

Permeability Prediction using multivariant structural regression

Matthew Andrew^{1,*}

¹Carl Zeiss Microscopy, 4385 Hopyard Rd., Pleasanton, CA, 94588, USA

Abstract. A novel method for permeability prediction is presented using multivariant structural regression. A machine learning based model is trained using a large number (2,190, extrapolated to 219,000) of synthetic datasets constructed using a variety of object-based techniques. Permeability, calculated on each of these networks using traditional digital rock approaches, was used as a target function for a multivariant description of the pore network structure, created from the statistics of a discrete description of grains, pores and throats, generated through image analysis. A regression model was created using an Extra-Trees method with an error of <4% on the target set. This model was then validated using a composite series of data created both from proprietary datasets of carbonate and sandstone samples and open source data available from the Digital Rocks Portal (www.digitalrocksporta.org) with a Root Mean Square Fractional Error of <25%. Such an approach has wide applicability to problems of heterogeneity and scale in pore scale analysis of porous media, particularly as it has the potential of being applicable on 2D as well as 3D data.

1 Introduction

Flow and transport in porous media are fundamentally rooted at the scale of the tiny tortuous pore pathways through which the flow takes place. As such, pore scale investigation is now a widely adopted tool across a range of disciplines associated with the examination of flow and transport. Such an approach has applications ranging from understanding the flow properties of porous ceramics (used as a catalytic substrate for vehicle emission reduction) [1], [2], examining battery electrolyte exchange [3], [4], characterizing geological formations for the purpose of understanding groundwater flow [5], carbon capture and storage [6]–[8] and (in its most industrially applied application) oil and gas recovery [9]–[12]. One of the most useful results from a holistic pore-scale characterization is the prediction of effective medium properties, such as permeability [13], relative permeability [14], capillary pressure curves [15], and effective acoustic and electrical properties [16], [17]. Traditional approaches to the prediction of effective properties from porous media (e.g. permeability, diffusivity or effective conductivity) focus on coupling the 3D structural imaging of porous media with the full physical simulation of the partial differential equations governing the property of interest [13], [18]. Once the domain and physics have been defined (with appropriate boundary conditions), the physics then converges over multiple iterations. Such an approach is contrasted to more traditional (legacy) approaches for effective property estimation, such as Kozeny-Carman or Kuwabara [19], [20] techniques which feed a relatively limited and difficult to measure set of structural properties into quasi-analytical models to make flow estimations. These are typically inaccurate (e.g. [21]), may require

significant (and arbitrary) correction factors and are challenging to effectively apply to real problems [22]. For example, different variants of the Kozeny-Carman equation require the knowledge of characteristic particle size, tortuosity or effective channel diameter – something typically ambiguous to measure in a real system of arbitrary pore network type.

The past 10 years have seen an explosion in the availability of open source machine learning and computer vision tools, enabling both in-depth multivariant classification and analysis of porous media structure [23], [24] and powerful new regression techniques, facilitating the usage of such multivariant statistics for effective prediction [25]. The application of these tools and techniques to flow and transport in porous media has been limited. In this study these techniques are used to present a new technique for prediction of Stokes flow (viscous dominated) permeability from pore-scale images of rocks using multivariant statistical regression coupled with extensive synthetic geometry generation and pore-scale computational simulation. Pore structures were parameterized using 34 distinct statistical parameters, derived from specific discretized structural properties. This multivariant model is validated on an extensive open source digital rock dataset, comprising carbonate and sandstone pore geometries, giving an average prediction error of around 25% for predicted permeability values ranging over 6 orders of magnitude (10 μ D to 10D).

2 Materials and Methods

One of the biggest challenges in the application of machine learning based computer vision techniques to

* Corresponding author: matthew.andrew@zeiss.com

digital rock analysis is that of data. Pore networks are complex, and the range of tools available for characterizing them is broad, so any multivariant characterization and prediction will require many datasets (potentially many thousands or millions) to effectively train while avoiding model overfitting, a difficult feat when each 3D image may take several hours to acquire. To overcome this obstacle, synthetic pore network generation techniques were used to generate a more extensive training set than would be available for real (imaged) structures. Early digital rock studies used synthetic pore space generation extensively to examine simple rock systems at the pore scale. While such an approach has been largely supplanted by direct imaging, synthetic techniques do present specific advantages, especially when examining mechanisms behind various processes while controlling the amount of heterogeneity [26], [27], or when trying to develop an holistic understanding of systems which are extremely challenging to characterize experimentally, such as the organic hosted pore networks present in unconventional shale reservoirs [28]. Synthetic pore spaces can either be constructed physically, usually by glass beads or etchings in glass (e.g. [29], [30]) or numerically using a pore space generator, using stochastic or object-based techniques, subject to various constraints (e.g. [31]–[33]).

