Flow Path Resistance in Heterogeneous Porous Media Recast into a Graph-Theory Problem

Z. Kanavas[1] · F. J. Pérez-Reche[2] · F. Arns[3] · V. L. Morales[1]

Abstract

This work aims to describe the spatial distribution of flow from characteristics of the underlying pore structure in heterogeneous porous media. Thousands of two-dimensional samples of polydispersed granular media are used to (1) obtain the velocity field via direct numerical simulations, and (2) conceptualize the pore network as a graph in each sample. Analysis of the flow field allows us to distinguish preferential from stagnant flow regions and to quantify how channelized the flow is. Then, the graph's edges are weighted by geometric attributes of their corresponding pores to find the path of minimum resistance of each sample. Overlap between the preferential flow paths and the predicted minimum resistance path determines the accuracy in individual samples. An evolutionary algorithm is employed to determine the "fittest" weighting scheme (here, the channel's arc length to pore throat ratio) that maximizes accuracy across the entire dataset while minimizing over-parameterization. Finally, the structural similarity of neighboring edges is analyzed to explain the spatial arrangement of preferential flow within the pore network. We find that connected edges within the preferential flow subnetwork are highly similar, while those within the stagnant flow subnetwork are dissimilar. The contrast in similarity between these regions increases with flow channelization, explaining the structural constraints to local flow. The proposed framework may be used for fast characterization of porous media heterogeneity relative to computationally expensive direct numerical simulations.

Article Highlights

1. A quantitative assessment of flow channeling is proposed that distinguishes pore-scale flow fields into preferential and stagnant flow regions.
2. Geometry and topology of the pore network are used to predict the spatial distribution of fast flow paths from structural data alone.
3. Local disorder of pore networks provides structural constraints for flow separation into preferential v stagnant regions and informs on their velocity contrast.

✉ V. L. Morales
   vermorales@ucdavis.edu

Extended author information available on the last page of the article

1 Introduction

Flow channelization is an ubiquitous phenomenon in subsurface media with implications in many engineering endeavors including groundwater management (Charbeneau 2006; Freeze and Cherry 1979; Šimnek et al. 2003; LeBlanc et al. 1991), oil and gas recovery (Middleton et al. 2015; Orr and Taber 1984; Wong 1988), and geotechnical engineering (Dullien 1992; Bear 1988). Observations of nonuniform flow distribution have been reported at the pore- and continuum-scales in laboratory studies (Kurotori et al. 2019; Datta et al. 2013; Carrel et al. 2018; Morales et al. 2017; Holzner et al. 2015; de Anna et al. 2021), field experiments (Goeppert et al. 2020; Bianchi et al. 2011; Zheng et al. 2011; Abelin et al. 1991; Guihéneuf et al. 2014; Le Borgne et al. 2006; Rasmuson and Neretnieks 1986; Winograd and Pearson 1976; Flury et al. 1994; Hendrickx and Flury 2001) and numerical simulations (Alim et al. 2017; Meyer and Bijeljic 2016; Tyukhova and Willmann 2016; Siena et al. 2014, 2019; Nissan and Berkowitz 2019; Puyguiraud et al. 2019; Kang et al. 2014; Bijeljic et al. 2013; Fiori et al. 2011, 2013; Zinn and Harvey 2003; Hyman 2020). The multi-scale structural heterogeneity of porous and fractured media gives rise to complex flow patterns that contain pronounced regions of preferential flow associated with fast local velocities, and areas of flow stagnation where velocities are slow (Siena et al. 2019). Independent of the scale considered, a well-accepted hallmark of channelized flow is the non-Gaussian distribution of the underlying velocity field, which has strong implications for anomalous transport and incomplete mixing (Alim et al. 2017; Dentz et al. 2011).

A formal quantitative definition of preferential flow is warranted to evaluate its occurrence in different samples and to identify the geologic properties that prompt it (Renard and Allard 2013; Hyman 2020; Le Borgne et al. 2006; Su et al. 2001). In weakly heterogeneous subsurface environments, relatively uniform flow is well described and upscaled by Darcy's law. For these media, pedotransfer functions provide a convenient way to estimate hydraulic properties from basic information of the geologic formation (Price et al. 1911; Carman 1937; Schaap et al. 2001). In strongly heterogeneous subsurface environments, however, neither the flow nor the medium's variability follow spatial stationarity rules (Chakraborty et al. 2020; Bradford et al. 2017). To accurately resolve the flow, direct numerical simulations of the governing equations in the detailed pore space are required. These calculations are associated with high computational costs and are not trivial to upscale (Dentz et al. 2011). As a consequence, the elusive link between medium structure attributes and complex flow distribution presents a challenge for accurate predictions of field-scale flow and transport problems. Relatable connectivity measures for static and dynamic metrics have been proposed to elucidate this link (Knudby and Carrera 2005; Renard and Allard 2013; Hobé et al. 2018; Hyman 2020; Rizzo and de Barros 2017). Static connectivity measures are defined from structural information, while dynamic connectivity measures reflect the system's response in its flow field and/or solute transport. In this framework, the principal aim is to infer dynamic flow properties from static connectivity measures (Knudby and Carrera 2005).

Recent work has focused on splitting the heterogeneous structure-flow relationship into its two main components—one for the preferential flow paths of high conductivity and another for low conductive layers of stagnant flow (Knudby and Carrera 2005). For the link in regions of high conductivity, Siena et al. (2019) propose an analytical function to describe how the velocity distribution in preferential flow paths is related to the pore size distribution and the spatial correlation. At the pore-scale, Jimenez-Martinez and Negre (2017) take advantage of network analysis and graph theory to build a model that

accounts for anisotropy in multiphase systems. Through eigenvector centrality metrics of static connectivity, the authors are able to infer reasonably well the fraction of the porous medium that behaves as stagnation zones and/or preferential paths. For gas-injection problems, Yeates et al. (2020) optimize a weighted graph representation of the pore space to predict preferential foam flow from shortest path analysis. At the continuum-scale, Rizzo and de Barros (2017) provide a connectivity link for heterogeneous permeability fields. In their work, the authors represent static connectivity as a graph in which the least resistance path is found and used to estimate the time of first arrival for risk analysis. By contrast, less attention has been paid to the link in regions of low conductivity. The work by de Anna et al. (2017) shows that local low velocities in disordered media are well described by a power-law distribution function with an exponent obtained from the pore throat size distribution. Similarly, Matyka et al. (2016) present work on the velocity distribution function in random porous media, where their distributions are described by a power-exponential distribution whose parameters depend on the porosity. For regions of low conductivity, it remains to be determined if the suggested conductivity links are valid in highly correlated porous matrices.

