



Water Resources Research

RESEARCH ARTICLE

10.1002/2015WR017937

Special Section:

Applications of percolation theory to porous media

Key Points:

- Topological persistence is a powerful and practical tool for characterizing rock microstructure
- Persistence data from sandstones contain a strong signal of the critical percolating sphere radius
- Limits on the connection are illustrated by a heterogeneous limestone sample

Correspondence to:

V. Robins,
vanessa.robins@anu.edu.au

Citation:

Robins, V., M. Saadatfar, O. Delgado-Friedrichs, and A. P. Sheppard (2016), Percolating length scales from topological persistence analysis of micro-CT images of porous materials, *Water Resour. Res.*, 52, 315–329, doi:10.1002/2015WR017937.

Received 3 AUG 2015

Accepted 22 DEC 2015

Accepted article online 28 DEC 2015

Published online 17 JAN 2016

Percolating length scales from topological persistence analysis of micro-CT images of porous materials

Vanessa Robins¹, Mohammad Saadatfar¹, Olaf Delgado-Friedrichs¹, and Adrian P. Sheppard¹
¹Department of Applied Mathematics, Research School of Physics and Engineering, Australian National University, Canberra, ACT, Australia

Abstract Topological persistence is a powerful and general technique for characterizing the geometry and topology of data. Its theoretical foundations are over 15 years old and efficient computational algorithms are now available for the analysis of large digital images. We explain here how quantities derived from topological persistence relate to other measurements on porous materials such as grain and pore-size distributions, connectivity numbers, and the critical radius of a percolating sphere. The connections between percolation and topological persistence are explored in detail using data obtained from micro-CT images of spherical bead packings, unconsolidated sand packing, a variety of sandstones, and a limestone. We demonstrate how persistence information can be used to estimate the percolating sphere radius and to characterize the connectivity of the percolating cluster.

1. Introduction

It has been long understood that the connectivity of different phases in a porous material (rock and water in an aquifer, for example) strongly influences physical properties of the system, particularly those related to transport (fluid flow, diffusion of contaminants, or mechanical force chains, for example). This relationship has been studied primarily through percolation theory, i.e., the study of cluster formation and growth under the addition of randomly distributed components [Broadbent and Hammersley, 1957].

Percolation theory has been applied extensively to the study of transport in porous materials, Sahimi [1993], with various adaptations to specific problems. Percolation length scales are correlated to the permeability of random porous media in Katz and Thompson [1986]. Permeability, k , is a geometrical property of the material relating to the rate of fluid flow which is controlled by the narrow constrictions in the pore phase. Permeability has the dimensions of area and may be thought of as representing the cross section of an effective channel for fluid flow through the pore space. Katz and Thompson showed that for sandstone and carbonate rocks, k is correlated to ℓ_c , a critical pore diameter corresponding to the diameter of the largest sphere that can percolate through the sample pore space:

$$k = c \ell_c^2 \sigma / \sigma_0, \quad (1)$$

where c is a constant, σ is the electrical conductivity of the brine-saturated sample, and σ_0 is the conductivity of bulk brine. This method for estimating the permeability has a significant merit as ℓ_c can also be experimentally measured from mercury intrusion or directly estimated from 3-D digital images (tomograms).

In this paper, we work with a simple geometric definition of percolation. The sample, X , is a bounded subset of \mathbb{R}^3 with two distinct phases, pore and grain. We define r_{pore} as the radius of the largest sphere that can move freely from one side of the pore space to the other and r_{grain} as the radius of the largest sphere that can move freely through the grain phase. This percolating sphere problem is equivalent to testing the connectivity of level sets of a real-valued function $f(x) : X \rightarrow \mathbb{R}$ defined by the shortest distance from each point in X to the pore-grain interface. By convention, we take distances to be negative in the pore and positive in the grain phase, and refer to f as the signed Euclidean distance transform (SED). The lower level sets of f are $L_f(c) = \{x \mid f(x) \leq c\}$, so that r_{pore} is the smallest value of c for which $L_f(c)$ spans the domain. The upper level sets are $U_f(c) = \{x \mid f(x) \geq c\}$ and r_{grain} is the largest value of c for which $U_f(c)$ spans. In a probabilistic setting where there are many realizations of the same random process, r_{pore} and r_{grain} are defined in a limit, so that with probability 1, we expect $L_f(c)$ will span if $c > r_{\text{pore}}$, and with probability 1 we expect $L_f(c)$

will not span if $c < r_{\text{pore}}$. Our data are obtained from X-ray micro-CT images of real porous and granular materials, and we do not attempt a full statistical analysis of their percolation phenomena, but work with the simple geometric definitions given above.

Topological quantities are attracting increasing attention in porous media studies, since connectivity plays such a dominant role in fluid transport and fluid trapping [Scholz *et al.*, 2012; Herring *et al.*, 2013; Wildenschild and Sheppard, 2013]. These studies have focused on computing topological measures such as the Euler characteristic and the Betti numbers of the pore space and relating them to transport. The Betti numbers, β_i , count the number of distinct i -dimensional cycles, i.e., components β_0 , loops β_1 , and enclosed voids β_2 , in a three-dimensional object. The Euler characteristic is the alternating sum of Betti numbers: $\chi = \beta_0 - \beta_1 + \beta_2$ for 3-D objects. The surface of a donut, for example, has $\beta_0 = 1$, $\beta_1 = 2$, and $\beta_2 = 1$. A connected network with V vertices and E edges has $\beta_0 = 1$ and $\beta_1 = E - V + 1$. Persistent homology is a mathematical theory that allows us to study topological quantities as an object grows [Verri *et al.*, 1993; Robins, 1999; Edelsbrunner *et al.*, 2002]. It has rapidly become an important tool for studying shape in application areas from digital images [Bendich *et al.*, 2010; Delgado-Friedrichs *et al.*, 2015] to dynamical systems Kurtuldu *et al.* [2011] to high-dimensional data-mining [Carlsson, 2009]. Rather than just computing the Betti numbers of an object, persistent homology allows us to track all the topological changes in the lower level sets of a real-valued function, $L_f(c)$, and presents this information in a set of *persistence diagrams*. The persistence diagrams give much more information than the Betti numbers as a function of c ; they allow us to see directly the range of c values over which each topological feature (component, loop, or void) exists. By applying persistence analysis to the SEDT of a segmented tomogram, we are able to capture a wealth of geometric information while also improving the robustness of topological measures. The purpose of this paper is to introduce topological persistence diagrams as a tool for characterizing structure in porous materials and to establish the connections between persistence and percolation.

This paper begins in section 2 by giving an intuitive description of persistence diagrams and their connection to geometric and topological properties of porous materials, using a random packing of spherical beads as an illustrative example. In section 3, we use a voxelized version of the “Swiss cheese model” to illustrate and elaborate on the connections between persistence and percolation. The key observation is that near the percolating threshold, the persistence algorithm can pair birth and death events that are spatially separated. This gives us a basis for estimating percolating radii from the persistence information. The following section 4 presents data from micro-CT images of sandstones that show qualitatively similar behavior to the Swiss cheese model, and other materials that do not.

2. Persistence Diagrams of Porous Materials

We describe how topological persistence diagrams are computed from micro-CT images and how to interpret this information as geometric properties of porous materials, using a packing of spherical acrylic beads as an illustrative example (Figure 1).

The micro-CT images used in this paper were obtained using machines built within the Department of Applied Mathematics at the ANU in Canberra. The hardware and software used in our facility is described in Sheppard *et al.* [2014]. The current paper examines simple porous materials consisting of just two phases so the X-ray density image is *segmented* to assign each voxel to either the grain or pore phase. Geometric characterization of this binary segmented image is achieved using the *signed Euclidean distance transform* (SEDT) defined by the shortest distance from a voxel to the interface between the two phases. A voxel that is face-adjacent to one of the opposite phase has a distance of 0.5 units. In this paper, we take negative distances within the pore phase and positive in the grain phase. The result is a function $f(x, y, z)$ defined on a regular rectangular grid of points $(x, y, z) \in \mathbb{Z}^3$ with the property that all local minima of f lie inside the pore phase and all local maxima in the grain phase.

We extend the SEDT f from a function defined on a grid of points to a discrete Morse function defined on a cubical cell complex as described in Robins *et al.* [2011]. This use of cubical cell complexes and discrete Morse functions allows us to correctly determine locations where the topology of level sets of f changes, and to build a combinatorial complex that captures this topology. Recall that the lower level sets of a function f are $L_f(c) = \{(x, y, z) \text{ such that } f(x, y, z) \leq c\}$. It is a central result in Morse theory that the topology of these lower level sets can only change as c passes through a critical value of f .