Two thousand, one hundred and ninety (2,190) synthetic intergranular pore networks were generated by modelling grains as convex polyhedral [34], randomly placed within the pore space under a range of packing conditions and statistical descriptions. Polyhedra were defined by picking vertices around a bounding ellipse, defined by 3 axes (L_1 , L_2 & L_3 , such that $L_1 < L_2 < L_3$). All geometries were given a nominal voxel size of $1\mu\text{m}$ – a single nominal voxel size was used as the resulting permeability simulation results were then scaled across multiple nominal voxel sizes using the Hagen-Poiseuille equation as described below.

To allow for a range of granular aspect ratios the axis lengths for each grain were picked from a normal statistical distribution, each defined by an average and standard deviation (μ_1 , μ_2 , μ_3 and σ_1 , σ_2 , σ_3). For simplicity the same distribution was used for the two shortest axes ($\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$). The ratio $L_3/L_{1,2}$ (defining average grain aspect ratio) was randomly defined at the beginning, being a uniformly distributed random variable between 1 and 3. Similarly target porosity (ϕ) was a uniformly distributed random variable in the range $0.05 < \phi < 0.7$, as were $\mu_{1,2}$ (in the range 15-60 voxels). σ_1 and σ_2 were determined using a coupled random variable in the range 5-35 voxels, and σ_3 an independently determined variable in the same range. Multiple geometric packing strategies were used, including (in order of increasing computational expense) allowing grains to overlap [28], removing grain overlap and the explicit physical modelling of gravitational settling. The entire range of target porosities could be effectively modelled when allowing grain overlap; however when using the other techniques (which achieved higher grain packing densities than when

allowing overlap) once the maximum packing density was reached grain networks were eroded and diluted to extend the networks over a wider array of porosities. Computationally cheaper techniques allowed for more complex statistical descriptions (along with larger total geometries) to be used, including multi size component grain distributions. A set of various different synthetic images are shown in figure 1.

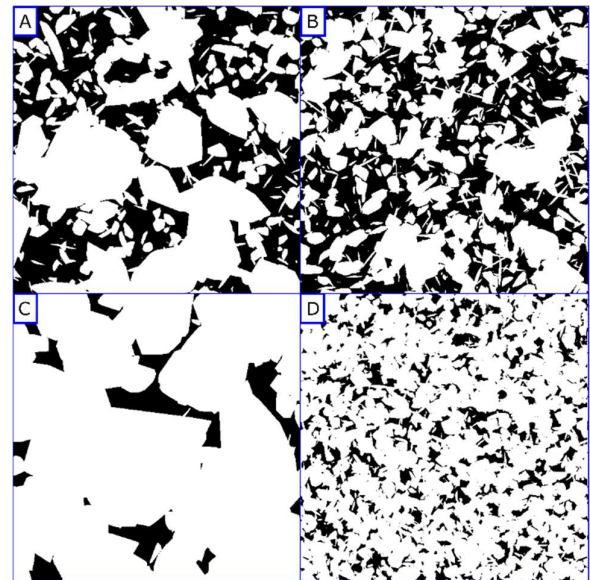


Figure 1 - A) A dual grain size system, with one grain population having a largest axis of $25\mu\text{m}$ and one population having a largest axis of $43\mu\text{m}$. B) A dual grain size system, with one grain population having a largest axis of $22\mu\text{m}$ and one population having a largest axis of $35\mu\text{m}$. C) A dual grain size system, with one grain population having a largest axis of $56\mu\text{m}$ and one population having a largest axis of $86\mu\text{m}$. D) A single grain size system with a largest axis length of $19\mu\text{m}$.