The principal focus of this study is to understand how flow is spatially distributed in and constrained by the underlying microstructure in samples of arbitrary heterogeneity. Our first aim is to predict the location and strength of preferential flow regions from pore network topology and channel geometry. Our second aim is to explain the structural constraints for preferential and stagnant flows that give rise to variable velocity distribution contrasts between these two regions. To do this, we consider an extensive dataset of two-dimensional samples of polydispersed porous media whose flow classification spans highly uniform to highly channelized.

The remainder of the paper is structured as follows. Section 2 details the methods used to extract static and dynamic connectivity measures and graph theory approaches used to find their link. Section 3 reports and discusses findings from the three major analyses. The first proposes a metric to quantify the degree of flow channelization and spatially distinguishes preferential from stagnant flow regions. The second reports the optimized shortest path (least resistance curve) to identify preferential flow paths from information of the underlying microstructure. The third shows how the (dis)order of the structural arrangement of the network constrains flow development. Section 4 provides conclusions and main take home messages.

2 Methods

2.1 Image Preparation and Flow Simulations

The porous medium consists of well-sorted Nafion grains with an average diameter of 3.6 mm that are wet-packed into a cubic flow cell of 3.8 cm in length (Morales et al. 2017). Micro-computed tomography is used to obtain the cross section images of a three-dimensional porous structure at a resolution of 25 μm, producing an image of 1540^3 voxels. Each tomographic image is cropped to exclude approximately 100 μm on all edges of the porous medium to exclude wall effects as much as possible (see Figure S1) while maximizing the structural information retained. Image segmentation identifies each voxel as either solid or void. To improve the connectivity in the 2D cross sections, each image is modified slightly. First, watershed segmentation is applied to ensure the pore space is permeable for flow

simulation. Next, select samples are subjected to varying steps of solid erosion (from 1 to 9 pixels) to create samples of variable pore geometry while preserving the original topology. Altogether, 3620 unique 2D samples are used in this study. The average pore throat across the 2D samples is 480 μm, the full pore-throat distribution is shown in Figure S2.

To obtain the Eulerian velocity field $\mathbf{u}(\mathbf{x})$ in each unique porous sample, we solve the Navier–Stokes equations using the finite element numerical model, COMSOL (www.comsol.com). In COMSOL, we simulate laminar flow where the inlet boundary condition is a line at the left border of the domain (broken where grains are at the edge) specified to achieve a mean velocity of 0.001 m/s. The outlet boundary condition is a line at the right border set at zero pressure. The top and bottom boundaries (perpendicular to the flow direction), as well as the grain surface boundary conditions, are set to no-slip.

2.2 Percolation Threshold and Flow Region Identification

A modified critical thresholding method is used to quantitatively evaluate the degree of flow channeling observed and to identify regions of preferential or stagnant flow (Otsu 1979; Ambegaokar et al. 1971; Kirkpatrick 1971). Briefly, points $\mathbf{r}$ of the pore space are excluded if the local velocity vector normalized by the mean of the modulus, $\mathbf{u}(\mathbf{r})/\langle|\mathbf{u}|\rangle$, is smaller than a given threshold p. That threshold value is iteratively lowered until a value $p = p_c$ is reached for which a continuous subdomain is first established that connects inlet and outlet boundaries (i.e., a percolating path is found). Pore space points are then classified as being part of the preferential or stagnant flow as follows:

$$\text{if } \mathbf{u}(\mathbf{r})/\langle|\mathbf{u}|\rangle \geq p_c \text{ then } \mathbf{u}(\mathbf{r}) = 1 \text{ (preferential) else } \mathbf{u}(\mathbf{r}) = 0 \text{ (stagnant)}.$$

The percolating subdomain along with any disconnected high-velocity regions is collectively classified as the preferential flow region (PFR). The remainder of the flow field is classified as the stagnant flow region (SFR). The percolating threshold (p_c) measures how channelized the flow is (high $p_c \gg 1$ corresponds to channelized flow, low $p_c \leq 1$ corresponds to uniform flow) and also indicates how much earlier particle breakthrough times might be expected in each sample. An advantage of this definition of flow channelization is that all flow fields are placed into a relative context using a dimensionless form of the percolation parameter.

Lastly, k-means clustering (Wilmott 2019) was applied to the values of p_c for each sample to group them into three characteristic flow regimes: uniform, intermediate, and channelized. The number of flow regimes in the dataset was estimated via the gap statistic (Tibshirani et al. 2001). This simplified classification of the flow will be useful for qualitative comparison of different samples. The authors note that p_c cutoffs found here for different flow classes are specific to this dataset.

2.3 Extracting Graphs from the Pore-Network

We use a graph-based approach to examine the topology and geometry of the pore space in a mathematical framework. Each porous media sample is first represented by its equivalent graph, $G(V, E)$, using a medial axis method [Skeletonize FIJI plugin (Arganda-Carreras et al. 2010)]. Here, V is a set of nodes and E is a set of edges. In this reduced representation of the pore space, the nodes in V correspond to pore bodies and edges in E correspond to pore channels. Then, the produced skeleton is used to determine the topological

characteristics and channel properties of the network with the method developed by Pérez-Reche et al. (2012).

From this approach, we obtain an adjacency matrix with stored topologic information about node degree (k), and edge geometric information, including arc-length (L), Euclidian distance (D) and pore width statistics ($S(x)$) (mean, minimum, maximum, range). Various combinations of the topological and geometric characteristics define the set $\mathcal{H}$ of network structural attributes used in our study (summarized in Table 1). These attributes are subsequently used to identify the pathway where channelized flow occurs from structural information alone.