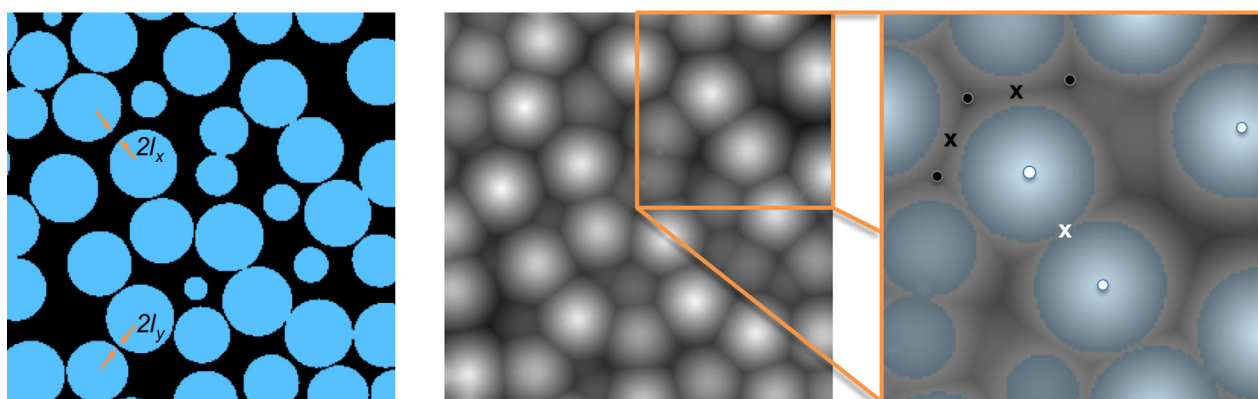


Figure 1. A single slice through a 3-D image of a packing of acrylic beads of diameter 1.59 mm. (left) Segmented image showing the critical percolating lengths for 2-D traversal in the x axis and y axis directions. (middle) Signed Euclidean distance transform image. (right) Close up of SEDT showing some local maxima (white dots), 2-saddles (white x), local minima (black dots), and 1-saddles (black x). The critical points are marked for illustrative purposes only, the true critical points do not all occur in the same slice.

The critical points are of four basic types: local minima (index-0), local maxima (index-3), index-1, and index-2 saddle points (the index of a critical point is the number of decreasing dimensions of the function in its neighborhood). The combinatorial cell complex consists of a 0-cell (vertex) for each local minimum, a 1-cell (edge) for each index-1 saddle, a 2-cell for each index-2-saddle, and a 3-cell for each local maximum. The 2-cells and 3-cells are patches and solids that are topologically simply connected like a disk and a solid ball, but can have geometrically irregular shape. The face-adjacencies between the cells in the complex are determined by gradient flow lines that start in the neighborhood of the index- k critical point and descend to an index- $(k - 1)$ critical point. The cells are assigned weights that are the critical values of their critical points. The weighted graph of vertices and edges from this cell complex is effectively a pore space network model as is commonly used in the study of fluid displacements in porous materials [Fatt, 1956; Bakke and Oren, 1997; Blunt, 2001; Prodanovic et al., 2007]. We call the subset of the cell complex that lies inside the pore space the *Morse skeleton*. This structure is analogous to a medial axis but removes the geometric noise and topological ambiguity associated with medial axes by applying a simplification process dictated by topological persistence [see Delgado-Friedrichs et al., 2015, for details].

In the example sphere packing image, local maxima of f occur at the center of each bead, index-2 saddle points at contacts between beads (and other locations), while index-1 saddle points typically define constrictions in the pore space ("throats") between "pores" that are detected as local minima of f . If we start with low values of c and increase with small steps, then each time we pass a local minimum a new component of $L_f(c)$ is created and a vertex is added to the combinatorial complex, \mathcal{C} . Passing an index-1 saddle means either two components of $L_f(c)$ merge into one, or a new loop is made; in either case an edge is added to \mathcal{C} . When passing through a 2-saddle, either a loop (one-dimensional hole) in $L_f(c)$ is filled in, or a void may be enclosed (e.g., a grain is isolated); in \mathcal{C} we add a 2-cell. A void in $L_f(c)$ is finally filled in by a local maximum and this is represented by the addition of a 3-cell to \mathcal{C} .

The above sequence of topological changes is called a *filtration* and the idea behind persistent homology is that each death event (i.e., merging two components, filling in a loop or a void) can be paired uniquely with an earlier birth event (the creation of a new component, 1-cycle or 2-cycle) [Edelsbrunner et al., 2002]. The (birth, death) pairs capture all topological changes in the lower level sets of a Morse function f . Note that a Morse function defined on an arbitrary manifold may have birth events that are never paired; these are called essential cycles. For an SEDT on a rectangular domain, there is only one essential cycle defined by its absolute minimum, the largest pore for the examples in this paper.

The homology of a filtration is conveniently summarized in the *persistence diagrams* (PD*i*) for each dimension i : plots of the critical values (b , d) for each (birth, death) pair. PD0 represents the births and deaths of connected components, PD1 the births and deaths of 1-cycles (loops), and PD2 the births and deaths of 2-cycles (voids). In any diagram, points near the diagonal have $d \approx b$, and these features are said to have low persistence. A small perturbation of the function can be made to remove the critical points that correspond to the paired birth and death event and a key result in persistent homology is that persistence diagrams are

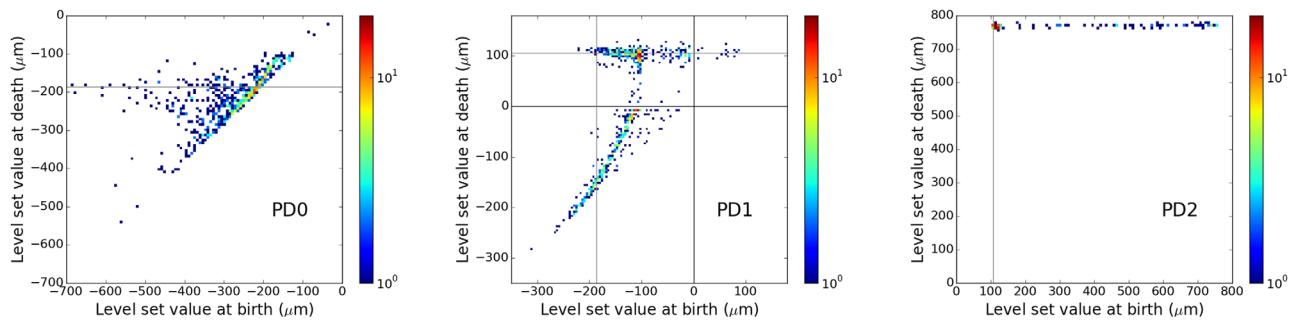


Figure 2. Persistence diagrams for a 512^3 voxel subset of the distance transform for a bead packing. The (b, d) points are binned in a 2-D histogram with logarithmic colorbar scale shown in the right (1 = dark blue, 20 = dark red). The gray horizontal and vertical lines show the relevant critical percolating length scales, $r_{\text{pore}} = 186.6 \mu\text{m}$ and $r_{\text{grain}} = 105.7 \mu\text{m}$. The hot spots and linear structures in PD1 and PD2 are due to the single size of sphere in this packing.

stable with respect to small changes in the parent function [Cohen-Steiner *et al.*, 2007]. The stability of the persistence diagrams is in stark contrast to the Betti numbers, which are very sensitive to small changes in function values, and is one of the main reasons a theory of persistent homology was developed.

Persistence diagrams derived from the SEDT of the sphere-packing image are shown in Figure 2. These illustrate some of the main features of PDs from simple porous materials, but the single-sized spherical beads mean the PDs display extra structure. A point (b, d) in PD0 means a connected component of the pore space is born at function-value b and dies at d . That is, there is a local minimum of f with value b and the component it lies inside joins with another component at an index-1 saddle point when the threshold reaches d . Birth values in PD0 therefore define a pore-size distribution equivalent to that derived from considering maximal inscribed sphere radii. The death values with $d < 0$ are throat radii for those throats in an optimal spanning tree of a medial-axis-style network model. Any points with $d > 0$ signal a disconnected pore space component.

A point (b, d) in PD1 means a 1-cycle (a loop) is created in $L_f(b)$ and filled in at $L_f(d)$. When $b < d < 0$ the 1-cycle is created and filled in the pore phase, so these pairs signal the existence of highly nonconvex pores that are not well modeled by simple spherical pores connected by cylindrical throats. In this case, a small difference $d - b$ means the pore is fairly flat like a pizza-base, and a larger difference suggests a more toroidal pore shape (see Figure 3, for a sketch). The persistence in this context is therefore a type of aspect-ratio for pores that are not well modeled by a sphere. In Figure 2, we see that such pores are in fact quite common in the bead packing and are due to close but not touching bead configurations present in the disordered packing of beads. When $b < 0 < d$ we have the typical situation of a 1-cycle that is a redundant pathway through a network model of the pore space, with b telling us the narrowest throat radius along this cycle, and d measuring a corresponding constriction in the grain phase (i.e., a grain-contact radius). If $0 < b < d$, the 1-cycle is born in the grain phase and signals a highly nonconvex structure in the grain phase.

Finally, a point (b, d) in PD2 means a 2-cycle is created in $L_f(c)$ when c passes through b and then filled in at a local maximum with value d . The value of the local maximum is the radius of the maximally inscribed sphere within a grain, and the value of the 2-saddle defines a grain-contact radius (analogous to a pore-throat radius). Birth values in PD2 with $b < 0$ signal a disconnected grain phase, unlikely in the interior of the image of a real porous material unless there has been a poor segmentation. For the sphere packing, we see most of the PD2 points concentrated around $(b, d) = (113, 780) \mu\text{m}$ because all spheres have the same size (their diameter is 1.59 mm). The grain-contact radius of $113 \mu\text{m}$ is perhaps larger than expected for point-contacts between spheres, but is a consequence of the relatively poor resolution in this public-domain tomographic image. The PD2 points with $d \approx 780 \mu\text{m}$ and $b > 113 \mu\text{m}$ are due to partial beads cropped off at the edge of the image.