Structural statistics were calculated on a range of 2D slices through each 3D volume. It was decided to make measurements in 2D rather than on the full 3D network as it was more computationally tractable (and trivially parallelizable) and allows for the same model to be applied on 3D networks (acquired with XRM and FIB-SEM) and 2D structures (such as those characterized using light or electron microscopy). The application of these techniques to 2D systems is an interesting target of future work, but beyond the scope of this paper. Inherent biases and errors associated with 2D rather than 3D measurements are inherently accounted for in the relative weightings of different measurements during statistical multivariant regression. Grains and pores were then separated using a watershed algorithm operating across an Euclidian distance map [35], [36], with object seeds filtered using the “h-maxima” algorithm (figure 2). While such a technique has its limitations when dealing with complex network geometries [37], relatively simple networks can be effectively characterized [38]. Perhaps more important was that a perfect granular separation was not required, just that any object separation errors were

consistent between training and validation series of data. For example, differential errors in grain separation for systems with large grains (relative to the voxel size) relative to systems with smaller grains (relative to the voxel size) are implicitly accounted for by the large array of different grain sizes in the training set. The same errors will be present in both the training and the final (imaged) datasets, and therefore the impact of any errors on prediction will be minimized.

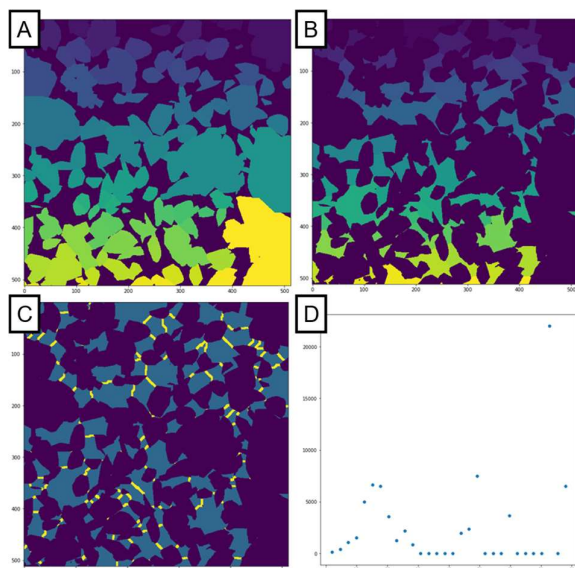


Figure 2 - Feature vector extraction. A) Grain separation. B) Pore separation. C) Throat identification (throats shown in yellow). D) Grain size distribution, showing a multimodal population distribution.

Once grains and pores were separated, population statistics were computed on features from these different populations, including grain, pore and pore throat sizes, shape averages, distributions, and network connectivity. This created a large (34 component) feature vector set for each network. The range of structural feature vectors (as well as their relative importance is shown in figure 3). The relative feature importance is defined as fraction of the decision tree associated with the permeability prediction decision tree [39], or the number of nodes in the decision graph associated with that feature. The 21 most important features had feature importances larger than 10% of the most important feature vector in the set (the standard deviation in the inscribed throat radius). Such a broad distribution in importances is an indication of both the complexity of the problem and the power of such a multivariate approach – a large number of different structural metrics are require to achieve a high quality prediction. An extensive evaluation of specific feature impacts in such a complex multivariate model, as well as an examination of the impact of changing numbers of input parameters and parameter selection on model performance would be an interesting target for future work. The requirement of the use of a large number of network characteristics is unsurprising, particularly given the requirement to make predictions using only features extracted in 2D. The goal for this approach is ultimately

to generate a model which can be applied on 2D data (for example light or electron microscopy), and a greater exploration of this is the target for future work. It is likely that, should features be extracted in 3D, a more accurate prediction could be made, potentially with a smaller number of extractable features, but this model would not have the potential of being (ultimately) transferred to 2D analytical techniques.