2.4 Path Prediction

2.4.1 Shortest Path Analysis

Each edge $e \in E$ of the pore space network connects two nodes and can have a weight w_e that we describe below as a linear combination of the defined structural attributes from Table 1 (e.g., distance, width, or local resistivity). Because the magnitude of each attribute is widely different, their population is normalized to span a [0,1] range by the maximum value observed in the full dataset. We shall denote the normalized quantities as H_i, where

Table 1 Structural attributes of individual channels used to bias the shortest path analysis

Interpretation	Structural attributes (H_i)
Arc length	L
Euclidean distance	D
Mean width[a]	$\bar{S}$
Mean width[a,b]	$\bar{S}^{\alpha}, \alpha = -2.5$
Minimum width (pore throat)[a]	S_m
Maximum width[a]	S_M
Range of width[a]	$S_M - S_m$
Exponential decay of pore throat[c]	$\exp(S_m^{-1})$
Mean resistance[d]	$L/\bar{S}$
Mean resistance$^{\alpha}$	$(L/\bar{S})^{\alpha}, \alpha = -2.5$
Minimum resistance	L/S_M
Maximum resistance	L/S_m
Poiseuille's law for resistance	$\left(\frac{\bar{S}}{2}\right)^4/L$
Curvature[a]	$1 - (D/L)$
Node degree[a,e]	k
Volume[a]	$L\left(\frac{S_m}{2}\right)^2 \pi$

[a]Channel weight was also tested in its inverse form

[b]From Yeates et al. (2020). Narrow throats result in greater edge weight than wide pore throats and the rate of change in weight is greater for narrow pore throats than wide pore throats

[c]From Jimenez-Martinez and Negre (2017). Similar effect as $\bar{S}^{-\alpha}$, but more amplified.

[d]From Ewing and Hunt (2009).

[e]k favors low connectivity regions, $1/k$ favors hubs in the network

$i = 1, \ldots, \#\mathcal{H}$ is the index for each quantity, following Table 1. The weight of a given edge is then expressed as a linear combination of the normalized structural attributes as follows:

$$w_e(\mathbf{x}) = \sum_{i \in \#\mathcal{H}} x_i H_i. \tag{1}$$

Here, $\mathbf{x} = \{x_i\}_{i=1}^{\#\mathcal{H}}$ are the coefficients whose values will be determined using a genetic algorithm as described below.

A path Γ in $G(V, E)$ is a sequence of nodes connected by edges. Assigning a weight to the edges generates a weighted graph $R(V, E)$, which is then used to approximate the path of minimum flow resistance. The authors note that the shortest path generally coincides with the path of minimum resistance, but not always. A source s and target t node are chosen to define the possible paths $\mathcal{P}_s^t$ on the graph $R(V, E)$. Since source and target nodes cannot be clearly identified in $R(V, E)$, we augmented each graph to include a communal source and a target node. For every channel that emanates near the inflow (outflow) boundary, an edge is added between the most distant node in the graph and the node representing the source (target) to use a single position of entry (exit) to the graph similar to the approach by Hyman (2020). Figure S3 illustrates the graph augmentation. Edges connected to the source or target node are given no weight and thereby do not contribute to a path's resistance. Then, the approximate minimum resistance is given by the minimum of an objective function defined as the sum of edge weights along all possible paths connecting source to target, viz. Rizzo and de Barros (2017):

$$\mathcal{R}(\mathbf{x}) = \min_{\Gamma \in \mathcal{P}_s^t} \sum_{e \in \Gamma} w_e(\mathbf{x}). \tag{2}$$

This problem can be solved using shortest path analysis (SPA) with Dijkstra's algorithm (Dijkstra 1959). The goal is to identify the most appropriate weighting scheme (i.e., the coefficients $\mathbf{x}$) for the graph such that the predicted shortest path mapped onto the PFR has maximal agreement.

2.4.2 Differential Evolution

A metaheuristic approach is used to search for potential weighting schemes to optimize $\mathcal{R}(\mathbf{x})$. More specifically, we use differential evolution (DE) which is a genetic algorithm that uses a greedy approach to minimize a given objective function by adjusting a population of coefficients $\mathbf{x}$ (Storn and Price 1997). In doing so, the algorithm evolves the coefficients in search of a better solution from a population of candidates. Briefly, a candidate solution is composed of a set of coefficients for different channel structural attributes which are mutated in each generation. Evolution starts from a population of randomly generated initial solutions whose fitness is measured in each generation. Fitness is here assessed via a modified Akaike information criterion (AIC), as described below. More fit solutions are stochastically selected from the current population and are both recombined and randomly mutated to form the next generation. The new generation of solutions is then used to iterate the algorithm. The generational process is terminated once the standard deviation of the population's fitness function falls below a user-defined threshold. In this work, the threshold was 0.01. More details about the DE implementation are provided in section S1 of SI.

The initial population contains n candidate solutions $\{m_k = \sum_{i \in \mathcal{H}_k} x_i H_i\}_{k=1}^n$, where $\mathcal{H}_k \subset \mathcal{H}$. The set of initial populations is explicitly designed to cover the entire parameter space, i.e., $\cup_{k=1}^n \mathcal{H}_k = \mathcal{H}$. In particular, the coefficients in an initial solution are randomly

sampled from a uniform distribution $U(0, 1]$. To prevent over-parameterization, each initial solution was limited to contain a maximum of three nonzero x_i coefficients chosen at random (i.e., the set $\mathcal{H}_k$ of structural attributes for a candidate solution j contains three elements from $\mathcal{H}$). The initial population was checked to ensure that every structural attribute appeared with a nonzero coefficient in at least one solution. Figure S4 illustrates an example initial population.

Fitness of a candidate solution is determined by the trade-off between goodness of fit of the weighting scheme across the entire dataset and the simplicity of the model. Goodness of fit is the arithmetic mean of accuracy for all samples in the dataset, μ_a. Accuracy, a_z, is evaluated for each individual sample and is estimated as the fraction of spatial overlap between the biased shortest path predicted (static connectivity) and the known PFR (dynamic connectivity). K, is the number of nonzero coefficients in the evolved solution given by $K = \#\{i | x_i > 0, i = 1, \dots, \#\mathcal{H}\}$. The modified AIC that the evolutionary algorithm minimizes is given by:

$$\text{AIC} = 2K - 2\ln(\mu_a). \tag{3}$$

To test each proposed weighting scheme, the samples were split into sets for training (80%) and testing (20%) sets via 5-fold cross-validation (see details in section S2 of SI). This step is implemented to assess how well the model results generalize to independent datasets. Model variability is captured by the accuracy distribution spread across the tested iterations.

2.5 Assortativity

To explain the spatial arrangement for preferential flow within a pore network, we measure the order between neighboring edges in terms of the conductance S_m. Our hypothesis is that edges that make up the preferential flow subnetwork are more similar (i.e., ordered in some way) than those that make up the stagnant flow subnetwork. To test this hypothesis, we define an assortativity coefficient, $r \in [-1, 1]$, using S_m as a scalar characteristic for edges (not for nodes as is typically done) (Newman 2018). Section S3 of the SI provides a detailed description of how assortativity is calculated. Positive values of r indicate order, and negative values indicate disorder between edges of different S_m. To compare and contrast the degree of order in the system, three measures of assortativity are obtained for each sample: the preferential flow subnetwork (network edges overlapping with the percolating cluster, r_1), the stagnant flow subnetwork (network edges overlapping with the stagnant regions, r_2) and the entire pore network, r_W.