We recap the above discussion of persistence by noting that a persistence pair in dimension i consists of two critical points of the Morse function f : a birth at $f(x_b, y_b, z_b) = b$ and a death at $f(x_d, y_d, z_d) = d$. The most common way to represent this information is through the persistence diagrams, PD_i , of the critical values (b, d) . We will see below that the distribution of distances between the birth and death critical points, $\|x_d - x_b\|$, also shows a strong signal of the critical percolating length scales.

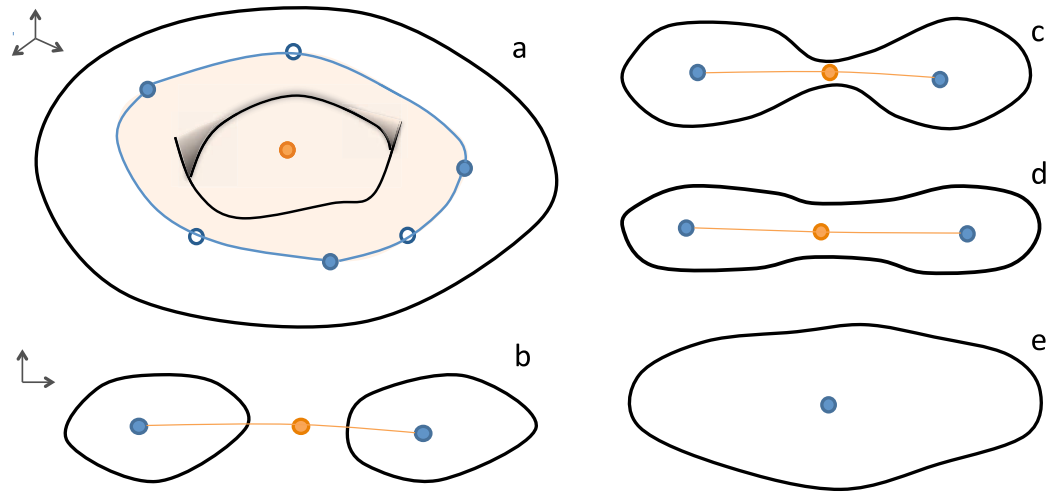


Figure 3. Diagrams to illustrate the connection between pore geometry and PD1 points. (a) This sketch shows a toroidal-shaped pore with the “typical” PD1 signature of $b < 0 < d$. The surface of torus is the interface between pore and grain. The pore contains three local minima (solid blue points, 0-cells in the combinatorial complex) and three 1-saddle points (open blue points on 1-cells of the cell complex) arranged in a loop. The orange point marks the location of a 2-saddle in the grain phase, it represents a contact between two “grains.” The corresponding 2-cell in the complex is colored pale orange. Suppose the critical values for the 1-saddles are $s_1 < s_2 < s_3 < 0$, and for the 2-saddle it is $t > 0$, then the PD1 point associated with this pore configuration is at (s_3, t) . (b) This sketch is a cross section through the toroidal pore above. The blue points and orange line now mark the intersection of the 1-cells and 2-cell with the sectioning plane. (c) Now imagine a pore where the grains are no longer in contact at the 2-saddle, but have become slightly separated. The 2-saddle is interior to the pore space and so $s_3 < t < 0$. (d) The grains are further apart in this configuration, so the persistence $(t - s_3)$ is smaller than for the pore in Figure 3c. (e) The 2-saddle is no longer present in the SEDT of this nearly convex pore.

3. Continuum Percolation and Persistence in the “Swiss Cheese” Model

The simplest mathematical model of percolating grains is the Swiss cheese (or Poisson-Boolean) model where points are generated in the unit cube with coordinates taken from the uniform random distribution on $[0,1]$ and spherical “grains” of radius r are centered on each point with no interaction, meaning they overlap freely. Values of the critical percolating volume fractions, $\phi_c = 1 - \exp(-\eta_c)$ have been computed to a high degree of accuracy by others [Lorenz and Ziff, 2001; Rintoul, 2000]:

$$\eta_{\text{grain}} = \frac{4}{3} \pi N r_{\text{grain}}^3 = 0.341889, \quad (2)$$

$$\eta_{\text{pore}} = \frac{4}{3} \pi N r_{\text{pore}}^3 = 3.5032 \text{ in the limit } N \rightarrow \infty. \quad (3)$$

We create a voxel image based on a Swiss cheese model by first generating 10^5 points with uniform random coordinates in the unit cube and then mapping them into a 1000 cubed voxel domain. These (mostly isolated) voxels are used as the grain phase seeds for an SEDT as described earlier. The percolating thresholds for the upper and lower level sets of this SEDT are $r_{\text{grain}} = -8.62$ and $r_{\text{pore}} = -19.22$ voxel units, respectively. In Figure 4, we show the persistence diagrams for the distance function that is positive in the grain and negative in the remainder of the domain, as is the case for the other data in this paper. The grains in this case are single voxels with finite width, so we see all the PD2 points have $d = 0.5$ voxel units. PD2 has a distribution of birth values, as expected for this purely random overlapping sphere model and in contrast to the bead packing described in the previous section. The PD0 births show a broad distribution of pore sizes, and a horizontal spike at a d value that coincides with the critical percolating radius for the pore phase. PD1 has a roughly triangular shape with tapering sides and we see that most (86%) of the (b, d) pairs lie within the region bounded by $b > r_{\text{pore}}$, $d < r_{\text{grain}}$, and $b < d$.

In Figure 5, we show distributions of the distances between the critical points associated with each (birth, death) persistence pair. These figures show very clearly that most of the creator and destroyer critical points are within 100 voxel units (5 times the mean nearest neighbor distance, or 0.1 of the unit cube edge), but that a small fraction of pairs become highly spatially separated (up to the domain width) at the critical percolating length scales. The highly separated pairs occur in a very narrow range of PD0 death values

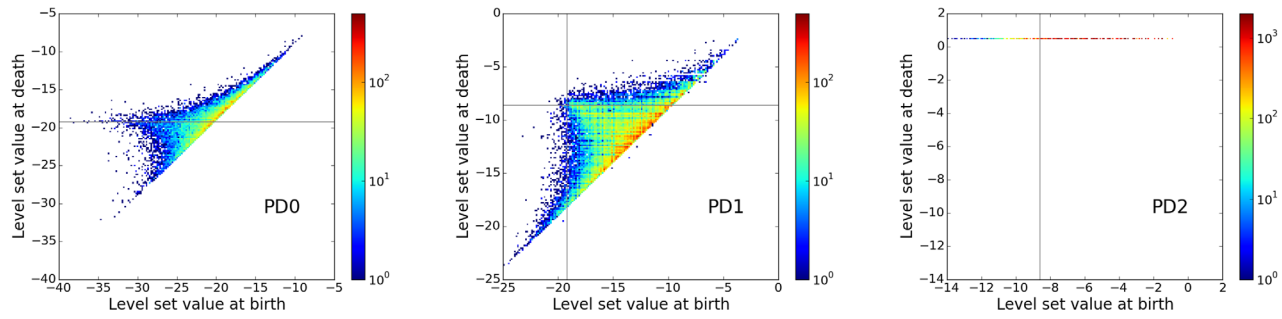


Figure 4. Persistence diagrams PD0, PD1, PD2 from the SEDT of 10^5 random point grains in a 1000^3 domain. The gray horizontal and vertical lines show critical percolating length scales, $r_{pore} = -19.22$ and $r_{grain} = -8.62$ voxel units. The value of r_{grain} here is the highest value of c for which the upper level sets of the SEDT span the domain.

centered on r_{pore} , the separation is maximal for PD1 birth values at r_{pore} and for PD1 death values at r_{grain} . There is a 4% difference between the birth value of the maximally separated PD2 pair and r_{grain} , and the peak in the distances is again narrowly defined.

We describe the persistence pairing algorithm in more detail in Appendix A, and make detailed connections to the critical percolating length scales. The key point is that the existence of a spanning cluster at the percolating length scale means the birth, death critical point pairs in PD0 (and PD1) can become separated. Also, we note that above the pore space percolation threshold most edges added to the filtration will create 1-cycles and we see that jump clearly in the PD1 diagram. By complementarity and duality, similar arguments apply to PD2 and the grain phase percolation threshold.

