



Data Article

3D Dataset of binary images: A collection of synthetically created digital rock images of complex media

Javier E. Santos*, Michael J. Pyrcz, Maša Prodanović

Hildebrand Department of Petroleum and Geosystems Engineering, The University of Texas at Austin, Austin, TX, USA

ARTICLE INFO

Article history:

Received 27 September 2021

Revised 22 December 2021

Accepted 3 January 2022

Available online 6 January 2022

Keywords:

Sphere-packs

Fractures media

Catalyst layers

Shales

Vuggs

ABSTRACT

Digital rock images are computational representations that capture the geometrical complexity of systems present ubiquitously in nature. In recent years, their use has become widespread due to the increasing availability of repositories, and open-source physics simulators and analysis tools. Here, we present a dataset of 3D binary geometries in a standardized format that represent a wide variety of geological and engineering systems. Our data is freely available at [1].

© 2022 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Specifications Table

Subject	Computers in Earth Sciences
Specific subject area	3D Digital Rock Images
Type of data	Image
How data were acquired	Computer simulations
Data format	3D arrays in HDF5 format
Parameters for data collection	After carrying-out the simulation to generate each image, all the isolated pore bodies were removed and a check was performed to assess connectivity along the third coordinate axis

(continued on next page)

* Corresponding author.

E-mail address: jesantos@utexas.edu (J.E. Santos).

Social media: (J.E. Santos), (M.J. Pyrcz), (M. Prodanović)

<https://doi.org/10.1016/j.dib.2022.107797>

2352-3409/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Description of data collection	The images are presented as binary 3D arrays of zeros and ones. The former represent the pore-space and the later the solid phase.
Data source location	Institution: The University of Texas at Austin
Data accessibility	Repository name: Digital Rocks Portal (DPR) Data identification number: 374 Direct URL to data: https://www.digitalrocksportal.org/projects/374 The code to create this dataset is part of the MPLBM-UT library [2].

Value of the Data

- The dataset includes 75 3D geometries (rock samples) representing a wide variety of depositional systems and diagenetical processes. This data has been used in a number of research articles [3–11].
- The collection is standardized with a high-performance file format (HDF5) that integrates its metadata (for portability), making it accessible with any programming language and popular scientific-visualization software. All the samples were post-processed to ensure 6-connectivity which allows flow simulations to converge faster.
- Researchers and users involved in direct simulations methods of physical processes can benefit from these images for benchmarking their simulations and inferring new physical insights with a standardized dataset. AI researchers will utilize this dataset of complex geometries to test supervised and unsupervised classification algorithms.
- The images can be used as inputs for physics-based simulations (e.g., single- and multi-phase flow [2,12,13], particle transport, nanoconfinement, electrical conductivity, and solid mechanics) to get new physical insights (e.g., scaling relationships, surrogate modeling, etc.), geometrical analysis (e.g., geostatistics, spatial statistics, etc.) [14], train supervised/unsupervised deep learning models, on samples that are representative of different subsurface settings.

1. Data Description

Flow and transport in porous media are critical to understand the geological processes of rock formation and applications such as the management of groundwater resources, carbon sequestration, enhanced oil recovery, and contaminant transport. Typical geological systems are composed of a broad spectrum of porous media with features such as porosity and permeability varying by orders of magnitude within an individual system. An example application is to utilize a digital rock as an analog to derive the relationship between porosity and permeability to impute missing permeability data from porosity data points, which are widely available.

The term ‘digital rock’ refers to a computational model of a porous medium in 2D or 3D whose source is either an actual image of a rock or soil (e.g., an x-ray microtomography image), or it is based on a model (e.g., calculated via process-based simulations of sediment transportation and deposition stochastic realization from a geostatistical model). These are commonly formatted in a regular grid or mesh that allow their use in a numerical simulator to forecast a transport process or property of interest. These simulations are then post-processed to obtain a bulk property that characterizes the rock in question, and this research area is often referred to as digital rock physics, or digital core analysis. Digital Rocks Portal [15] provides a curated library of digital rocks representing a wide variety of domains ranging from imaged biofilms in soil to texturally equilibrated models of salt pore spaces. Digital Rocks Portal stores both actual images from a variety of imaging techniques and representative porous media models. Digital Rocks Portal is organized in units called *projects*.