3 Results and Discussion

In this section, we report the measurements of connectivity in virtual experiments to infer dynamic flow quantities from static structural measures.

3.1 Quantitative Degree of Flow Channelization

The percolating threshold metric (p_c) proposed here allows objective quantification of the spectrum of flow channelization in any sample. From a dynamic standpoint, this approach

measures the strength of the percolating flow path that connects the boundaries of a given flow field. It is reasonable to expect that, in porous media, only a fraction of the porosity participates in bulk flow, however, uniform this is. Intuitively, one would expect increasingly strong preferential flow paths (determined by large p_c) to be restricted to a diminishing fraction of the pore space. An illustrative example is provided in Fig. 1 (Right) for two velocity fields of uniform and channelized flow and the spatial extent of the preferential pathway. The quantitative relationship between p_c and the fraction of porosity occupied by the PFR, ϕ_1, is illustrated in Fig. 1 (Left). We find that ϕ_1 decreases exponentially with p_c and increases linearly with porosity ϕ. Since the PFR in our dataset is limited to a fraction $\phi_1 \in (0, \phi)$ of the sample, this exponential relationship captures the notion that only a fraction of the pore space can carry the bulk flow within the preferential flow paths for channelized flows of a given strength. While additional testing on numerous independent datasets is required to generalize this exponential relationship, similar behavior was observed for different 2D geometries of granular porous media, as shown in Figure S7.

The samples in our dataset are grouped into one of three pre-defined flow classes: *uniform* ($p_c \leq 1.1$), *intermediate* ($1.1 < p_c < 2.0$), or *channelized* ($p_c \geq 2.0$) (see Figure S8 for the full histogram). This flow classification allows one to contrast velocity probability distribution in the PFR versus the SFR for the extreme cases as shown in Fig. 2. The uniform flow samples (red) show that velocities experienced in the PFR (solid lines) are similarly distributed as those in the SFR (dashed lines). Conversely, the channelized flow samples (green) show large differences in the velocity distributions of the two flow regions. While this is not a surprising finding, it permits an explicit definition for the strength and spatial location of preferential flow paths, and the expected velocity distributions associated with PFR and SFR.

## 3.2 Prediction of Shortest Paths via Differential Evolution

The DE-SPA approach identified the fittest edge weighting scheme as $w_e(\mathbf{x}) = \sum_{i \in \#\mathcal{H}} x_i H_i$ with $\mathcal{H} = \{L/S_m\}$ across all five iterations of the cross-validation scheme. An exemplary weighting scheme evolution for one tested iteration is shown in Fig. 3 (Top). In
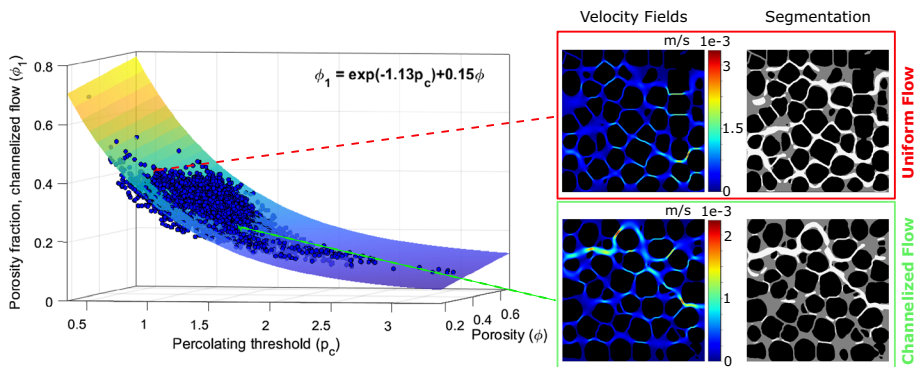

Fig. 1 (Left) Exponential function relating the porosity fraction associated with preferential flow, ϕ_1, the normalized velocity percolation threshold, p_c, and the porosity, ϕ. (Right) Velocity field for two typical realizations of pore structures with comparable porosity $\phi \sim 0.34$. Included is the corresponding segmentation as per the velocity percolation threshold. This distinguishes the system's preferential flow region (in white) from the stagnant flow region (in gray)

Fig. 2 Probability distribution function (PDF) for the average channel velocity determined at the narrowest width of each channel in a subset of samples whose flow is characterized as highly uniform (red) or highly channelized (green). The line type indicates whether the distribution pertains to velocities in the preferential flow region (solid) or stagnant flow region (dashed)
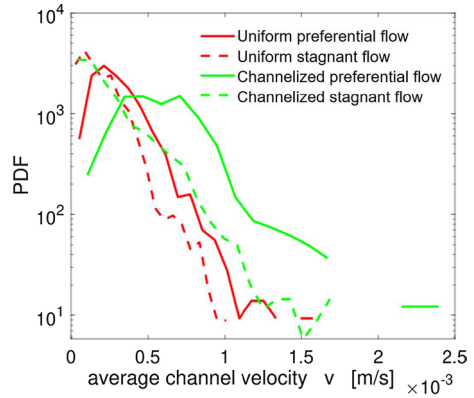