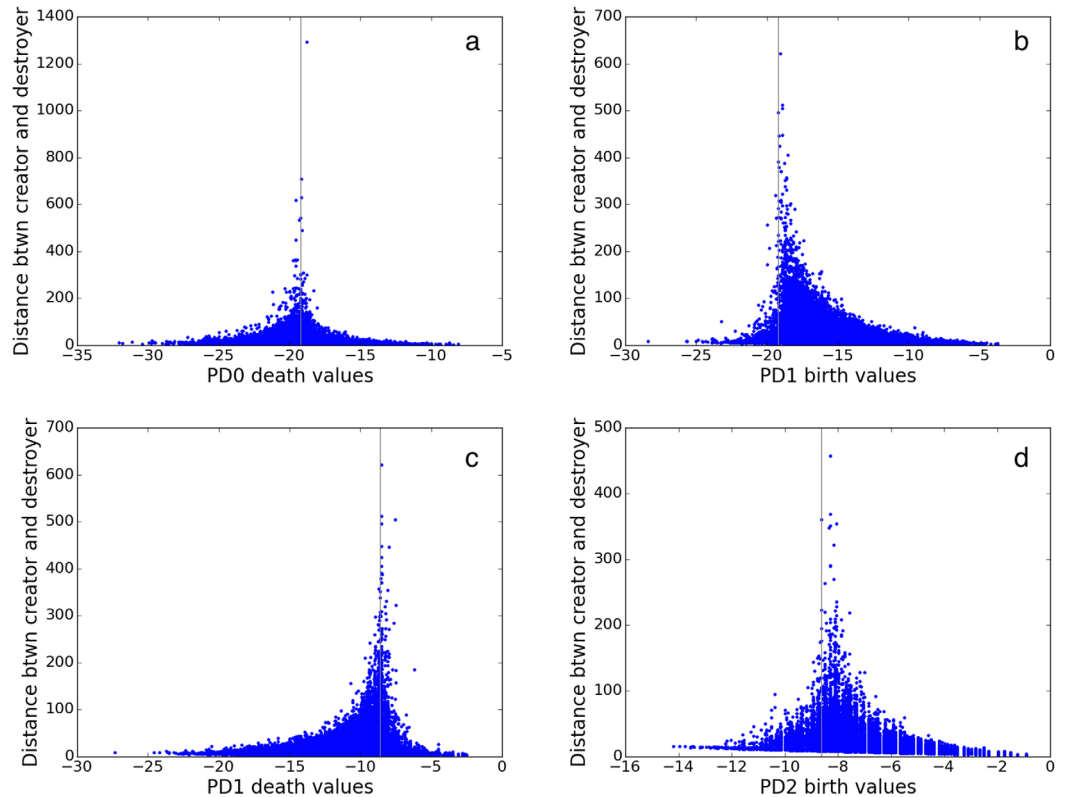


Figure 5. Distances between creator and destroyer persistence pairs for the SEDT of 10^5 random point grains. (a) PD0 deaths versus separation. (b) PD1 births versus separation. (c) PD1 deaths versus separation. (d) PD2 births versus separation. The gray lines mark the relevant critical percolating length scales, $r_{pore} = -19.22$ and $r_{grain} = -8.62$ voxel units.

Table 1. Percolation Length Scales and Estimates From Maximal Persistence ($d - b$) and Maximal Separation $|w - v|$ Persistence Pairs With Birth $f(v) = b$ and Death $f(w) = d^a$

Sample Porosity	Image Size Voxel Res. (μm)	r_{pore}	r_{grain}	PD0 max($d - b$)	PD1 max($d - b$)	PD2 max($d - b$)
				PD0 max $ w - v $	PD1 max $ w - v $	PD2 max $ w - v $
Swiss cheese	1000 × 1000 × 1000	−19.22	−8.62	−18.81	−20.78, −8.27	−14.20
	1			−18.81	−19.04, −8.50	−8.27
RCP spheres	512 × 512 × 512	−186.6	105.7	−182.7	−8.7, 495.5	103.1
37.9%	17.5			−184.2	−174.5, 116.0	114.8
Sand pack	512 × 512 × 512	−41.3	29.7	−41.3	−25.9, 256.8	33.3
36.3%	9.2			−41.3	−40.4, 32.1	59.0
Castlegate	512 × 512 × 512	−10.9	43.7	−6.8	−2.8, 99.4	46.0
20.6%	5.6			−13.0	−9.7, 42.3	54.9
Bentheimer	1283 × 1283 × 1792	−13.5	41.1	−10.6	−5.2, 68.7	44.9
21.2% (Ben2,1)	4.26			−12.0	−13.2, 39.8	39.8
Bentheimer	1287 × 1287 × 1808	−16.5	42.3	−14.9	2.1, 120.5	46.7
23.0% (Ben2,2)	4.27			−14.9	−14.9, 40.8	41.6
Bentheimer	1131 × 1131 × 1760	−14.2	56.4	−11.1	−5.1, 77.6	56.1
18.9% (BTH)	2.96			−13.6	−11.1, 51.3	56.1
Berea	1039 × 1039 × 1600	−7.8	24.3	−6.2	−5.1, 35.7	25.8
19.6%	1.9			−7.8	−5.4, 23.8	25.5
Clashach1	1273 × 1273 × 1860	−13.7	37.2	−11.9	26.1, 91.7	36.5
18.9%	2.64			−13.8	−13.8, 34.3	36.4
Clashach2	1235 × 1235 × 1856	−12.2	38.8	−4.7	9.5, 85.5	40.2
17.6%	2.72			−4.7	−9.9, 36.2	40.3
Mt Gambier	512 × 512 × 512	−31.2	39.3	−28.3	−5.2, 106.0	40.6
43.6%	3.0			−41.3	−4.5, 25.7	29.9

^aFor PD0, the death value of these maximal pairs is an estimate of r_c for the pore space. For PD1, the birth value estimates r_c pore and the death value r_c for the grain phase. For PD2, the birth value is also an estimate for r_c grain. Estimates that are within $\pm 10\%$ of the computed percolating radii are highlighted in bold.

4. Persistence and Percolation for Micro-CT Imaged Samples

In this section, we compare persistence diagrams and critical percolating length scales for a variety of examples. Results from micro-CT images of an unconsolidated sand pack and several sandstones show qualitatively similar features to the Swiss cheese model. We also give results from a micro-CT image of a limestone as an example where the connection between persistence and percolating length scales is less straightforward.

For all the samples, we estimate percolating length scales from the persistence information in a number of ways. Our data for the Swiss cheese model and arguments in Appendix A suggest the following candidates for r_{pore} :

1. PD0 death value of the pair with maximal persistence, $d - b$.
2. PD0 death value of the pair with maximal separation, $|\mathbf{x}_d - \mathbf{x}_b|$.

and for r_{grain} :

3. PD2 birth value of the pair with maximal persistence, $d - b$.
4. PD2 birth value of the pair with maximal separation, $|\mathbf{x}_d - \mathbf{x}_b|$.

We also test the relationships between birth and death values of maximally persistent and separated pairs in PD1 and r_{pore} and r_{grain} .

Values of critical percolating radii and the above quantities for all samples are given in Table 1. The death values of maximally persistent 0-cycles are within 10% of r_{pore} for five out of 11 samples. Death values of maximally separated 0-cycles are a better estimate with seven out of 11 samples within 10% of r_{pore} . In almost all samples, the death values of maximally persistent and separated 0-cycles are larger (less negative) than the critical percolating value of r_{pore} . This is to be expected since the pairing of vertices and edges that are separated in value and/or space is most likely to occur when two large components $L_i(c)$ are merged. We also know that the largest pore (most negative local minimum) is never paired. The appendix explains this in more detail.

The birth values of maximally persistent 2-cycles are within 10% of r_{grain} in nine out of 11 samples. Birth values of maximally separated 2-cycles are within 10% of r_{grain} in 8 out of 11 data sets. The PD2 estimates of percolation do not show the same systematic error as observed for PD0. One reason for this is because there is no unpaired local maximum.

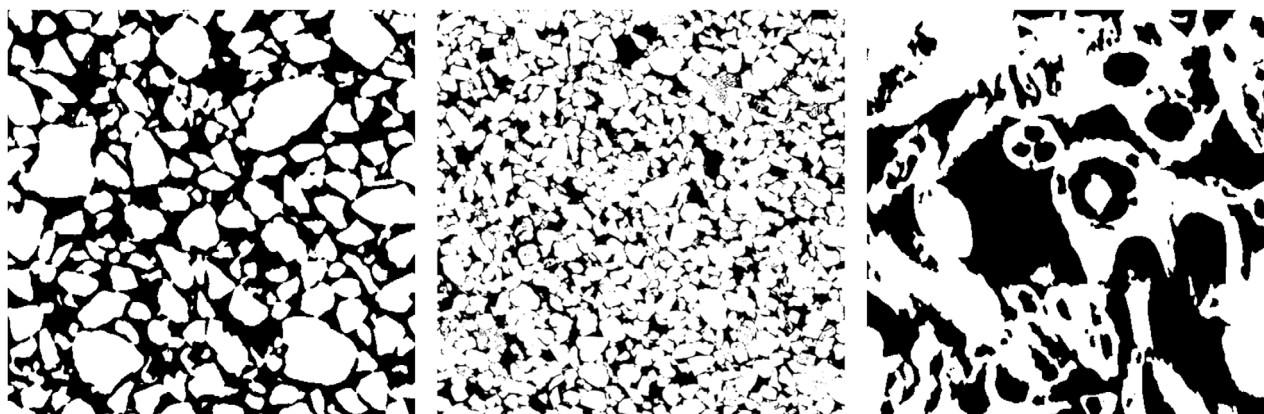


Figure 6. Slices through the segmented volume images of some porous rocks. (left) The unconsolidated fluvial sand pack, image dimension is 4.71 mm (512 voxels) square. (middle) A Bentheimer sandstone, image dimension is 5.46 mm (1283 voxels) square. (right) Mt Gambier limestone, image dimension is 1.54 mm (512 voxels) square.

Birth and death values of maximally persistent 1-cycles only match the percolating length scales in the Swiss cheese model. Six samples have their maximally separated 1-cycles within 10% of both r_{pore} and r_{grain} and an additional four samples (10 in total) have the death value of the maximally separated 1-cycles within 10% of r_{grain} .