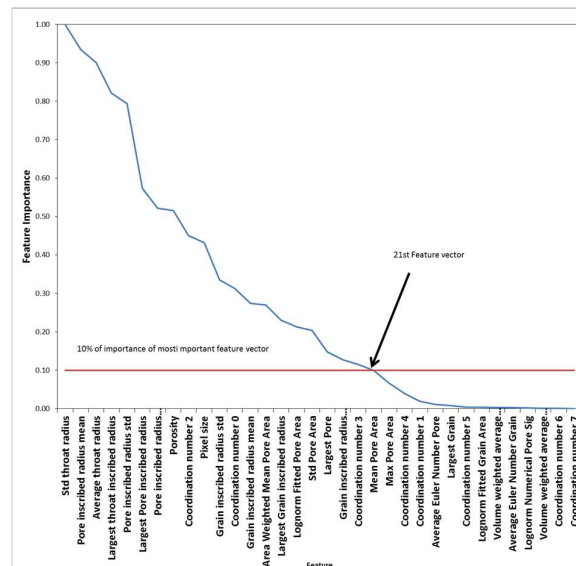


Figure 3 - distribution of feature vector importances.

It is interesting that the most important parameter (the parameter which is responsible for the greatest number of nodes in the forest of decision trees) is that of the standard deviation in inscribed pore throat radii. This corresponds to whether or not the porous medium is homogenous (has a uniform fluid velocity distribution) or heterogeneous (has a more channelized fluid flow distribution). As the flow rate is non-linearly sensitive to changes in throat size (equation 1, below) it is not surprising that the distribution in throat size has a strong control on flow rate. Similarly it is unsurprising that the second and third most important features are the average inscribed pore and pore throat radii respectively.

Flow was then simulated on each of these synthetic volumes using traditional digital rock techniques [40], giving the permeability for each network. Each of the synthetic networks had a nominal voxel size of 1µm; however, a much broader range of pore network statistics (with associated permeabilities) could be generated by examining analytical solutions to the Navier-Stokes equation within the viscous dominated regime (such as the Hagen-Poiseuille equation, equation 1)

$$\frac{\Delta P}{\Delta L} = \frac{8\mu}{r^2} v \quad (1)$$

where $\Delta P/\Delta L$ is the pressure gradient, μ is the viscosity, r the pipe radius and v the (linear) fluid average velocity. For any (defined) structure, a Stokes flow solution to the Navier-Stokes equation will predictably scale as the square of the spatial length scale. As such any change in

nominal voxel size (for a topologically and topographically defined network) should change permeability values similarly. An arbitrarily extended set of network statistics could therefore be created, scaling each (dimensioned) measure appropriately with the voxel size, and the permeability values with its square. An extended set of 219,000 network statistics were created using this technique, with nominal voxel sizes ranging from 10nm to 10um and predicted permeability values varying over 10 orders of magnitude (from 10^{-25}m^2 to 10^{-7}m^2 , with a 10-90 percentile range of $1.2 \times 10^{-18}\text{m}^2$ – $9.25 \times 10^{-12}\text{m}^2$) (figure 4).

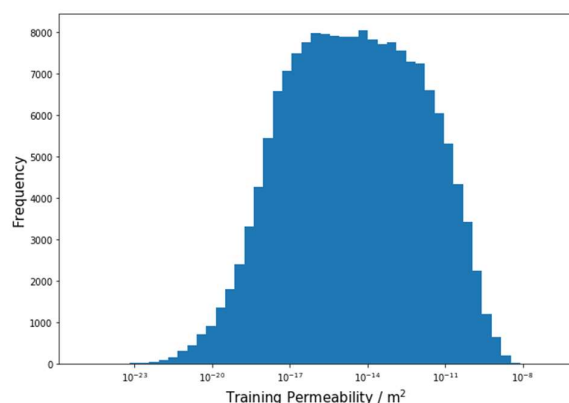


Figure 4 - distribution of training permeability values.

At first inspection it seems like such an approach adds little to the overall training dataset – they are just generated by extrapolation. The key to understanding its utility is that the regression is designed to be performed on real (measured) datasets with real pixel sizes. The same network will have different permeability values if you change its nominal voxel size, and any model must be trained on volumes spanning the entire range of pixel sizes of interest. One way to extend the pixel size range of the training set is to randomly assign pixel sizes, however this will decrease the training dataset density at any specific spatial length scale. Instead Hagen-Poiseuille extrapolation was used as it allows for all the training datasets to be available to the training algorithm at all spatial lengthscales.