The dataset collection (Digital Rocks Portal project) described here provides 3D porous and fractured models that systematically vary a range of their properties and can be used for benchmarking simulation algorithms, including systematic training of deep learning algorithms predicting transport properties [10,11]. Datasets are in the form of binary images that have designated pore (void) and solid (grain) numerical cells or voxels, and are organized in 10 groups (or

samples, in Digital Rocks Portal jargon). Many have already been used in previously published work though not necessarily shared. The groups are described as follows:

Group 1 contains discretized representation of experimental measurement of a packing of identical spheres (Sample 1, the original is stored at the portal [16]). This sample is then systematically changed to mimic diagenesis processes. One morphological sphere dilation and six morphological erosions are numerically performed to obtain samples with higher (Sample 0) and lower porosity (Sample 2–8) respectively, but that still exhibit a similar pore structure to the original sample. These samples come in both 256^3 and 480^3 .

Group 2 contains self-similar fractures running through the middle of the volume. This group contains five fractures with varying surface roughness representing different materials. These five fractures are downsampled three times to create fractures with smaller aperture that still exhibit the same degree of surface roughness. All the downsampled fractures have the same mean aperture (hence, porosity) respectively.

Group 3 contains packings of randomly placed spheres. Unlike Group 1, these spheres are allowed to overlap slightly to achieve a target porosity. These images are well-suited to represent catalyst layers [17]. This group contains samples with six porosities (10, 15, 20, 25, 30, and 35%) with three realizations each (since the algorithm uses a random number seed to start the simulation). **Group 7** contains samples created with the same algorithm and with different seeds.

Group 4 contains two process-based reconstructions of nanoporous organic matter in shale [5]. These samples have a porosity of 20 and 30%.

Group 5 contains four fractured samples. Unlike Group 2, the fractures have an adjacent permeable matrix [3,4].

Group 6 contains two samples with a linear gradient in the porosity of their slices in the z-direction.

Group 8 contains a realistic single fracture on a solid domain. The roughness of this sample is heterogeneous and the aperture field is not self-affine.

Group 9 Randomly placed overlapping spheres with an embedded synthetic fracture along the z-axis.

Group 10 Voxelized sphere packs where a certain number of solid grains were removed to simulate a vuggy domain.

The highest and lowest porosity values (defined as number of zero-entries over image size) in each group can be visualized in Fig. 1. A selection of samples is shown in Fig. 2.

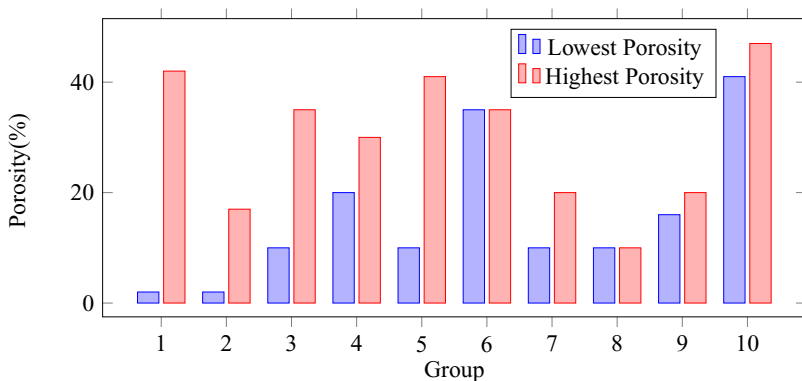


Fig. 1. Lowest and highest porosity sample values in each group.