Fig. 3 (Top) Evolving weights for different structural attributes tested (individual attributes are shown in different colors) at the end of each differential evolution generation. (Bottom) Evolution of corresponding model performance parameters (AIC and mean accuracy, μ_a) at the end of each generation. Data shown are for a single iteration of the 5-fold cross-validation. Convergence is reached after 39 generations. Section S1 of the SI provides more details on the convergence criteria used
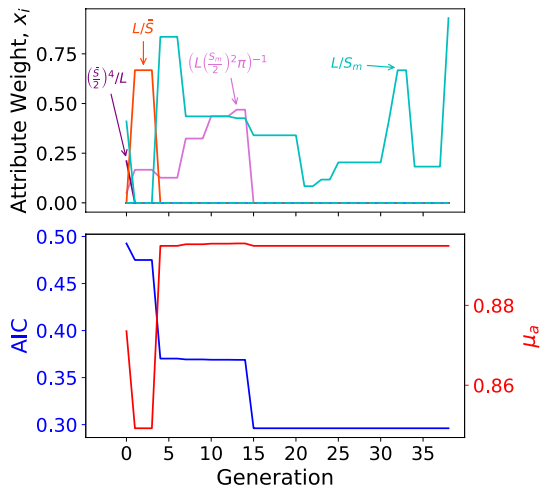
this example, the most fit solution from the initial population was $w_e = \sum_{i \in \#\mathcal{H}} x_i H_i$ with $\mathcal{H} = \{L/S_{\mathrm{m}}, \left(\frac{\bar{S}}{2}\right)^4/L, \left(L\left(\frac{S_{\mathrm{m}}}{2}\right)^2 \pi\right)^{-1}\}$, shown for generation 0. At each subsequent generation, the most fit solutions are recombined and mutated. As can be seen, within four generations, the algorithm determined that the contribution by the attributes $\{L/\bar{S}, \left(\frac{\bar{S}}{2}\right)^4/L\}$ did not produce a better result. That is, AIC decreased and μ_a increased, but not as much as a mutated solution with different coefficients [see Fig. 3 (Bottom)]. Likewise, various combinations of contributions by the characteristic $\{\left(L\left(\frac{S_{\mathrm{m}}}{2}\right)^2 \pi\right)^{-1}\}$ were explored, but removed from the working population by generation 15. It is worth noting that from generations 14–15, the evolved solution became simpler (depending on one attribute instead of two) at the minor expense of μ_a (loss of 6×10^{-4} accuracy). Altogether, this goodness of fit and penalty for over-parameterization yielded the lowest AIC, which is the objective function this approach aims to minimize.

A pictorial view of the shortest paths corresponding to potential solutions at select generations is shown in Fig. 4. Generations $g = \{0, 1, 4, 15\}$ are shown by different line types. Results are provided for three representative samples from the *uniform*, *intermediate* and *channelized* flow class. In these figures, the solid grains are shown in black,
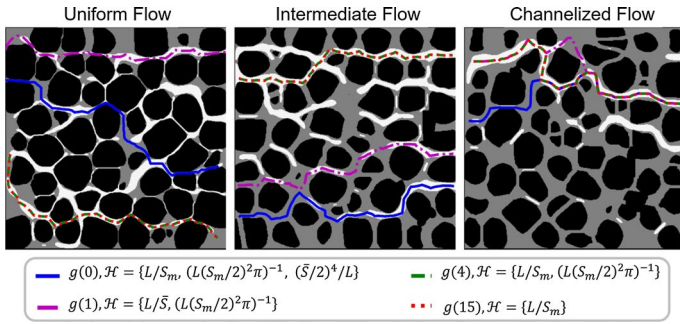
Fig. 4 Comparison of different weighted shortest paths in typical structures of uniform, intermediate and channelized flow class. Solid grains are in black, preferential flow regions are in white, and stagnant flow regions are in gray. The shortest paths shown correspond to the evolving weighing scheme at select generations of Fig. 3

the SFR in gray, and the PFR in white. Generally, the shortest paths become better aligned with the preferential flow region in all samples as generations increase. Note that, the predicted shortest path between generations 4 and 15 is the same for the samples shown, demonstrating that a simpler model performs equally well. We also assess the possible correlation between accuracy of the shortest path predicted and the strength of the flow channelization. A low anticorrelation (Spearman correlation coefficient of -0.06, see Figure S9) is found between a_z and p_c that verifies the unbiased ability to accurately predict the preferential flow path in samples of all channelized flow strength.

To assess the performance of our graph-based approach in identifying the minimum resistance path, we compare with a null unweighted model $w_e = 1$ for all edges. This comparison sheds light on flow-structure dependence on the pore geometry while preserving the underlying topology. Figure 5 illustrates the complementary cumulative distribution function (CCDF) of the accuracy a_z in each sample for the fittest weighting scheme based on L/S_m and the null model in each of the 5-fold cross-validation sets. From these distributions, it is evident that the accuracy of the optimized weighting scheme (yellow) significantly outperforms the null model (blue). Neither weighting scheme can capture the full percolating path for all samples considered in this study. However, the improved performance is notable at, e.g., CCDF $= 0.5$ where one finds that $a_z \gtrsim 0.6$ for the null model whereas $a_z \gtrsim 0.9$ for the most fit weighting scheme.

Fig. 5 Complementary cumulative distribution function (CCDF) of individual sample accuracy a_z from the shortest path analyses for graphs weighted by $w_e(\mathbf{x}) = \sum_{i \in \#\mathcal{H}} x_i H_i$ with $\mathcal{H} = \{L/S_m\}$ (the most fit solution) or $w_e = 1$ (the null model). Individual lines correspond to each of the 5-fold cross-validation sets
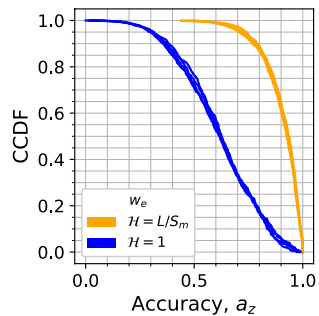
3.3 Inferring Velocity Distribution from Network Structure

To understand why flow becomes channelized, we evaluate the networks' edge assortativity in terms of their pore throat size, S_m. Assortative mixing ($r > 0$) is found in all samples' PFR and SFR subnetworks, as well as in their full pore network (see Fig. 6). For all samples, the contrast in assortativity between subnetworks corresponding to the PFR and the SFR is $r_1/r_2 > 1$. This corroborates our hypothesis that the underlying pore structure where preferential flow develops is structurally more ordered than that where flow is stagnant.

To evaluate how this contrast changes with flow channelization, we investigate the correlation of r_1/r_2 with p_c, as shown in Fig. 6. A significant positive correlation between these parameters is found (Spearman coefficient of 0.68). From these data, we can infer that samples of characteristically high p_c (with channelized flow) tend to exhibit large contrast between subnetwork assortativity ($r_1/r_2 \gg 1$). This suggests that channelized flow is constrained to edges that are structurally ordered (relative to disordered edges corresponding to stagnant flow). Consequently, flow paths would greatly differ between the two flow regions. In contrast, samples of characteristically low p_c (with uniform flow) tend to display small contrast between subnetwork assortativity ($r_1/r_2 \sim 1$). This implies that uniform flow is not structurally constrained given that edges in the two subnetworks are similarly ordered. Hence, flow paths in the two flow regions would bear high resemblance. This structural arrangement and flow channelization relationship is consistent with the velocity variability observed in PFR and SFR of uniform (red) vs channelized (green) flow fields (see Fig. 2) and sheds light on the relationship between static and dynamic connectivity properties.