3 Results and discussion

This extended dataset of multivariate structural statistics was split into “training” and “testing” sets (with 80% of the data belonging to the training set, and 20% belonging to the test set). The training set was then regressed against the natural logarithm of the permeability using open-source machine learning tools (www.scikit-learn.org) using a randomly seeded Extra-Trees (decision tree based) ensemble algorithm [41]. Total training time on a 40-core workstation was around 12 seconds. A natural logarithmic representation of permeability was used for regression as it allowed prediction errors to be evenly distributed across the entire range of permeability – a linear representation would have concentrated prediction errors strongly towards the largest permeabilities,

overfitting these values and underfitting the rest of the dataset. Another way of viewing this is that a logarithmic representation essentially forces the algorithm to minimize fractional (rather than absolute) errors in the regression model. This model was then tested on the test set, finding a Root-Mean-Square-Fractional-Error (RMSFE) of <4% (figure 5).

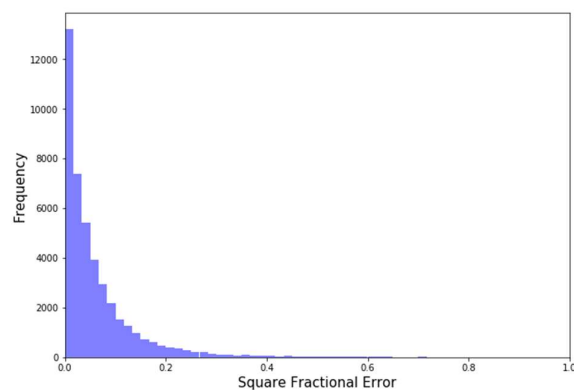


Figure 5 - Root Mean Square Fractional Error (RMSFE) distribution

One major question when applying this model (trained on synthetic networks) to real datasets is whether the synthetic images really represent real imaged pore networks. A composite set of real digital rock images to validate the predictive model for more general use. This composite set of 36 images consisted of a range of open source volumes taken from the Digital Rock Portal (www.digitalrockportal.org), previously published data, and new nano-CT datasets from micritic carbonate microporosity. This constituted two qualitatively different pore networks (the intergranular pore networks common in sandstones and micritic microporosity common in dual porosity carbonate reservoirs). While there are significant differences in chemistry and scale for the networks, in both cases the grain networks consist of euhedral to sub-hedral sub-angular clasts, so they may have a similar relationship between pore structure and flow properties and have grains which are well represented by convex polyhedra.

The unsegmented datasets were segmented using state-of-the-art machine learning based segmentation [24] (Zen Intellesis). Flow was then simulated on each of the segmented datasets, with predicted permeabilities ranging over 6 orders of magnitude ($10\mu\text{D}$ to 10D). These predictions were then compared with multivariate predictions with a RMSFE of <25% across the entire range (figure 6).

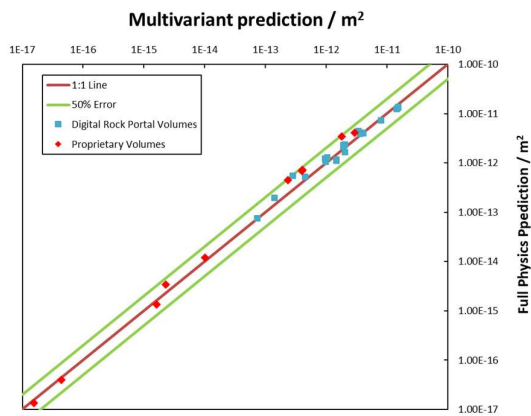


Figure 6 - correlation between simulated permeabilities and permeability predicted using multivariate regression.

Such a result is extremely encouraging – while the prediction error was higher in the (real) validation datasets than the (synthetic) test datasets, the prediction error was still relatively small. Predictions were accurate across multiple pore network types (both intergranular sandstone pore networks and micritic microporosity), and a wide range of length scales and permeability values.

This technique offers multiple advantages over both traditional (full physics) simulations and legacy estimation techniques. First, while analysing and simulating flow on the thousands of training images was computationally expensive, applying it to each imaged dataset was relatively cheap. As such it is significantly faster to apply than traditional techniques, with a large volume prediction taking less than a minute on standard lab computational resources, where a full physics simulation of the same result might take many hours. Perhaps the most important advantage of this technique, however, is that it has the potential of greatly expanding the forms of data applicable to quantitative flow prediction, as the features used for both regression and prediction can be extracted in 2D as well as 3D. This allows for effective property prediction from structural analysis of 2D data (e.g. light or electron microscopy), which can be acquired over much larger areas (allowing a much better characterization of structural heterogeneity), can be acquired much faster, and offers qualitatively richer data (giving information about mineralogy, geochemistry or texture not present when using 3D imaging techniques).

