Permeability prediction for natural porous rocks through feature selection and machine learning

Jinlong Fu, Hywel R. Thomas, and Chenfeng Li

Abstract

The relationships between macroscopic properties and microstructural characteristics are of great significance for natural porous rocks, based on which transport properties can be directly predicted from measurable pore microstructures. However, the explicit establishment of such microstructure-property mappings appears to be difficult, due to the intricacy, stochasticity and heterogeneity of pore network systems. In this paper, a data-driven framework is developed to excavate the inherent microstructure-permeability linkage, where multiple techniques are integrated together, including stochastic reconstruction, microstructure characterization, pore-scale simulation, feature selection and machine learning. Large numbers of digital rock samples are generated from micro-CT imaging and the stochastic reconstruction algorithm. The pore microstructures are quantitatively characterized by using various morphological descriptors. High-fidelity lattice Boltzmann simulations are conducted to evaluate the permeability values of these samples. The optimal set of morphological descriptors that best represents permeability is identified through a performance-oriented feature selection process, and then a machine learning-based surrogate model is constructed to implicitly link permeability with microstructural features. This model can efficiently predict permeability with a broad range of values, and it exhibits great superiority over commonly used empirical/analytical relations in terms of predictive accuracy, general applicability and data-driven evolvability. Besides, deep insights into the underlying microstructure-permeability correlation can also be obtained from the feature selection results, due to the good interpretability of morphological descriptors.

Keywords: Porous rocks; Permeability prediction; Microstructure characterization; Pore-scale simulation; Feature selection; Machine learning.

Introduction

Permeability quantifies the ability of a porous medium to transmit fluid (Bear, 2013), which is a fundamental parameter to understanding the transport behaviors of fluid flows. And it also plays a critical role in many geological applications, including oil and gas recovery, geothermal energy exploitation, CO₂ underground storage, radioactive waste disposal and contaminant hydrogeology. The permeable pore spaces within different geologic materials are often highly distinctive, leading to an extremely broad range of permeability values varying up to 13 orders of magnitude (Luhmann et al., 2017). Essentially, the macroscopic physical properties of porous media strongly depend on the microstructural characteristics. And the structure-property relationship is one of the most elementary issues in porous media research, with the expectation of predicting hydraulic, mechanical, electrical and thermal properties from the measurable microstructural informatics. However, the intricacy, stochasticity and heterogeneity inherent in natural porous rocks make it difficult to accurately and rapidly evaluate permeability, especially for tight rocks with low porosity. Therefore, a deep insight into the dependence of permeability
on microstructural characteristics is always being pursued, aiming to achieve a reliable and efficient method for permeability prediction (Blair et al., 1996; Coker et al., 1996; Xu and Yu, 2008; Berg, 2014; Fu et al., 2020a).

Laboratory measurement is the routine way to determine permeability, where fluid flow is driven by a constant pressure difference to pass through a rock core and permeability is then evaluated according to Darcy’s law when the fluid flow reaches the steady state (Bear, 2013). High cost and long waiting time are the main limitations of experimental measurement in practical application, especially for tight rocks. Besides, great efforts have been made to incorporate permeability using analytical or empirical models, such as the famous semi-empirical Kozeny-Carman relation (Fu et al., 2020b) and many variants deriving from it (Berryman and Blair, 1987; Costa, 2006; Berg, 2014; Wei et al., 2018b). Generally, these models rely on specific microstructural characteristics of porous media, such as porosity, specific surface area, tortuosity, characteristic length, pore size, constriction factor, fractal dimension and others. Despite the simplicity and convenience in actual practice, analytical models are often highly idealized and empirical models usually contain adjustable parameters of uncertainty, which make them only appropriate and reliable for specific or ideal pore microstructures and often break down for natural rocks with complicated pore networks.

Recently, the digital rock physics (DRP) technique has been rapidly developed to be an alternative to laboratory measurement, which is convenient and reliable to characterize microstructural attributes and compute petrophysical properties (Andrä et al., 2013; Fu et al., 2020a). The DRP technique applies advanced microscopy imaging facility (Blunt et al., 2013; Anovitz and Cole, 2015), such as X-ray micro-computed tomography (micro-CT) and focused ion beam scanning electron microscopy (FIB-SEM), to obtain 3D visualization of pore microstructures at the microscale, which provides a new routine to explore microscopic physics of transport phenomena. After the image acquisition, high-fidelity numerical simulations of transport processes can be performed on the digital microstructures to evaluate corresponding transport properties or explore specific physical phenomena (Andrä et al., 2013; Blunt, 2017).