The data presented in Table 1 strongly support our argument that there is a relationship between percolation thresholds and persistence data but the maximally persistent or maximally separated pairs are not the best signifiers, due to their sensitivity. Further numerical and probabilistic studies are needed to establish a more accurate relationship.

The rest of this section discusses the different samples in detail.

4.1. Unconsolidated Fluvial Sand Pack

Our first example is a somewhat poorly sorted unconsolidated fluvial sand pack from southern Australia. The segmented image has a porosity of 36.3%, with dimensions $512 \times 512 \times 512$ voxels and a voxel size of $9.184 \mu\text{m}$ (see Figure 6a). The persistence diagrams are displayed in Figure 7. As expected, we see a broad range of pore and grain sizes in the PD0 birth and PD2 death values, respectively. The percolation length scale for the grain phase matches the narrow spike in the PD2 diagram, with the most persistent 2-cycle having a birth value very close to the independently computed percolating sphere radius (33.3 versus $29.7 \mu\text{m}$). The PD0 diagram shows a dual behavior to PD2, with the death value of the most persistent 0-cycle matching the pore space percolating radius of $41.3 \mu\text{m}$.

Looking at PD1 of the sand pack, we see a significant number, 30%, of (b, d) pairs with $b < d < 0$ signifying the existence of highly nonconvex pores, and only a small fraction, 2.2% of (b, d) points with $0 < b < d$, confirming that the grains are approximately convex. Similar to the Swiss cheese model and in contrast to the sphere packing, the PD1 (b, d) pairs for the sand pack fill out a triangular region bounded by the percolating length scales and the $b = d$ diagonal. This densely filled, roughly triangular region in PD1 is also seen in the sandstones.

Plots of the distances between paired critical points are given in Figure 8. Again, there is extremely close agreement between the percolation thresholds and narrow spikes in the distributions. In Figure 8d, we see a number of outlier points that broaden the main peak, and have lead to a significant discrepancy between the maximally separated birth value and r_{grain} . These outlier points are probably due to the very wide distribution of grain sizes in the sand pack.

4.2. Consolidated Sandstones

We have computed persistence diagrams and critical percolating length scales for seven samples covering four different sandstone types: Bentheimer from Germany, Berea from Ohio, Castlegate from Utah, and Clashach from Scotland (see Figure 6b, for a representative cross section of one segmented image and Figure 9 for persistence diagrams of four samples). The persistence diagrams for these sandstones share broadly similar features with each other, and with the sand pack. The most significant difference between the

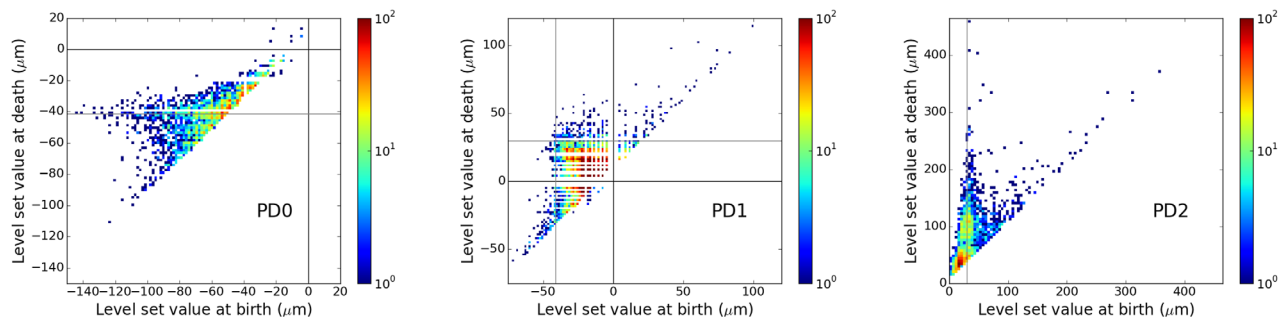


Figure 7. Persistence diagrams for a 512^3 voxel subset of the distance transform for an unconsolidated sand packing. The (b, d) points are binned in a 2-D histogram with logarithmic color bar scale shown in the right. The gray horizontal and vertical lines show the critical percolating length scales, $r_{pore} = -41.3 \mu\text{m}$ and $r_{grain} = 29.7 \mu\text{m}$.

unconsolidated sand packing and the consolidated sandstones is that the latter have a much smaller proportion of PD1 points with $b < d < 0$ (2.5% for the Castlegate sample, for example) and a much larger proportion with $0 < b < d$ (36% for Castlegate). Recall that the latter points signify highly nonconvex components in the grain phase. This is naturally explained for the sandstones by the sedimentation of the narrowest throats fusing three or more grains together, and is a significant marker of blocked flow pathways. The disappearance of the highly nonconvex pores may be explained by the compaction and sedimentation processes forcing the near-contacts in a sand pack that created the $b < d < 0$ PD1 points to become true contacts in the grain phase so that now $b < 0 < d$ in the sandstone.

The relationship between the percolating length scales and the distribution of PD1 points is similar in all the sandstones. The point (r_{pore}, r_{grain}) defines four regions in the PD1 diagram, with the upper left an empty zone where almost no PD1 points are found, the lower right containing the vast majority of PD1 points (94–98% for these samples), and the other two areas containing a marginal number of PD1 points.

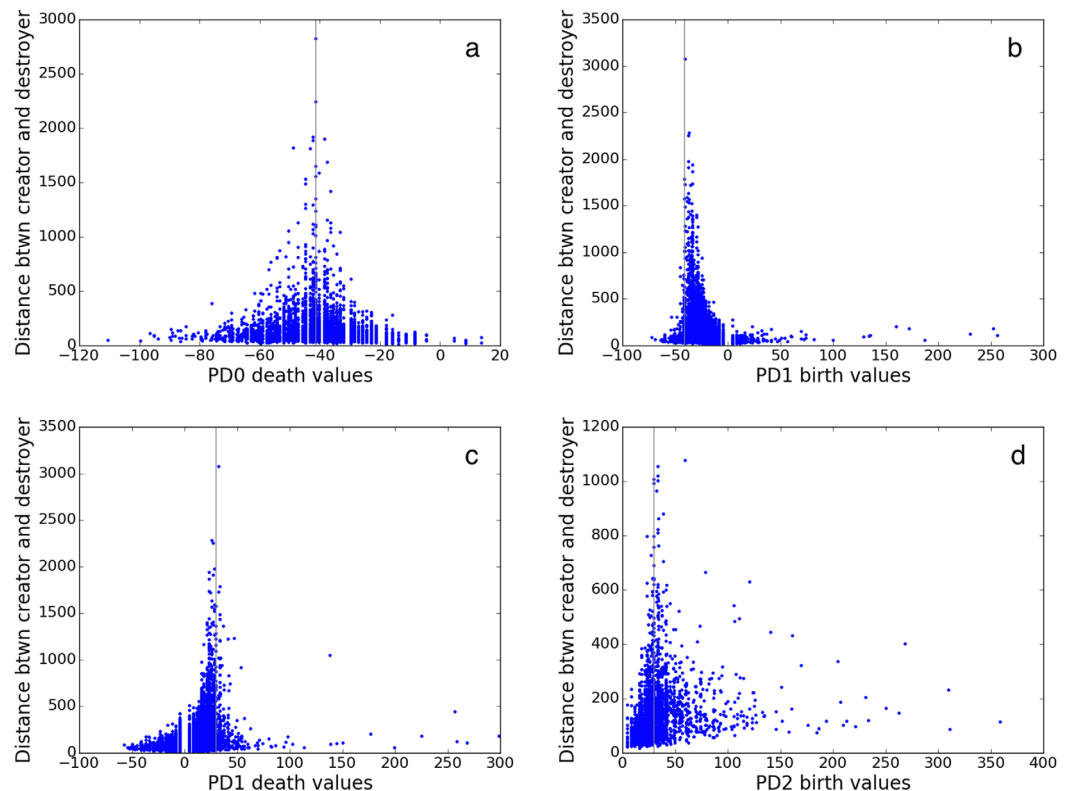


Figure 8. Distances between creator and destroyer persistence pairs for the sand pack (vertical and horizontal axes are μm). (a) PD0 deaths versus separation. (b) PD1 births versus separation. (c) PD1 deaths versus separation. (d) PD2 deaths versus separation. The gray lines mark the relevant critical percolating length scales, $r_{pore} = -41.3 \mu\text{m}$ and $r_{grain} = 29.7 \mu\text{m}$.