A significant drawback to traditional techniques for permeability estimation (like Kozeny-Carman) is the challenging and arbitrary nature of the measurement of their constituent parameters [42]. A significant advantage of multivariate regression-based techniques is that all inherent errors in measurement are included when performing the regression - the minimization of the prediction error implicitly includes the minimization of the impact of any systematic measurement errors. The general approach for the prediction of effective medium properties based on multivariate structural statistics can be extended to other systems through the incorporation of either other synthetic pore network generation routines or

re-enforcement learning with real (imaged) data. There is also significant scope for future development in the extraction of structural statistics, including the usage of multi-point statistical characterization [43] or recently developed persistent homology metrics [44]. Future work will focus on the examination of the usage of such large area predictive statistical techniques to validate and inform upscaling techniques, allowing for integration of multiple scales of data in single predictive models.

4 Conclusions

A novel method for permeability prediction using multivariate statistical regression is presented. Over 2,000 synthetic geometries were created by modelling granular media using convex polyhedral. Full physical simulations were performed on all of these datasets. The resulting permeability values were used as a target function for a high dimensional description of the volumes, created from structural statistics of rock grains, pores and throats. The regression model was trained with an error of less than 4% in the test data. Validation of this model was performed on a composite series of 36 independent datasets, taken from both an open source data source (www.digitalrockportal.org) and proprietary data from carbonate microporosity. These datasets range in pore system from carbonate microporosity to intergranular sandstone pores, with pixel sizes ranging from 10µm to 32nm and permeabilities ranging over 6 orders of magnitude from 10µD to 10D. Predictions in this validation set showed a RMSFE of less than 25% over the entire range.

Differences in error between the validation and test data provide an interesting insight into the validity of the synthetic datasets in training multivariate models – while they do not perfectly represent real (imaged) data, they are close enough to accurately predict data across length scales. Interesting future work could focus on expanding the training dataset to include networks representative of other pore systems, either through an expansion of synthetic generation techniques or through the incorporation of real imaged geometries.

Such an approach to permeability prediction is extremely promising for addressing many problems associated with pore scale analysis of geological materials, particularly those associated with scale and heterogeneity. By reducing the relative computational complexity of simulation larger geometries can be efficiently analysed. Perhaps more exciting is the application of these tools (trained on 2D slices through 3D volumes) to other pore structural analytical technologies operating in 2D. These techniques are able to address much larger linear length scales, thereby enabling for a much more effective assessment of structural heterogeneity. They also allow for the incorporation of analytical information not accessible in 3D.

Acknowledgements

I thank ZEISS microscopy and Math2Market GmbH. for the equipment and software access required for this work. Many thanks to Dr. Hannah Menke for assistance in the preparation of this manuscript.