The correlation between a global measures of static and dynamic connectivity, here taken as the assortativity of the full pore network, r_W, and p_c, respectively, is explored in Figure S10 (top). Their correlation is weak (Spearman coefficient of 0.32), which is insufficient to properly assess the flow heterogeneity from global structural characteristics. The correlation between global and local connectivity, r_W and r_1/r_2, respectively, is shown in Figure S10 (bottom), finding no significant correlation (Spearman coefficient of 0.05). This suggests that it is not possible to infer the distribution of flow from the structural order of the entire system.

Fig. 6 Scatter plot of the assortativity ratio, r_1/r_2, and percolation threshold, p_c, of individual samples, colored according to r_W value. A positive correlation between the terms is found, as indicated by the Spearman coefficient ρ of 0.68 and a p value of 0.00. The data suggest that the order between pore network regions associated with preferential vs stagnant flow grows in contrast for samples of increasing flow channelization
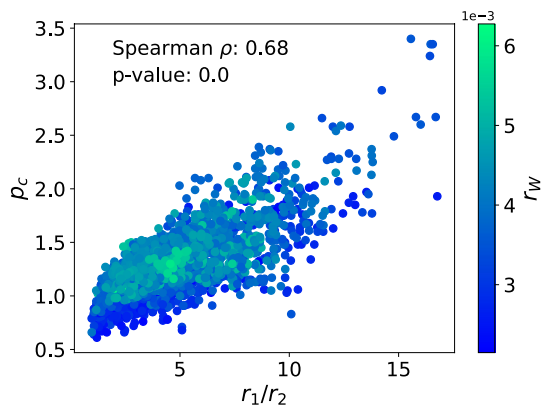
4 Conclusion

Resolving the full flow field in porous media is computationally expensive and requires detailed information of the pore structure. The precise link between pore-scale flow and the underlying structural attributes has extensively been studied but remains elusive. In this work, we presented and analyzed a set of methods designed to probe how the pore structure impacts flow channelization in two-dimensional pore networks. Diverse samples were used with developed flow ranging from very uniform to very channelized. Each sample was characterized by its flow properties under steady state and its pore-space recast as a graph with unique topology and channel geometry. We used these measurements to (1) quantify the spectrum of flow channeling occurring in each sample based on a percolation parameter, (2) define regions of the flow field that describe preferential and stagnant flow by velocity thresholds, (3) link the location of preferential pathways to structural properties of the graph through a biased shortest path analysis, and (4) explain the structural constraints for preferential and stagnant flow formation by evaluating the (dis)order in regions of the pore network. Specifically, graphs weighted by the channel's arc length to pore throat ratio (L/S_{m}) gave the most accurate representation of structural flow resistance when the path length was minimized. Likewise, the structural order of networks based on edge conductivity, S_{m}, delineated accurately where and how much contrast there is between preferential and stagnant flow regions. Based on this flow-structure relationship, we found greater order in the pore subnetwork associated with preferential flow compared to the subnetwork corresponding to the stagnant flow. This structural order contrast increased with flow channelization. Ascertaining how pore structural properties trigger preferential pathway formation, even in simple porous media like the ones here studied, is critical to characterizing many subsurface processes of critical importance at the pore and the field scale.

References

Abelin, H., Birgersson, L., Moreno, L., Widén, H., Ågren, T., Neretnieks, I.: A large-scale flow and tracer experiment in granite: 2. Results and interpretation. Water Resour. Res. **27**(12), 3119 (1991). https://doi.org/10.1029/91WR01404

Alim, K., Parsa, S., Weitz, D.A., Brenner, M.P.: Local pore size correlations determine flow distributions in porous media. Phys. Rev. Lett. **119**(14), 144501 (2017). https://doi.org/10.1103/PhysRevLett.119.144501

Ambegaokar, V., Halperin, B., Langer, J.: Hopping conductivity in disordered systems. Phys. Rev. B **4**(8), 2612 (1971). https://doi.org/10.1103/PhysRevB.4.2612

Arganda-Carreras, I., Fernández-González, R., Muñoz-Barrutia, A., Ortiz-De-Solorzano, C.: 3D reconstruction of histological sections: application to mammary gland tissue. Microsc. Res. Tech. **73**(11), 1019 (2010). https://doi.org/10.1002/jemt.20829

Bear, J.: Dynamics of Fluids in Porous Media. Courier Corporation, North Chelmsford (1988)

Bianchi, M., Zheng, C., Wilson, C., Tick, G.R., Liu, G., Gorelick, S.M.: Spatial connectivity in a highly heterogeneous aquifer: from cores to preferential flow paths. Water Resour. Res. (2011). https://doi.org/10.1029/2009WR008966

Bijeljic, B., Raeini, A., Mostaghimi, P., Blunt, M.J.: Predictions of non-Fickian solute transport in different classes of porous media using direct simulation on pore-scale images. Phys. Rev. E **87**(1), 013011 (2013). https://doi.org/10.1103/PhysRevE.87.013011

Bradford, S.A., Leij, F.J., Schijven, J., Torkzaban, S.: Critical role of preferential flow in field-scale pathogen transport and retention. Vadose Zone J. **16**(4), 1 (2017). https://doi.org/10.2136/vzj2016.12.0127

Carman, P.: Fluid flow through granular beds. Trans. Inst. Chem. Eng. **15**, S32 (1937). https://doi.org/10.1016/S0263-8762(97)80003-2

Carrel, M., Morales, V.L., Dentz, M., Derlon, N., Morgenroth, E., Holzner, M.: Pore-scale hydrodynamics in a progressively bioclogged three-dimensional porous medium: 3-D particle tracking experiments and stochastic transport modeling. Water Resour. Res. **54**(3), 2183 (2018). https://doi.org/10.1002/2017WR021726

Chakraborty, P., Das, B.S., Vasava, H.B., Panigrahi, N., Santra, P.: Spatial structure, parameter nonlinearity, and intelligent algorithms in constructing pedotransfer functions from large-scale soil legacy data. Sci. Rep. **10**(1), 15050 (2020). https://doi.org/10.1038/s41598-020-72018-2

Charbeneau, R.J.: Groundwater Hydraulics and Pollutant Transport, vol. OCLC, p. 1026182712. Waveland, Long Grove (2006)