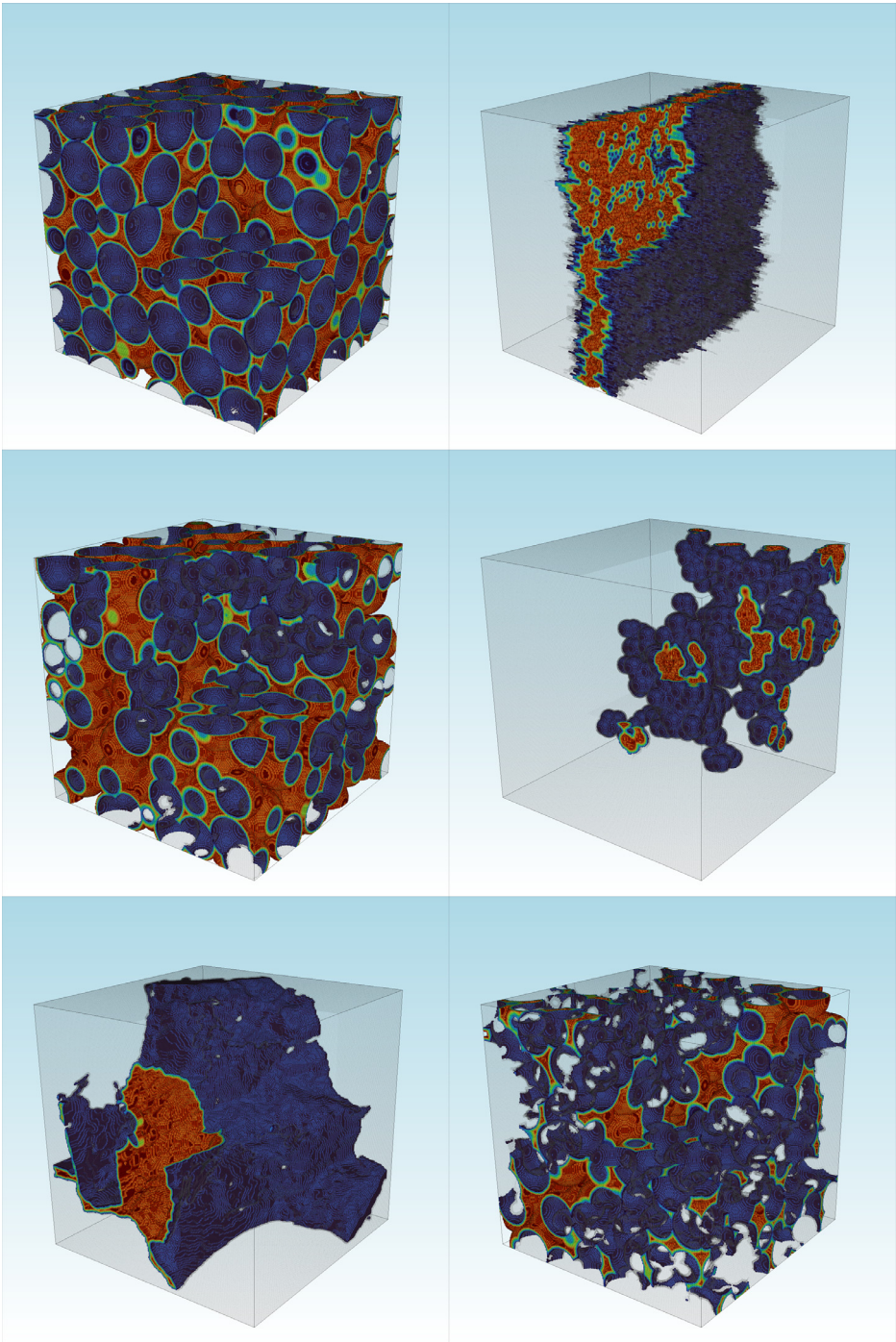


Fig. 2. 3D plots of dataset samples. Each subplot shows the Euclidean distance of the pore space for visualization purposes. The code to reproduce this figure is shown in [Listing 1](#).

```

1 import urllib, hdf5storage, scipy.ndimage, vedo # import libs
2
3 # download image from the Digital Rocks Portal
4 drp = 'https://www.digitalrocksporal.org/'
5 url = drp + 'projects/374/images/309544/download/'
6 urllib.request.urlretrieve( url, im_name:='374_01_00_256.mat' )
7
8 # read-in image
9 bin_im = hdf5.storage.loadmat(im_name)['bin']
10
11 # obtain the Euclidean distance of the pore-space
12 edist = scipy.ndimage.morphology.distance_transform_edt(bin_im==0)
13
14 # remove upper corner
15 half_size = bin_im.shape[0]//2
16 edist[ -half_size:, -half_size:, :half_size] = 0
17 bin_im[-half_size:, -half_size:, :half_size] = -1
18
19 # plot using Vedo
20 plot = vedo.Volume(edist).legosurface(vmin=1, vmax=5, cmap='turbo')
21 plot += vedo.Volume(bin_im).legosurface(vmin=1, vmax=2).c('lightgray').
    opacity(0.05)
22 vedo.plotter(plot)

```

Listing 1. Python code to download and plot the binary images as shown in Fig. 2.

2. Experimental Design, Materials and Methods