References

- [1] G. Bruno, A. Efremov, B. Wheaton, I. Bobrikov, V. G. Simkin, and S. Misture, "Micro- and macroscopic thermal expansion of stabilized aluminum titanate," *J. Eur. Ceram. Soc.*, 2010.
- [2] G. Bruno *et al.*, "On the stress-free lattice expansion of porous cordierite," *Acta Mater.*, 2010.
- [3] P. R. Shearing, L. E. Howard, P. S. Jørgensen, N. P. Brandon, and S. J. Harris, "Characterization of the 3-dimensional microstructure of a graphite negative electrode from a Li-ion battery," *Electrochem. commun.*, 2010.
- [4] D. Kehrwald, P. R. Shearing, N. P. Brandon, P. K. Sinha, and S. J. Harris, "Local Tortuosity Inhomogeneities in a Lithium Battery Composite Electrode," *J. Electrochem. Soc.*, 2011.
- [5] R. Dann, M. Turner, M. Close, and M. Knackstedt, "Multi-scale characterisation of coastal sand aquifer media for contaminant transport using X-ray computed tomography," *Environ. Earth Sci.*, 2011.
- [6] H. P. Menke, M. G. Andrew, M. J. Blunt, and B. Bijeljic, "Reservoir condition imaging of reactive transport in heterogeneous carbonates using fast synchrotron tomography - Effect of initial pore structure and flow conditions," *Chem. Geol.*, vol. 428, pp. 15–26, 2016.
- [7] M. Andrew, B. Bijeljic, and M. J. Blunt, "Pore-by-pore capillary pressure measurements using X-ray microtomography at reservoir conditions: Curvature, snap-off, and remobilization of residual CO₂," *Water Resour. Res.*, vol. 50, no. 11, pp. 8760–8774, 2014.
- [8] S. Iglauer, S. Favretto, G. Spinelli, G. Schena, and M. J. Blunt, "X-ray tomography measurements of power-law cluster size distributions for the nonwetting phase in sandstones," *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 82, pp. 10–12, 2010.
- [9] M. J. Blunt *et al.*, "Pore-scale imaging and modelling," *Adv. Water Resour.*, vol. 51, pp. 197–216, 2013.
- [10] a. V. Ryazanov, K. S. Sorbie, and M. I. J. van Dijke, "Structure of residual oil as a function of wettability using pore-network modelling," *Adv. Water Resour.*, vol. 63, pp. 11–21, 2014.
- [11] J. Schmatz, J. L. Urai, S. Berg, and H. Ott, "Nano-scale imaging of pore-scale fluid-fluid-solid contacts in sandstone," *Geophys. Res. Lett.*, p. n/a-n/a, 2015.
- [12] D. Silin, L. Tomutsa, S. M. Benson, and T. W. Patzek, "Microtomography and Pore-Scale Modeling of Two-Phase Fluid Distribution," *Transp. Porous Media*, vol. 86, pp. 495–515, 2011.
- [13] H. Andrä *et al.*, "Digital rock physics benchmarks-part II: Computing effective properties," *Comput. Geosci.*, vol. 50, pp. 33–43, 2013.
- [14] T. Ramstad, N. Idowu, C. Nardi, and P. E. Øren, "Relative Permeability Calculations from Two-Phase Flow Simulations Directly on Digital Images of Porous Rocks," *Transp. Porous Media*, 2012.
- [15] I. Shikhov and C. H. Arns, "Evaluation of Capillary Pressure Methods via Digital Rock Simulations," *Transp. Porous Media*, 2015.
- [16] M. A. Knackstedt, S. Latham, M. Madadi, A. Sheppard, T. Varslot, and C. Arns, "Digital rock physics: 3D imaging of core material and correlations to acoustic and flow properties," *Lead. Edge*, 2009.
- [17] Y. Keehm, T. Mukerji, and A. Nur, "Computational rock physics at the pore scale: Transport properties and diagenesis in realistic pore geometries," *Lead. Edge*, 2002.
- [18] H. Andrä *et al.*, "Digital rock physics benchmarks-Part I: Imaging and segmentation," *Comput. Geosci.*, vol. 50, pp. 25–32, 2013.
- [19] P. C. Carman, "Fluid flow through granular beds," *Trans. Inst. Chem. Eng.*, vol. 15, pp. 150–166, 1937.
- [20] S. Kuwabara, "The Forces experienced by Randomly Distributed Parallel Circular Cylinders or Spheres in a Viscous Flow at Small Reynolds Numbers," *J. Phys. Soc. Japan*, 1959.
- [21] S. Mauran, L. Rigaud, and O. Coudevylle, "Application of the carman-kotenzy correlation to a high-porosity and anisotropic consolidated medium: The compressed expanded natural graphite," *Transp. Porous Media*, 2001.
- [22] P. N. Singh and W. W. Wallender, "Effects of Adsorbed Water Layer in Predicting Saturated Hydraulic Conductivity for Clays with Kozeny–Carman Equation," *J. Geotech. Geoenvironmental Eng.*, 2008.
- [23] L. Mosser, O. Dubrule, and M. J. Blunt, "Reconstruction of three-dimensional porous media using generative adversarial neural networks," *Phys. Rev. E*, 2017.
- [24] M. Andrew, "A quantified study of segmentation techniques on synthetic geological XRM and FIB-SEM images," *Comput. Geosci.*, vol. 22, no. 6, pp. 1503–1512, 2018.
- [25] O. Sudakov, E. Burnaev, and D. Koroteev, "Driving digital rock towards machine learning: Predicting permeability with gradient boosting and deep neural networks," *Comput. Geosci.*, vol. 127, pp. 91–98, 2019.
- [26] J. Bear, "Dynamics of Fluids in Porous Media," *Soil Sci.*, 2006.
- [27] P. S. Ringrose, a. W. Martinius, and J. Alvestad, "Multiscale geological reservoir modelling in practice," *Geol. Soc. London, Spec. Publ.*, vol. 309, no. 1, pp. 123–134, 2008.