Pore-network modeling (PNM) and direct numerical simulation (DNS) are the two primary pore-scale computing approaches to mimicking transport processes happening inside porous rocks. PNM simplifies the complicated pore space into a topologically representative network of pore bodies interconnected by pore throats with ideal shapes (such as sphere and cylinder), according to specific criteria (Xiong et al., 2016; Blunt, 2017). The transport behaviors within each network element are described by semi-analytical laws (such as Hagen-Poiseuille law), which makes the computation considerably fast and enables multi-scale modeling to incorporate strong heterogeneity in large volumes. PNM is naturally appropriate for capillary-controlled transport processes. However, the transport property results estimated from PNM are often less accurate, due to the simplification of the complicated pore space. It is still of great challenge to gain reliable results from PNM by identifying the microstructural features relevant to network modeling and ignoring the remaining parts to simplify the computational complexity (Xiong et al., 2016).

On the contrary, DNS directly discretizes the raw pore space into computing elements by preserving pore geometry (voxels can be used as the computing elements), and then transport equations (such as Navier-Stokes or Laplace equations) are numerically solved or approximated on the computational meshes (Andrä et al., 2013; Blunt et al., 2013). Lattice Boltzmann method (LBM), finite element method (FEM) and finite volume method (FVM) are commonly used to approximate or solve transport equations at the pore scale. Generally, DNS can provide direct insight into the impact of pore microstructure on transport properties, but it has severe limitations in computational intensity. The 3D digital microstructure with large representative size and high resolution usually contains hundreds of millions, even billions of computational elements (or voxels). As a consequence, massively parallel programming, long computing time, high-performance computing (HPC) platform and large data storage are usually required to run such hug numerical simulations (Liu et al., 2016; Saxena et al., 2017). The compute-intensive nature of DNS makes it difficult to accommodate all the details of pore microstructures and involve all the relevant transport physics.

Generally, it is a brute-force way to explore the relationships between transport properties and microstructural characteristics by solely using pore-scale simulations. Currently, many attempts have been made to built the surrogate microstructure-property models through artificial intelligence, in order to rapidly and accurately predict macroscopic properties from the measurable microstructural informatics. Due to the powerful capacities in massive data analysis and hidden rule exploration, machine/deep learning algorithms are becoming more and
more popular in this field, especially the convolutional neural network (CNN). Typically, CNN (Agrawal and Choudhary, 2019) is capable of automatically extracting task-related features from spatial data such as images through its convolution layers, avoiding the manual feature selection procedure, and it has achieved tremendous success in the computer vision filed. Therefore, many similar studies have been conducted to construct CNN-based surrogate models for permeability prediction, where the 2D or 3D digital images of porous microstructures are directly used as the input data (Srisutthiyakorn, 2016; Wu et al., 2018; Sudakov et al., 2019; Kamrava et al., 2020; Tian et al., 2020a; Hong and Liu, 2020; Tsembely et al., 2020). Besides, CNN has also been applied to establish the linkages between microstructures and other macroscopic properties/behaviors for various heterogeneous materials, including effective thermal conductivity (Wei et al., 2018a), effective elastic moduli (Cecen et al., 2018; Li et al., 2019), effective diffusivity (Wu et al., 2019), P/S-wave velocity (Karimpouli and Tahmasebi, 2019), formation factor (Rabbani et al., 2020) and fluid velocity filed (Santos et al., 2020).

However, several critical issues and challenges have not been properly solved for the CNN-based surrogate modeling of physical properties in porous media research. (1) Dimensional homogeneity problem. In some studies (Li et al., 2019; Kamrava et al., 2020; Tian et al., 2020a), the “linkages” between digital images and physical-property variables are directly built through CNN without nondimensionalization process, which is contrary to the principle of dimensional homogeneity. (2) Computational problem. High computational complexity and excessive memory requirement are the inherent defects of the 3D CNN, which strictly limit both the quality and quantity of 3D training images. However, a reliable prediction model needs a large amount of representative elementary volumes (REVs) with high quality to feed it, and such high-intensive computation usually demands HPC platform. (3) Feature extraction problem. Kernels (convolving windows) are applied across the input image to extract local features, but the internal connection of components as well as the relative spatial relationships are not captured by the convolution layers of CNN. It means that the global features of porous media (such as long-distance connectivity and topological information) that are crucial to transport properties are rarely used. (4) Lack of spatial invariance. The internal representation of a pore microstructure in CNN is not independent of view angle, which means rotation of input image can affect the prediction result. This issue can be solved through data augmentation, but the computational expense of DNN model training will be dramatically increased. (5) Overfitting problem. The CNN model is easily over-fitted due to a large database for training. (6) Low-level interpretability. The complicated CNN architecture, formed by a deep stack of distinct layers, is often referred to as a “black box”, because it is difficult to understand the underlying mechanics and no inherent way exists to interpret how features influence a particular prediction. (7) Lack of flexibility. Once a CNN model is fitted for physical property prediction, both the size and resolution of input images are fixed, which is very inflexible for the common cases where adjustments of image size or resolution have to be made without losing information.