All the image groups were created synthetically to represent different natural settings. The main motivation to create this dataset was to have a diversity of benchmark samples to perform transport simulations and geometric analysis. All the samples are discretised as 256^3 volumes and some have an additional 480^3 volume. All of these are presumed unit-less and can be scaled to any physical length of interest (for instance, spheres diameters can be scaled by a relevant mean grain diameter). The pore-space is represented with zeros and the solid matrix with ones. Most of the images have been used in projects related to a wide range of earth-science applications.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

Acknowledgments

We would like to acknowledge the contributions of Dr. Yin Ying, Dr. Christopher J. Landry, and Dr. Hasan J. Khan to this project.

References

- [1] J. E. Santos, Y. Yin, M. Prodanovic, A. Gigliotti, N. Lubbers, H. Khan, 3D Collection of Binary Images, 2021a, doi:[10.17612/nxgk-k611](https://doi.org/10.17612/nxgk-k611). (<http://www.digitalrocksporal.org/projects/374a>).
- [2] J. E. Santos, A. Bihani, A. Gigliotti, MPLBM-UT: Single- and Multi-phase simulations in PM, 2021b, [10.5281/zenodo.4665905](https://doi.org/10.5281/zenodo.4665905).
- [3] M. Prodanovic, S.L. Bryant, Physics-Driven interface modeling for drainage and imbibition in fractures, SPE J. 14 (03) (2009) 532–542, doi:[10.2118/110448-PA](https://doi.org/10.2118/110448-PA).