- [28] M. Andrew, “Comparing organic-hosted and intergranular pore networks: topography and topology in grains, gaps and bubbles,” *Geol. Soc. London, Spec. Publ.*, vol. 484, pp. SP484–4, 2018.
- [29] T. Phenrat, H. J. Kim, F. Fagerlund, T. Illangasekare, R. D. Tilton, and G. V. Lowry, “Particle size distribution, concentration, and magnetic attraction affect transport of polymer-modified Fe0 nanoparticles in sand columns,” *Environ. Sci. Technol.*, 2009.
- [30] F. De Smedt and P. J. Wierenga, “Solute Transfer Through Columns of Glass Beads,” *Water Resour. Res.*, 1984.
- [31] P. E. Øren and S. Bakke, “Process based reconstruction of sandstones and prediction of transport properties,” *Transp. Porous Media*, vol. 46, no. 2–3, pp. 311–343, 2002.
- [32] S. Bakke and P.-E. Øren, “3-D Pore-Scale Modelling of Sandstones and Flow Simulations in the Pore Networks,” *SPE J.*, vol. 2, no. 02, pp. 136–149, 1997.
- [33] T. Lerdahl, P. E. Oren, and S. Bakke, “A predictive network model for three-phase flow in porous media,” *SPE/DOE Improv. Oil Recover. Symp.*, 2000.
- [34] P. A. Cundall, “Formulation of a three-dimensional distinct element model-Part I. A scheme to detect and represent contacts in a system composed of many polyhedral blocks,” *Int. J. Rock Mech. Min. Sci.*, 1988.
- [35] A. Rabbani, S. Jamshidi, and S. Salehi, “An automated simple algorithm for realistic pore network extraction from micro-tomography images,” *J. Pet. Sci. Eng.*, vol. 123, pp. 164–171, 2014.
- [36] M. G. Andrew, H. P. Menke, M. J. Blunt, and B. Bijeljic, “The Imaging of Dynamic Multiphase Fluid Flow Using Synchrotron-Based X-ray Microtomography at Reservoir Conditions,” *Transp. Porous Media*, vol. 110, pp. 1–24, 2015.
- [37] A. Q. Raeini, B. Bijeljic, and M. J. Blunt, “Generalized network modeling: Network extraction as a coarse-scale discretization of the void space of porous media,” *Phys. Rev. E*, vol. 96, no. 1, 2017.
- [38] H. Dong and M. J. Blunt, “Pore-network extraction from micro-computerized-tomography images,” *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 80, no. 3, 2009.
- [39] Gilles Louppe, “Understanding Random Forests: From Theory to Practice,” 2014.
- [40] S. Linden, A. Wiegmann, and H. Hagen, “The LIR space partitioning system applied to the Stokes equations,” *Graph. Models*, vol. 82, pp. 58–66, 2015.
- [41] P. Geurts, D. Ernst, and L. Wehenkel, “Extremely randomized trees,” *Mach. Learn.*, 2006.
- [42] H. P. Menke, Y. Gao, S. Linden, and M. G. Andrew, “Using nano-XRM and high-contrast imaging to inform micro-porosity permeability during Stokes-Brinkman single and two-phase flow simulations on micro-CT images,” *Adv. Water Resour.*
- [43] H. Okabe and M. J. Blunt, “Prediction of permeability for porous media reconstructed using multiple-point statistics,” *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.*, 2004.
- [44] A. L. Herring, V. Robins, and A. P. Sheppard, “Topological Persistence for Relating Microstructure and Capillary Fluid Trapping in Sandstones,” *Water Resour. Res.*, 2019.