COMSOL multiphysics ®. www.comsol.com

Datta, S.S., Chiang, H., Ramakrishnan, T.S., Weitz, D.A.: Spatial fluctuations of fluid velocities in flow through a three-dimensional porous medium. Phys. Rev. Lett. **111**(6), 064501 (2013). https://doi.org/10.1103/PhysRevLett.111.064501

de Anna, P., Quaife, B., Biros, G., Juanes, R.: Prediction of the low-velocity distribution from the pore structure in simple porous media. Phys. Rev. Fluids **2**(12), 124103 (2017). https://doi.org/10.1103/PhysRevFluids.2.124103

de Anna, P., Pahlavan, A.A., Yawata, Y., Stocker, R., Juanes, R.: Chemotaxis under flow disorder shapes microbial dispersion in porous media. Nat. Phys. **17**(1), 68 (2021). https://doi.org/10.1038/s41567-020-1002-x

Dentz, M., Le Borgne, T., Englert, A., Bijeljic, B.: Mixing, spreading and reaction in heterogeneous media: a brief review. J. Contam. Hydrol. **120–121**, 1 (2011). https://doi.org/10.1016/j.jconhyd.2010.05.002

Dijkstra, E.W.: A note on two problems in connexion with graphs. Numer. Math. **1**(1), 269 (1959). https://doi.org/10.1007/BF01386390

Dullien, F.A.L.: Porous Media: Fluid Transport and Pore Structure, 2nd edn. Academic Press, Cambridge (1992)

Ewing, R., Hunt, A.: Percolation Theory for Flow in Porous Media. Lecture Notes in Physics, vol. 771. Springer, Berlin Heidelberg (2009). https://doi.org/10.1007/978-3-540-89790-3

Fiori, A., Jankovic, I., Dagan, G.: The impact of local diffusion upon mass arrival of a passive solute in transport through three-dimensional highly heterogeneous aquifers. Adv. Water Resour. **34**(12), 1563 (2011). https://doi.org/10.1016/j.advwatres.2011.08.010

Fiori, A., Dagan, G., Jankovic, I., Zarlenga, A.: The plume spreading in the MADE transport experiment: Could it be predicted by stochastic models?: Made prediction. Water Resour. Res. **49**(5), 2497 (2013). https://doi.org/10.1002/wrcr.20128

Flury, M., Flühler, H., Jury, W.A., Leuenberger, J.: Susceptibility of soils to preferential flow of water: a field study. Water Resour. Res. **30**(7), 1945 (1994). https://doi.org/10.1029/94WR00871

Freeze, R.A., Cherry, J.A.: Groundwater. Prentice-Hall, Hoboken (1979)

Goeppert, N., Goldscheider, N., Berkowitz, B.: Experimental and modeling evidence of kilometer-scale anomalous tracer transport in an alpine karst aquifer. Water Res. **178**, 115755 (2020). https://doi.org/10.1016/j.watres.2020.115755

Guihéneuf, N., Boisson, A., Bour, O., Dewandel, B., Perrin, J., Dausse, A., Viossanges, M., Chandra, S., Ahmed, S., Maréchal, J.: Groundwater flows in weathered crystalline rocks: impact of piezometric variations and depth-dependent fracture connectivity. J. Hydrol. **511**, 320 (2014). https://doi.org/10.1016/j.jhydrol.2014.01.061

Hendrickx, J.M., Flury, M.: Conceptual Models of Flow and Transport in the Fractured Vadose Zone. National Research Council, The National Academies Press, Washington (2001)

Hobé, A., Vogler, D., Seybold, M.P., Ebigbo, A., Settgast, R.R., Saar, M.O.: Estimating fluid flow rates through fracture networks using combinatorial optimization. Adv. Water Resour. **122**, 85 (2018). https://doi.org/10.1016/j.advwatres.2018.10.002

Holzner, M., Morales, V.L., Willmann, M., Dentz, M.: Intermittent Lagrangian velocities and accelerations in three-dimensional porous medium flow. Phys. Rev. E **92**(1), 013015 (2015). https://doi.org/10.1103/PhysRevE.92.013015

Hyman, J.D.: Flow channeling in fracture networks: characterizing the effect of density on preferential flow path formation. Water Resour. Res. (2020). https://doi.org/10.1029/2020WR027986

Jimenez-Martinez, J., Negre, C.F.A.: Eigenvector centrality for geometric and topological characterization of porous media. Phys. Rev. E **96**(1), 013310 (2017). https://doi.org/10.1103/PhysRevE.96.013310

Jiménez-Martínez, J., Le Borgne, T., Tabuteau, H., Méheust, Y.: Impact of saturation on dispersion and mixing in porous media: photobleaching pulse injection experiments and shear-enhanced mixing model. Water Resour. Res. **53**(2), 1457 (2017)

Kang, P.K., de Anna, P., Nunes, J.P., Bijeljic, B., Blunt, M.J., Juanes, R.: Pore-scale intermittent velocity structure underpinning anomalous transport through 3-D porous media. Geophys. Res. Lett. **41**(17), 6184 (2014). https://doi.org/10.1002/2014GL061475

Kirkpatrick, S.: Classical transport in disordered media: scaling and effective-medium theories. Phys. Rev. Lett. **27**(25), 1722 (1971). https://doi.org/10.1103/PhysRevLett.27.1722

Knudby, C., Carrera, J.: On the relationship between indicators of geostatistical, flow and transport connectivity. Adv. Water Resour. **28**(4), 405 (2005). https://doi.org/10.1016/j.advwatres.2004.09.001

Kurotori, T., Zahasky, C., Hosseinzadeh Hejazi, S.A., Shah, S.M., Benson, S.M., Pini, R.: Measuring, imaging and modelling solute transport in a microporous limestone. Chem. Eng. Sci. **196**, 366 (2019). https://doi.org/10.1016/j.ces.2018.11.001

Le Borgne, T., Bour, O., Paillet, F., Caudal, J.P.: Assessment of preferential flow path connectivity and hydraulic properties at single-borehole and cross-borehole scales in a fractured aquifer. J. Hydrol. **328**(1), 347 (2006). https://doi.org/10.1016/j.jhydrol.2005.12.029

LeBlanc, D.R., Garabedian, S.P., Hess, K.M., Gelhar, L.W., Quadri, R.D., Stollenwerk, K.G., Wood, W.W.: Large-scale natural gradient tracer test in sand and gravel, Cape Cod, Massachusetts: 1. Experimental design and observed tracer movement. Water Resour. Res. **27**(5), 895 (1991). https://doi.org/10.1029/91WR00241

Matyka, M., GoÃ â€šembiewski, J., Koza, Z.: Power-exponential velocity distributions in disordered porous media. Phys. Rev. E **93**(1), 013110 (2016). https://doi.org/10.1103/PhysRevE.93.013110

Meyer, D.W., Bijeljic, B.: Pore-scale dispersion: bridging the gap between microscopic pore structure and the emerging macroscopic transport behavior. Phys. Rev. E **94**(1), 013107 (2016). https://doi.org/10.1103/PhysRevE.94.013107