- [4] M. Prodanovic, S.L. Bryant, Z.T. Karpyn, Investigating matrix/fracture transfer via a level set method for drainage and imbibition, *SPE J.* 15 (01) (2010) 125–136, doi:[10.2118/116110-PA](https://doi.org/10.2118/116110-PA). <https://onepetro.org/SJ/article/15/01/125/192097/Investigating-Matrix-Fracture-Transfer-via-a-Level>.
- [5] C.J. Landry, M. Prodanović, P. Eichhubl, Direct simulation of supercritical gas flow in complex nanoporous media and prediction of apparent permeability, *Int J Coal Geol* 159 (2016) 120–134, doi:[10.1016/j.coal.2016.03.015](https://doi.org/10.1016/j.coal.2016.03.015). <https://linkinghub.elsevier.com/retrieve/pii/S0166516216300659>.
- [6] Y. Yin, Z. Qu, C. Zhu, J. Zhang, Visualizing gas diffusion behaviors in three-dimensional nanoporous media, *Energy & Fuels* 35 (3) (2021) 2075–2086, doi:[10.1021/acs.energyfuels.0c03529](https://doi.org/10.1021/acs.energyfuels.0c03529).
- [7] E. Guiltinan, J.E. Santos, Q. Kang, Residual saturation during multiphase displacement in heterogeneous fractures with novel deep learning prediction, in: *Unconventional Resources Technology Conference (URTEC)*, Society of Petroleum Engineers (SPE), Austin, 2020, doi:[10.15530/urtec-2020-3048](https://doi.org/10.15530/urtec-2020-3048).
- [8] E.J. Guiltinan, J.E. Santos, M.B. Cardenas, D.N. Espinoza, Q. Kang, Two-phase fluid flow properties of rough fractures with heterogeneous wettability: analysis with lattice Boltzmann simulations, *Water Resour Res* (2020), doi:[10.1029/2020WR027943](https://doi.org/10.1029/2020WR027943).
- [9] J. Santos, M. Prodanovic, C. Landry, H. Jo, Determining the impact of mineralogy composition for multiphase flow through hydraulically induced fractures, in: *SPE/AAPG/SEG Unconventional Resources Technology Conference 2018*, URTC 2018, 2018, doi:[10.15530/urtec-2018-2902986](https://doi.org/10.15530/urtec-2018-2902986).
- [10] J.E. Santos, D. Xu, H. Jo, C.J. Landry, M. Prodanović, M.J. Pyrcz, PoreFlow-Net: a 3D convolutional neural network to predict fluid flow through porous media, *Adv Water Resour* 138 (February) (2020) 103539, doi:[10.1016/j.advwatres.2020.103539](https://doi.org/10.1016/j.advwatres.2020.103539). <https://linkinghub.elsevier.com/retrieve/pii/S0309170819311145>.
- [11] J.E. Santos, Y. Yin, H. Jo, W. Pan, Q. Kang, H.S. Viswanathan, M. Prodanović, M.J. Pyrcz, N. Lubbers, Computationally efficient multiscale neural networks applied to fluid flow in complex 3D porous media, 0123456789, Springer Netherlands, 2021, doi:[10.1007/s11242-021-01617-y](https://doi.org/10.1007/s11242-021-01617-y).
- [12] J.E. McClure, Z. Li, M. Berrill, T. Ramstad, The LBPM software package for simulating multiphase flow on digital images of porous rocks, *Comput. Geosci.* 25 (3) (2021) 871–895, doi:[10.1007/s10596-020-10028-9](https://doi.org/10.1007/s10596-020-10028-9).
- [13] J. Gostick, M. Aghighi, J. Hinebaugh, T. Tranter, M.A. Hoeh, H. Day, B. Spellacy, M.H. Sharqawy, A. Bazylak, A. Burns, W. Lehnert, A. Putz, OpenPNM: a Pore network modeling package, *Computing in Science & Engineering* 18 (4) (2016) 60–74, doi:[10.1109/MCSE.2016.49](https://doi.org/10.1109/MCSE.2016.49). <http://ieeexplore.ieee.org/document/7478437/>.
- [14] J. Gostick, Z. Khan, T. Tranter, M. Kok, M. Agnaou, M. Sadeghi, R. Jervis, Porespy: a Python toolkit for quantitative analysis of porous media images, *Journal of Open Source Software* 4 (37) (2019) 1296, doi:[10.21105/joss.01296](https://doi.org/10.21105/joss.01296).
- [15] M. Prodanovic, M. Esteva, M. Hanlon, G. Nanda, P. Agarwal, Digital Rocks Portal: a repository for porous media images, 2015, <https://www.digitalrockportal.org/>. 10.17612
- [16] J. Finney, M. Prodanovic, Finney Packing of Spheres, 2016, (<http://www.digitalrockportal.org/projects/47>). 10.17612/P78G69
- [17] K.J. Lange, P.-C. Sui, N. Djilali, Pore scale simulation of transport and electrochemical reactions in reconstructed PEMFC catalyst layers, *J Electrochem Soc* 157 (10) (2010) B1434, doi:[10.1149/1.3478207](https://doi.org/10.1149/1.3478207).