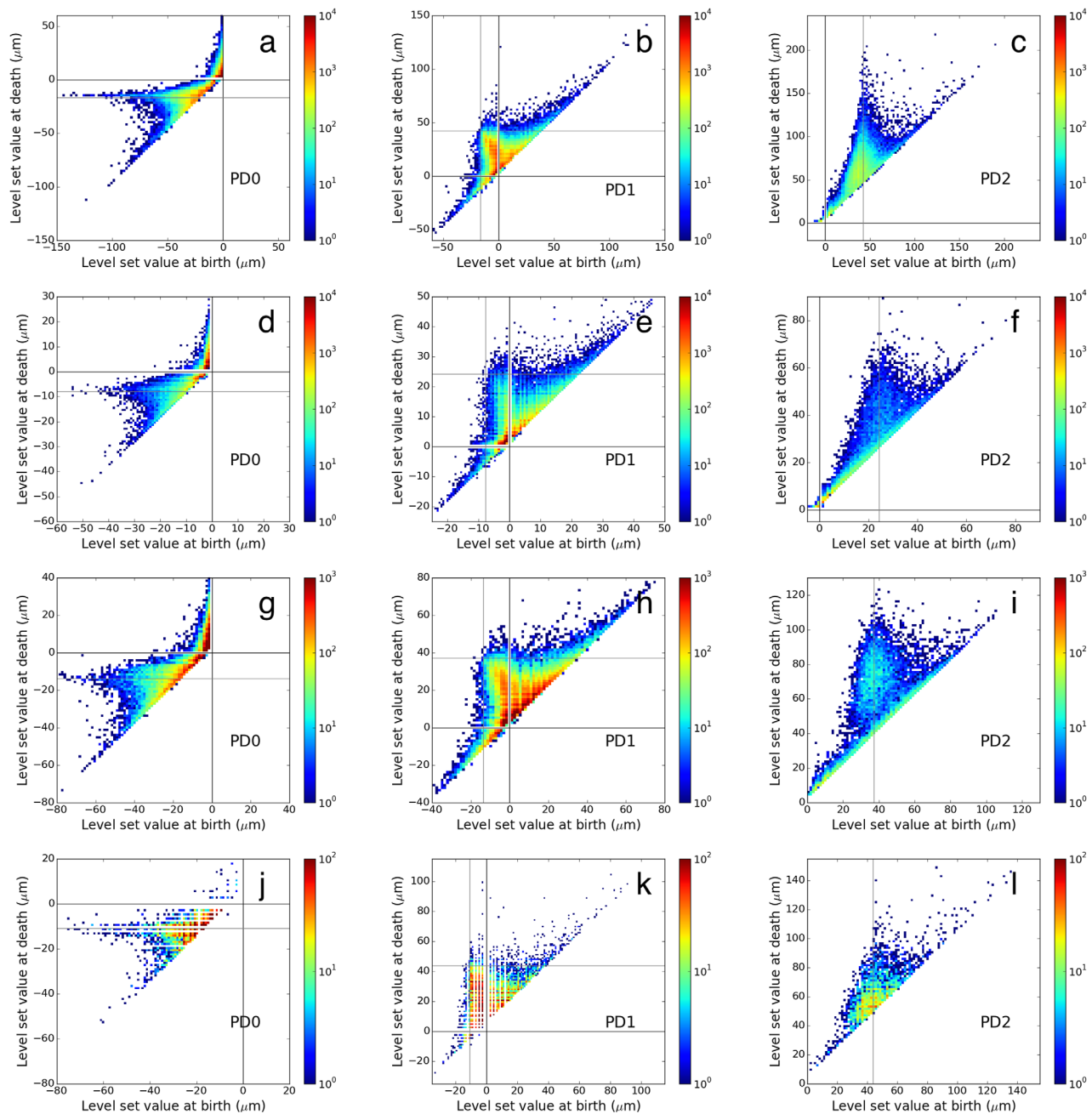


Figure 9. Persistence diagrams of four sandstone samples. (a–c) PD0, PD1, PD2 for Bentheimer (Ben2,2), (d–f) same for Berea, (g–i) Clashach1, and (j–l) Castlegate. The (b, d) points are binned in a 2-D histogram with logarithmic colorbar scale shown in the right. The gray horizontal and vertical lines show the critical percolating radii in μm .

Plots of the separations between creator and destroyer critical points in persistence pairs are shown for a single sample in Figure 10. Again we see that the narrow peaks occur close to the relevant percolation thresholds, but the single maximal point can be significantly offset. A more sophisticated technique for determining the location of the persistence peaks would improve our estimates of percolation thresholds.

4.3. Mt Gambier Limestone

The Mt Gambier limestone sample is a counterpoint to the above sandstones and demonstrates the utility of persistence diagrams in giving a comprehensive picture of the pore space geometry and topology.

A 2-D section and the persistence diagrams are shown in Figures 6c and 11; the distances between persistence pairs are summarized in Figure 12. This sample has no clearly defined length scales in the grain or

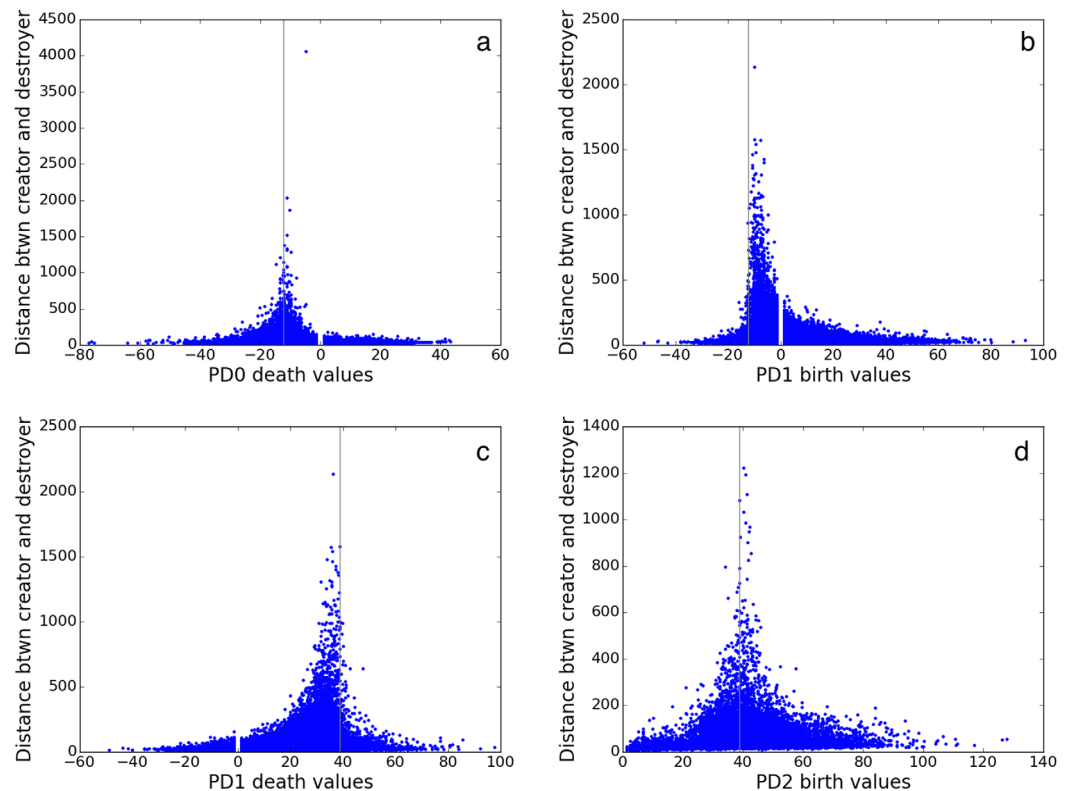


Figure 10. Distances between creator and destroyer persistence pairs for the Clashach2 sandstone (vertical and horizontal axes are in μm). (a) PD0 deaths versus separation. (b) PD1 births versus separation. (c) PD1 deaths versus separation. (d) PD2 births versus separation. The gray lines mark the relevant critical percolating length scales, $r_{\text{pore}} = -12.2 \mu\text{m}$ and $r_{\text{grain}} = 38.8 \mu\text{m}$.

pore sizes, a larger degree of similarity between the two phases and the percolation length scales do not provide the tight bound on the majority of PD1 points that we saw in the sandstones. The discrepancy between the percolation lengths and the dense region of PD1 points to the lack of homogeneity and suggests that the spanning cluster found by the percolation algorithm might be less robust to perturbation than those for the sandstones. For example, for c just larger than r_{pore} , the first Betti number of $L_f(c)$ is still small, telling us that there are few redundant pathways through the pore space at this length scale. Only when c is significantly larger (i.e., closer to 0) is there a jump in the number of redundant paths through the pore space.

The maximally persistent and separated births and deaths are given in Table 1 and surprisingly, there is agreement to within 10% for the maximally persistent PD0 death value and r_{pore} , and the maximally persistent PD2 birth and r_{grain} . We also calculate the dominant persistence length scales for the Mt Gambier data using the top 1% of persistence pairs in each dimension, and list results in Table 2. These show that the mean PD2 births are within 10% of r_{grain} , but mean PD0 deaths are now at a significantly smaller length, at $-22 \mu\text{m}$ compared to $r_{\text{pore}} = -31.2$. The data from PD1 birth and death values do not correspond to either of the percolating radii.

5. Conclusions

Persistent homology provides robust and rigorous definitions of topological descriptors that quantify connectivity over a range of length scales. Persistence analysis is therefore a powerful method to characterize the structure of a porous material in a way that is relevant for transport processes. By applying persistence analysis to the signed Euclidean distance transform of a binary image, we are able to incorporate a wealth of geometric information while also improving the robustness of topological measures.

