, Germany: Springer.
 17. Akolkar A, Petrasch J. 2012 Tomography-based characterization and optimization of fluid flow through porous media. *Transp. Porous Media* **95**, 535–550. (doi:10.1007/s11242-012-0060-7)
 18. Legland D, Kiu K, Devaux MF. 2007 Computation of Minkowski measures on 2D and 3D binary images. *Image Anal. Stereol.* **26**, 83–92. (doi:10.5566/ias.v26.p83-92)
 19. Scholz C, Wirner F, Götz J, Rüde U, Schröder-Turk G, Mecke K, Bechinger C. 2012 Permeability of porous materials determined from the Euler characteristic. *Phys. Rev. Lett.* **109**. (doi:10.1103/PhysRevLett.109.264504)
 20. Vogel HJ, Weller U, Schlter S. 2010 Quantification of soil structure based on Minkowski functions. *Comput. Geosci.* **36**, 1236–1245. (doi:10.1016/j.cageo.2010.03.007)
 21. Petrasch J, Meier F, Friess H, Steinfeld A. 2008 Tomography based determination of permeability, Dupuit–Forchheimer coefficient, and interfacial heat transfer coefficient in reticulate porous ceramics. *Int. J. Heat Fluid Flow* **29**, 315–326. (doi:10.1016/j.ijheatfluidflow.2007.09.001)
 22. Patankar S, Spalding D. 1972 A calculation procedure for heat, mass and momentum transfer in three-dimensional parabolic flows. *Int. J. Heat Mass Trans.* **15**, 1787–1806. (doi:10.1016/0017-9310(72)90054-3)
 23. Linstrom PJ, Mallard WG. 2013 NIST chemistry webbook. NIST standard reference database number 69. National Institute of Standards and Technology, Gaithersburg, MD 20899, USA. <http://webbook.nist.gov> (accessed 10 October 2013).
 24. 2013 enGrid—open source mesh generation. enGits GmbH Haldenweide 12, 79674 Todtnau, Germany. <http://engits.eu/en/engrid> (retrieved 26 April 2013).
 25. Chiu SN, Stoyan D, Kendall WS, Mecke J. 2013 *Stochastic geometry and its applications*, 3rd edn. Wiley series in probability and mathematical statistics. London, UK: Wiley.
 26. Jeulin D. 2005 Random structures in physics. In *Space, Structure and Randomness*, vol. 183 of *Lecture Notes in Statistics*, pp. 183–219. New York, NY: Springer.
 27. Andersson L, Jones AC, Knackstedt MA, Bergström L. 2011 Permeability, pore connectivity and critical pore throat control of expandable polymeric sphere templated macroporous alumina. *Acta Mater.* **59**, 1239–1248. (doi:10.1016/j.actamat.2010.10.056)
 28. Vogel HJ. 2000 A numerical experiment on pore size, pore connectivity, water retention, permeability, and solute transport using network models. *Eur. J. Soil Sci.* **51**, 99–105. (doi:10.1046/j.1365-2389.2000.00275.x)
 29. Samouëlian A, Vogel H-J, Ippisch O. 2007 Upscaling hydraulic conductivity based on the topology of the sub-scale structure. *Adv. Water Resour.* **30**, 1179–1189. (doi:10.1016/j.advwatres.2006.10.011)
 30. Comaniciu D, Ramesh V, Meer P. 2003 Kernel-based object tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**, 564–577. (doi:10.1109/TPAMI.2003.1195991)
 31. Reyes-Aldasoro CC, Bhalerao A. 2006 The Bhattacharyya space for feature selection and its application to texture segmentation. *Pattern Recognit.* **39**, 812–826. (doi:10.1016/j.patcog.2005.12.003)
 32. Fukunaga K. 1990 *Introduction to statistical pattern recognition*. San Diego, CA: Academic Press.
 33. Duda RO, Hart PE, Stork DG. 2000 *Pattern classification*, 2 edn. London, UK: Wiley-Interscience.
 34. Bhattacharyya A. 1943 On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc.* **35**, 99–109.
 35. Lilliefors HW. 1967 On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *J. Am. Stat. Assoc.* **62**, 399–402. (doi:10.1080/01621459.1967.10482916)