According to the discussion above, it is clear that the poor explainability of CNN doesn’t contribute to a good understanding of the microstructure-property linkages. In contrast, simple regression algorithms, including linear regression, decision tree, random forest, support vector machine and shallow neural network, are much easier to be interpreted, which can be used to reveal the underlying mechanisms of the microstructure-property linkages. Additionally, the manual feature-extraction free of CNN doesn’t constitute a comparative advantage over the other machine learning algorithms that require predefined feature variables, especially in porous media research. Because various morphological descriptors that statistically characterize pore microstructures have already been properly designed. Compared with the unreadable features extracted by CNN, morphological descriptors characterize porous microstructures from multiple perspectives with clear physical meanings, and these descriptors can compose the feature pool to construct the surrogate microstructure-property relationships through simple regression algorithms. Feature selection can be conducted to identify the morphological descriptors that are significant to permeability and remove the abundant and irrelevant ones, though which the microstructural complexity is reduced to a limited number of descriptive parameters related to permeability, and then a feature-effective prediction model with excellent performance can be achieved, as illustrated in Figure 1. What’s more important is that the dependence of permeability on microstructural characteristics can also be well interpreted through the feature selection process, providing a deep insight into the implicit microstructure-permeability linkage.
Figure 1: The data-driven framework to investigate the dependence of permeability on microstructural characteristics for natural porous rock, and it primarily contains six modules: (1) Digital rock acquisition, (2) Stochastic microstructure reconstruction, (3) Microstructural characterization, (4) Feature selection, (5) Pore-scale simulation (LBM), and (6) Model construction.

Although simple supervised machine learning algorithms were adopted to model physical properties of porous media in previous studies (van der Linden et al., 2016; Wang et al., 2019; Erofeev et al., 2019; Tian
et al., 2020b; Röding et al., 2020), they didn’t achieve the purpose to yield insightful understandings of the linkages between macroscopic properties and microstructural characteristics, and corresponding pore-scale behaviors are still poorly understood. This study distinguishes itself from previous efforts in five aspects: (1) Plenty of 3D digital rocks with diverse morphologies were acquired from micro-CT at high-resolution level, which are used to construct the predictive model to ensure its generalization ability. (2) A large number of 3D microstructure samples are stochastically reconstructed by preserving statistical equivalence, morphological similarity and transport properties, which are used as the raw data to capture the stochasticity in permeability modeling. (3) A wide variety of morphological descriptors are collected through an extensive literature survey for the purpose of comprehensive characterization of pore microstructures. (4) High-fidelity simulations of pore-scale fluid flow passing through the representative elementary volumes (REV) of digital microstructures are performed on the HPC platform, to reliable permeability values. (5) Feature selection is conducted to gain the optimal feature set for predictive model construction and enhance the model interpretability.

In this work, the dependence of permeability on microstructural characteristics of natural porous rocks is fundamentally investigated through feature selection and machine learning (shallow neural network). Various morphological descriptors are collected from an extensive literature survey, and they quantitatively characterize pore microstructures from global, local, geometrical and topological perspectives. These morphological descriptors compose the raw feature pool, from which microstructural characteristics that are significant to permeability are recognized and picked out through a performance-oriented process of feature selection.

The optimal set of morphological descriptors (selected features) contains absolute and connected porosity, specific surface area, geometrical tortuosity, characteristic length, constriction factor, average pore throat size and hydraulic pore diameter. These selected microstructural features not only optimally characterize permeability, but also provide deep insights into the underlying microstructure-permeability relationship, which can be interpreted as follows: (1) Absolute and connected porosity represent the entire void fraction and the permeable portion conducting fluid flow respectively; (2) Specific surface area approximately reflects the fluid-solid interface where adhesive friction is applied to fluid flow; (3) Characteristic length, pore throat size and hydraulic pore diameter are indicators with different emphases to represent the average distance between opposite solid walls, where fluid flow is constricted between them; (4) Geometrical tortuosity quantifies the prolongation degree of streamlines due to the sinuosity of percolating pore path; (5) Constriction factor measures the cross-section fluctuation along pore channels that leads to divergence and convergence of streamlines at the expansion and shrinkage positions respectively.

Based on the optimal set of selected features, the machine learning-based surrogate model is constructed to accurately and efficiently predict permeability for natural porous rocks. This model possesses an excellent generalization performance in predicting permeability value spanning 4 orders of magnitude with the average error less than 6.92%. Compared with commonly used empirical/analytical relations, the machine learning-based prediction models exhibit great superiorities in terms of predictive accuracy and general applicability. Basically, this study provides a novel data-driven framework to fundamentally understand the microstructure-property linkages hidden in porous media, and it can be straightforwardly applied to study different macroscopic properties (such as effective diffusivity, thermal conductivity, formation factor and effective elastic moduli) by involving significant microstructural informatics, although only permeability is covered in this paper.

References