We have shown that much can be inferred about a porous material by dividing PD1 (the information about rings or loops) into quadrants. The influence of a loop on the transport properties of a material will be quite

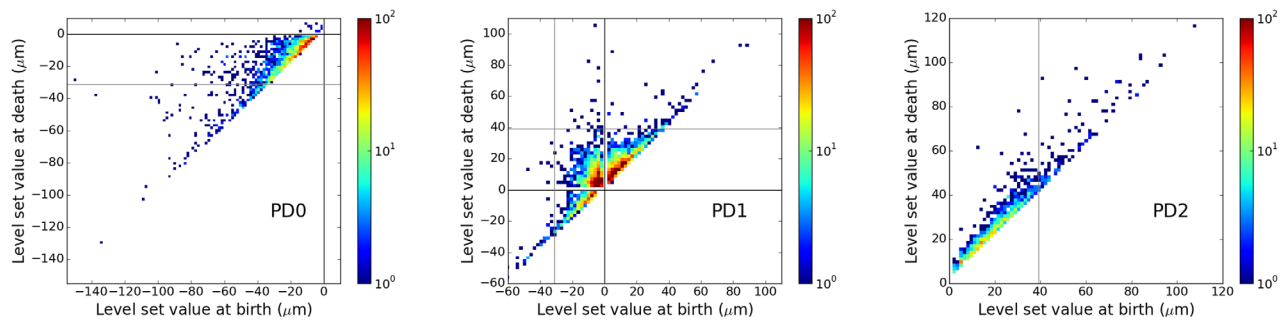


Figure 11. Persistence diagrams for a 512^3 volume of a distance transform for the Mt Gambier limestone. The (b, d) points are binned in a 2-D histogram with logarithmic colorbar scale shown in the right. The gray horizontal and vertical lines show the critical percolating length scales.

different depending on whether its birth and death values are positive or negative. Negative values correspond to critical points lying within the pore space, while positive values are within the grain material. The relative number of PD1 points with $b < d < 0$ or $0 < b < d$ tells us about the level of compaction or cementation in a sandstone or granular packing. The sandstones have many more PD1 points in the grain phase than the unconsolidated sand packing, for example.

For granular packs and well-sorted sandstones, we find a strong correlation between the percolation threshold and the emergence of large numbers of highly persistent loops in the pore space. The persistence lifetimes, i.e., $(d - b)$ values, of pore and grain phase components (PD0 and PD2) are peaked at the critical percolating-sphere radii for the respective phases. We also see that scatter plots of the spatial separation between birth and death critical points are sharply peaked at the relevant percolating radii. This is the first evidence of a connection between persistent homology and percolation, and an explanation of why these quantities are related is given in Appendix A.

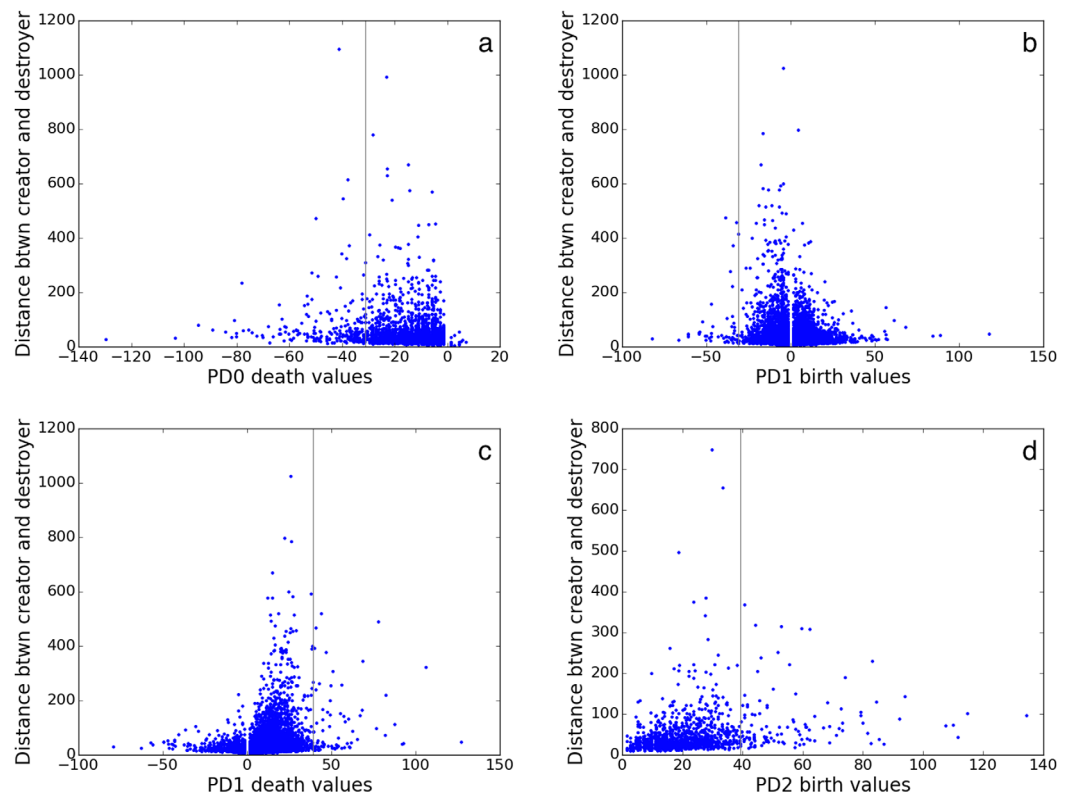


Figure 12. Distances between creator and destroyer persistence pairs for the Mt Gambier limestone sample (vertical and horizontal axes are in μm). (a) PD0 deaths versus separation. (b) PD1 births versus separation. (c) PD1 deaths versus separation. (d) PD2 births versus separation. The gray lines mark the relevant critical percolating length scales, $r_{\text{pore}} = -31.2 \mu\text{m}$ and $r_{\text{grain}} = 39.3 \mu\text{m}$.

Table 2. Persistence Length Scales for Top 1% Pairs From Mt Gambier Data^a

	Estimated Quantity	Value (μm)
.Percolating sphere radius	r_{pore}	−31.2
	r_{grain}	39.3
Top 1% persistence lifetimes	Mean PD0 deaths	−21.6
	Mean PD1 births	−14.4
	Mean PD1 deaths	37.1
	Mean PD2 births	36.2
Top 1% pair separations	Mean PD0 deaths	−22.6
	Mean PD1 births	−10.1
	Mean PD1 deaths	26.9
	Mean PD2 births	38.3

^aPersistence estimates that are within 10% of the percolating sphere radii are highlighted in bold.

We have estimated percolation radii from the birth and death values of persistence pairs with maximal lifetimes and maximal spatial separation and find good agreement with independent percolation calculations on the images. Maximal quantities are notoriously sensitive to errors, so a more accurate method for estimating percolating radii from persistence information would be to use mean birth and death values from a larger group of the most highly persistent or separated pairs. This entails choosing criteria for selecting the larger group, such as the top $p\%$ of pairs, or all

pairs above a suitable threshold on the persistence or separation. Such analysis should improve results for the relatively homogenous samples such as the sandstones.

The connection between persistence and percolation is less clear in highly heterogeneous systems, exemplified by the Mt Gambier limestone sample. In such systems, the numerical dominance of smaller features can overwhelm the signal from the main transport pathways. The persistence analysis does not recover the percolation length scales in this example. A larger sample might improve the picture. However, it is likely that the lack of a smooth distribution of grain and pore sizes in a heterogeneous sample obscures the simple relationship observed between percolation and persistence length scales of the homogenous sandstones.

In terms of computational cost, building the persistence diagrams is naturally more expensive than calculating the percolation threshold directly. The advantage of the persistence information is that it provides a more complete picture of the geometric and topological structure of the sample (PD0 births give the maximally inscribed sphere pore-size distribution, PD0 deaths and PD1 births provide the throat radii, PD2 deaths are the grain-sizes and PD2 births and PD2 deaths are grain-contact radii). This additional information may help provide a purely geometric characterization of the permeability of porous materials. A study of simulated quasi-2-D porous materials found that the effective conductivity term σ/σ_0 in the Katz-Thompson formula (1) could be replaced by $\left(\frac{1-\chi_a}{N}\right)^\alpha$, the “genus per grain,” Scholz *et al.* [2012]. To extend that work to real 3-D, porous materials requires the robust quantification of pore space connectivity and the identification of grains. Both these quantities can be easily extracted from the persistence diagrams.

Appendix A: The Persistence Pairing Algorithm and its Connection to Percolation

The topological changes in the lower level sets of the SEDT, $L_\lambda(c)$, are represented by the filtration of a cell complex known as the *Morse complex*. As mentioned in section 2, the Morse complex is defined by mapping the critical points of the SEDT to topological elements (vertices, edges, patches, solids) in a cell complex. Local minima are vertices, or 0-cells. Index-1 saddle points become edges or 1-cells. Such a saddle point has two descending gradient flow-paths that terminate at local minima, these vertices are the end points of the edge. The graph of vertices and edges with negative critical values in the SEDT is analogous to a network model of the pore space. An index-2 saddle point becomes a 2-cell, and its adjacent 1-cells are those 1-saddles that are reached by descending flow paths from the 2-saddle. A local maximum is a 3-cell. We note that each 2-saddle has exactly two ascending gradient flow paths that terminate at local maxima. This adjacency of 3-cells via 2-cells defines a dual graph that is effectively a grain-contact network. Alternatively, the dual graph can be defined from the filtration of upper level sets of the SEDT adding local maxima, 2-saddles, 1-saddles, and minima to $U_\lambda(c)$ as c decreases from the global maximum. Dual models such the above have been discussed for porous materials, for example by Glantz and Hilpert [2007], and are related to the duality between Voronoi partitions and Delaunay triangulations of point patterns.