Middleton, R.S., Carey, J.W., Currier, R.P., Hyman, J.D., Kang, Q., Karra, S., Jiménez-Martínez, J., Porter, M.L., Viswanathan, H.S.: Shale gas and non-aqueous fracturing fluids: opportunities and challenges for supercritical CO2. Appl. Energy **147**, 500 (2015). https://doi.org/10.1016/j.apenergy.2015.03.023

Morales, V.L., Dentz, M., Willmann, M., Holzner, M.: Stochastic dynamics of intermittent pore-scale particle motion in three-dimensional porous media: experiments and theory: particle motion dynamics in porous media. Geophys. Res. Lett. **44**(18), 9361 (2017). https://doi.org/10.1002/2017GL074326

Newman, M.E.J.: Networks, 2nd edn. Oxford University Press, Oxford (2018)

Nissan, A., Berkowitz, B.: Anomalous transport dependence on Péclet number, porous medium heterogeneity, and a temporally varying velocity field. Phys. Rev. E **99**(3), 033108 (2019). https://doi.org/10.1103/PhysRevE.99.033108

Orr, F.M., Taber, J.J.: Use of carbon dioxide in enhanced oil recovery. Science **224**(4649), 563 (1984). https://doi.org/10.1126/science.224.4649.563

Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. **9**(1), 62 (1979). https://doi.org/10.1109/TSMC.1979.4310076

Pérez-Reche, F.J., Taraskin, S.N., Otten, W., Viana, M.P., Costa, L.D.F., Gilligan, C.A.: Prominent effect of soil network heterogeneity on microbial invasion. Phys. Rev. Lett. **109**(9), 098102 (2012). https://doi.org/10.1103/PhysRevLett.109.098102

Price, W.G., Potter, A., Thomson, T.K., Smith, G.E.P., Hazen, A., Beardsley, R.C.: Discussion on dams on sand foundations. Trans. Am. Soc. Civ. Eng. **73**(3), 190 (1911). https://doi.org/10.1061/TACEAT.0002320

Puyguiraud, A., Gouze, P., Dentz, M.: Stochastic dynamics of Lagrangian pore-scale velocities in three-dimensional porous media. Water Resour. Res. **55**(2), 1196 (2019). https://doi.org/10.1029/2018WR023702

Rasmuson, A., Neretnieks, I.: Radionuclide transport in fast channels in crystalline rock. Water Resour. Res. **22**(8), 1247 (1986). https://doi.org/10.1029/WR022i008p01247

Renard, P., Allard, D.: Connectivity metrics for subsurface flow and transport. Adv. Water Resour. **51**, 168 (2013). https://doi.org/10.1016/j.advwatres.2011.12.001

Rizzo, C.B., de Barros, F.P.J.: Minimum hydraulic resistance and least resistance path in heterogeneous porous media: minimum hydraulic resistance. Water Resour. Res. **53**(10), 8596 (2017). https://doi.org/10.1002/2017WR020418

Schaap, M.G., Leij, F.J., van Genuchten, M.T.: rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. J. Hydrol. **251**(3), 163 (2001). https://doi.org/10.1016/S0022-1694(01)00466-8

Siena, M., Riva, M., Hyman, J.D., Winter, C.L., Guadagnini, A.: Relationship between pore size and velocity probability distributions in stochastically generated porous media. Phys. Rev. E **89**(1), 013018 (2014). https://doi.org/10.1103/PhysRevE.89.013018

Siena, M., Iliev, O., Prill, T., Riva, M., Guadagnini, A.: Identification of channeling in pore-scale flows. Geophys. Res. Lett. **46**(6), 3270 (2019). https://doi.org/10.1029/2018GL081697

Šimnek, J., Jarvis, N.J., Van Genuchten, M.T., Gärdenäs, A.: Review and comparison of models for describing non-equilibrium and preferential flow and transport in the vadose zone. J. Hydrol. **272**(1–4), 14 (2003). https://doi.org/10.1016/S0022-1694(02)00252-4

Storn, R., Price, K.: Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. J. Glob. Optim. **11**, 341 (1997). https://doi.org/10.1023/A:1008202821328

Su, G.W., Geller, J.T., Pruess, K., Hunt, J.R.: Solute transport along preferential flow paths in unsaturated fractures. Water Resour. Res. **37**(10), 2481 (2001). https://doi.org/10.1029/2000WR000093

Tibshirani, R., Walther, G., Hastie, T.: Estimating the number of clusters in a data set via the gap statistic. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **63**(2), 411 (2001). https://doi.org/10.1111/1467-9868.00293

Tyukhova, A.R., Willmann, M.: Conservative transport upscaling based on information of connectivity: transport upscaling using connectivity. Water Resour. Res. **52**(9), 6867 (2016). https://doi.org/10.1002/2015WR018331

Wilmott, P.: Machine Learning: An Applied Mathematics Introduction, 1st edn. Panda Ohana Publishing, London (2019)

Winograd, I.J., Pearson, F.J.: Major carbon 14 anomaly in a regional carbonate aquifer: possible evidence for megascale channeling, South Central Great Basin. Water Resour. Res. **12**(6), 1125 (1976). https://doi.org/10.1029/WR012i006p01125

Wong, P.: The statistical physics of sedimentary rock. Phys. Today **41**(12), 24 (1988). https://doi.org/10.1063/1.881138

Yeates, C., Youssef, S., Lorenceau, E.: Accessing preferential foam flow paths in 2D micromodel using a graph-based 2-parameter model. Transp. Porous Med. **133**(1), 23 (2020). https://doi.org/10.1007/s11242-020-01411-2

Zheng, C., Bianchi, M., Gorelick, S.M.: Lessons learned from 25 years of research at the MADE site. Groundwater **49**(5), 649 (2011). https://doi.org/10.1111/j.1745-6584.2010.00753.x

Zinn, B., Harvey, C.F.: When good statistical models of aquifer heterogeneity go bad: a comparison of flow, dispersion, and mass transfer in connected and multivariate Gaussian hydraulic conductivity fields. Water Resour. Res. (2003). https://doi.org/10.1029/2001WR001146

Authors and Affiliations

Z. Kanavas[1] · F. J. Pérez-Reche[2] · F. Arns[3] · V. L. Morales[1]

[1] Department of Civil and Environmental Engineering, University of California, Davis, USA

[2] Institute for Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen, UK

[3] Institute of Environmental Engineering, ETH Zürich, Zurich, Switzerland