The lower level set filtration of the Morse complex is an ordering of all the cells by the critical values of their associated critical points, we refer to the critical value as the weight of the cell. Properties of Morse functions ensure that faces are added before the higher-dimensional cells they are adjacent to. We now describe the persistence pairing of 1-cells with 0-cells as the filtration grows. When a vertex (0-cell, local

minimum) is added it automatically creates a new component, i.e., a new zero-dimensional homology class. Each time an edge (1-cell, 1-saddle) is added, it must either join two components (causing the death of a zero-dimensional homology class) or create a loop (a new one-dimensional homology class). If the edge joins two distinct components, it is paired with an as-yet unpaired vertex from these two components that is as close as possible in weight. At the beginning of the growth process, most edges will join isolated vertices together, and in this case the edge is paired with the adjacent vertex of highest weight. If an edge joins two components with more than a single vertex each, elementary combinatorics guarantees that each connected component has a single unpaired vertex, so the edge that bridges them is paired with the unpaired vertex of highest weight. It follows that the vertex with minimal weight (the largest pore) remains unpaired by the persistence algorithm.

The growth of the filtration is exactly the same growth used when testing for percolation. The percolation algorithm finds the lowest weight for which there is a path from one side of the sample to the other. Near the percolation threshold, r_{pore} , it is known that there are clusters of all sizes present. So it is possible for some edges with weights around r_{pore} to join two large components and the persistence algorithm then pairs such edges with vertices that may be far away in space and in weight-value.

Past the percolation threshold, the edges added are more likely to create loops (they are PD1 births), or to join a smaller component to the spanning cluster, and the smaller component is more likely to have its unpaired vertex closer in weight to that of the new edge. This means both the persistence lifetimes $d - b$ and the pair separations $\|\mathbf{x}_d - \mathbf{x}_b\|$ in PD0 reach their maximum for values of d near r_{pore} .

A similar argument applied to the dual structure of 3-cells and 2-cells of the Morse complex (i.e., the grains and grain-contacts) establishes the observed relationship between PD2 points and the grain phase percolation length scale: persistence lifetimes and pair separations reach their maximum for birth values near r_{grain} .

Next we discuss the observed structure in the PD1 diagram, focusing on the reason for the empty quadrant to the upper left of the percolation thresholds and the densely filled triangular region to the lower right of the percolation point. Each edge in the forward (lower level set) filtration of the Morse complex must either be a death event in PD0 or a birth event in PD1. This explains why most points in PD1 have birth $b > r_{pore}$: more cycles are created after the spanning cluster has formed. Similarly, from the complementary dual perspective, we see that most points in PD1 will have death values $d < r_{grain}$. On average then, we expect to see most pairs in the region satisfy $r_{pore} < b < d < r_{grain}$. Also, when the birth and death values have a continuous distribution, (as is certainly true for the Swiss cheese model and a good approximation for the sandstones, but not the sphere packing or the limestone) the (b, d) points fill the region to the lower right of (r_{pore}, r_{grain}) .

Finally we consider the almost empty quadrant $b < r_{pore}, d > r_{grain}$. If a persistence pair in PD1 has $b < r_{pore}$, we know that the 1-cycle must belong to a finite (nonspanning) component of $L_f(b)$, with b as its narrowest throat (i.e., b is the least negative value of the SEDT in this cycle). Similarly, if $d > r_{grain}$ the cycle in the "grain" phase must belong to a finite (nonspanning) component of $U_f(d)$ with d as the narrowest point in $U_f(d)$ along this cycle. These two cycles, one in $L_f(b)$ and one in $U_f(d)$ are linked for $b < c < d$. The existence of two linked cycles, born outside the spanning clusters and persisting for a large range of length scales is highly unlikely in random homogenous structures such as the sandstone samples or the Swiss cheese model.

Acknowledgments

This work was funded in part by ARC grant DP110102888 (to A.P.S. and V.R.), and ARC Future Fellowships FT100100470 (A.P.S.) and FT140100604 (V.R.). Financial support was also provided by the member companies of the ANU/UNSW Digicore Consortium, who have also contributed samples for this study. We thank Lincoln Paterson for providing the Mt Gambier limestone sample. The segmented binary image data for the sand pack, limestone, Castlegate sandstone, and the random close pack beads are available for download from Sheppard and Schröder-Turk [2005]. Code for building the Morse complex from a digital image and computing the persistence diagrams is available in the *diamorse* package [Delgado-Friedrichs, 2015].

References

- Bakke, S., and P. Oren (1997), 3-D pore-scale modelling of sandstones and flow simulations in the pore networks, *SPE J.*, 2(2), 136–149.
- Bendich, P., H. Edelsbrunner, and M. Kerber (2010), Computing robustness and persistence for images, *IEEE Trans. Vis. Comput. Graph.*, 16, 1251–1260.
- Blunt, M. J. (2001), Flow in porous media: Pore-network models and multiphase flow, *Curr. Opin. Colloid Interface Sci.*, 6(3), 197–207, doi:10.1016/S1359-0294(01)00084-X.
- Broadbent, S. R., and J. M. Hammersley (1957), Percolation processes, *Proc. Cambridge Philos. Soc.*, 53(3), 629–641, doi:10.1017/S0305004100032680.
- Carlsson, G. (2009), Topology and data, *Bull. Am. Math. Soc.*, 46(2), 255–308.
- Cohen-Steiner, D., H. Edelsbrunner, and J. Harer (2007), Stability of persistence diagrams, *Discrete Comput. Geom.*, 37, 103–120.
- Delgado-Friedrichs, O. (2015), Diamorse: Digital image analysis using discrete Morse theory and persistent homology, The Australian National University, Canberra. [Available at <https://github.com/AppliedMathematicsANU/diamorse>.]
- Delgado-Friedrichs, O., V. Robins, and A. Sheppard (2015), Skeletonization and partitioning of digital images using discrete Morse theory, *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(3), 654–666, doi:10.1109/TPAMI.2014.2346172.

- Edelsbrunner, H., D. Letscher, and A. Zomorodian (2002), Topological persistence and simplification, *Discrete Comput. Geom.*, 28(4), 511–533.
- Fatt, I. (1956), The network model of porous media, *Trans. AIME* 207, 144–181.
- Glantz, R., and M. Hilpert (2007), Dual models of pore spaces, *Adv. Water Resour.*, 30(2), 227–248.
- Herring, A. L., E. J. Harper, L. Andersson, A. Sheppard, B. K. Bay, and D. Wildenschild (2013), Effect of fluid topology on residual nonwetting phase trapping: Implications for geologic CO₂ sequestration, *Adv. Water Resour.*, 62, 47–58, doi:10.1016/j.advwatres.2013.09.015.
- Katz, A. J., and A. H. Thompson (1986), Quantitative prediction of permeability in porous rock, *Phys. Rev. B*, 34(11), 8179–8181, doi:10.1103/PhysRevB.34.8179.
- Kurtuldu, H., K. Mischaikow, and M. F. Schatz (2011), Measuring the departures from the Boussinesq approximation in Rayleigh–Bénard convection experiments, *J. Fluid Mech.*, 682, 543–557, doi:10.1017/jfm.2011.244.
- Lorenz, C., and R. Ziff (2001), Precise determination of the critical percolation threshold for the three dimensional Swiss cheese model using a growth algorithm, *J. Chem. Phys.*, 114(8), 3659–3661.
- Prodanovic, M., W. Lindquist, and R. Seright (2007), 3d image-based characterization of fluid displacement in a Berea core, *Adv. Water Resour.*, 30(2), 214–226, doi:10.1016/j.advwatres.2005.05.015.
- Rintoul, M. (2000), Precise determination of the void percolation threshold for two distributions of overlapping spheres, *Phys. Rev. E*, 62(1), 68–72.
- Robins, V. (1999), Towards computing homology from finite approximations, *Topol. Proc.*, 24, 503–532.
- Robins, V., P. J. Wood, and A. P. Sheppard (2011), Theory and algorithms for constructing discrete Morse complexes from grayscale digital images, *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(8), 1646–1658.
- Sahimi, M. (1993), Flow phenomena in rocks: From continuum models to fractals, percolation, cellular automata, and simulated annealing, *Rev. Mod. Phys.*, 65(4), 1393–1534, doi:10.1103/RevModPhys.65.1393.
- Scholz, C., F. Wirner, J. Götz, U. Rüe, G. E. Schröder-Turk, K. Mecke, and C. Bechinger (2012), Permeability of porous materials determined from the Euler characteristic, *Phys. Rev. Lett.*, 109, 264504.
- Sheppard, A., et al. (2014), Techniques in helical scanning, dynamic imaging and image segmentation for improved quantitative analysis with X-ray micro-CT, *Nucl. Instrum. Methods Phys., Sect. B*, 324, 49–56, doi:10.1016/j.nimb.2013.08.072.
- Sheppard, A. P., and G. E. Schröder-Turk (2005), The network generation comparison forum, The Australian National University, Canberra. [Available at http://people.physics.anu.edu.au/~aps110/network_comparison.]
- Verri, A., C. Uras, P. Frosini, and M. Ferri (1993), On the use of size functions for shape analysis, *Biol. Cybern.*, 70(2), 99–107, doi:10.1007/BF00200823.
- Wildenschild, D., and A. P. Sheppard (2013), X-ray imaging and analysis techniques for quantifying pore-scale structure and processes in subsurface porous medium systems, *Adv. Water Resour.*, 51, 217–246, doi:10.1016/j.advwatres.2012.07.018.