Supporting Information for “Diffusion Model-Based Generation of Three -dimensional Multiphase Fluid Pore-Scale Images”

***Linqi Zhu*[a,*], *Branko Bijeljic*[a], *Martin J. Blunt*[a]**

[a]Department of Earth Science and Engineering, Imperial College London, SW7 2BP, London, United Kingdom

Introduction Here, we present additional results, starting with the formulae for the two-point correlation function and the two-point cross-correlation function. We then display the two-point correlation and cross-correlation functions for the real images. Finally, we provide an overall results table for the image generation of Bentheimer sandstone multiphase images, including a comparison of parameters.

1 Formulae

The two-point correlation function, $S_2(r)$, is defined as follows:

$$S_2(r) = \frac{1}{V}\int_V \chi_\phi(\mathbf{x})\chi_\phi(\mathbf{x}+\mathbf{r})\,d\mathbf{x} \tag{S1}$$

where $\chi_\phi(\mathbf{x})$ is the characteristic function for the phase ϕ at location $\mathbf{x}$, $\mathbf{r}$ is the displacement vector between two points, and V is the volume over which the integration is performed.

The characteristic function, $\chi_\phi(\mathbf{x})$, is defined as:

$$\chi_\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \text{ is in phase } \phi, \\ 0 & \text{otherwise.} \end{cases}$$

In this context, the characteristic function indicates whether a point $\mathbf{x}$ belongs to a specific phase ϕ. It takes the value of 1 if the point is in phase ϕ and 0 otherwise.

The two-point cross-correlation function, $C_2(r)$, is defined as follows:

$$C_2(r) = \frac{1}{V}\int_V \chi_{\phi_1}(\mathbf{x})\chi_{\phi_2}(\mathbf{x}+\mathbf{r})\,d\mathbf{x} \tag{S2}$$

where $\chi_{\phi_1}(\mathbf{x})$ and $\chi_{\phi_2}(\mathbf{x}+\mathbf{r})$ are the characteristic functions for the phases ϕ_1 and ϕ_2 at locations $\mathbf{x}$ and $\mathbf{x}+\mathbf{r}$, respectively, $\mathbf{r}$ is the displacement vector between two points, and V is the volume over which the integration is performed.

The characteristic functions, $\chi_{\phi_1}(\mathbf{x})$ and $\chi_{\phi_2}(\mathbf{x}+\mathbf{r})$, are defined as:

$$\chi_{\phi_1}(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \text{ is in phase } \phi_1, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\chi_{\phi_2}(\mathbf{x}+\mathbf{r}) = \begin{cases} 1 & \text{if } \mathbf{x}+\mathbf{r} \text{ is in phase } \phi_2, \\ 0 & \text{otherwise.} \end{cases}$$

In this context, the characteristic function indicates whether a point $\mathbf{x}$ or $\mathbf{x}+\mathbf{r}$ belongs to a specific phase (ϕ, ϕ_1, or ϕ_2). It takes the value of 1 if the point is in the specified phase and 0 otherwise.

2 Algorithm

2.1 Training Phase

The training phase of a Denoising Diffusion Implicit Model (DDIM) involves learning a noise prediction network that predicts the noise added to the data during the diffusion process.

Algorithm 1 Training algorithm for DDIM

Require: Training data set X, number of epochs N, batch size m, number of diffusion steps T
Ensure: Trained noise prediction network ϵ_θ
  for epoch $= 1, N$ **do**
    for each batch **do**
      Sample batch of m examples $\{x^{(1)}, \ldots, x^{(m)}\}$ from data generating distribution $p_{\text{data}}(x)$
      for each $x^{(i)}$ in batch **do**
        Sample a timestep $t \sim \text{Uniform}(\{1, \ldots, T\})$
        Sample noise $\epsilon \sim \mathcal{N}(0, I)$
        Generate noisy data $x_t^{(i)} = \sqrt{\alpha_t} x^{(i)} + \sqrt{1 - \alpha_t} \epsilon$
        Predict the noise $\epsilon_\theta(x_t^{(i)}, t)$
        Compute the loss: $\mathcal{L} = \|\epsilon - \epsilon_\theta(x_t^{(i)}, t)\|^2$
        Update the network parameters by descending the gradient: $\nabla_\theta \mathcal{L}$
      end for
    end for
  end for

2.2 Generation (reconstruction) Phase

Once the DDIM has been trained, the noise prediction network can be used to generate new data samples by reversing the diffusion process.

Algorithm 2 Data generation using a trained DDIM

Require: Trained noise prediction network ϵ_θ, number of samples n, number of diffusion steps T
Ensure: Generated data set Y
Initialize an empty data set Y
for $i = 1, n$ **do**
  Sample noise $x_T \sim \mathcal{N}(0, I)$
  for $t = T, 1, -1$ **do**
    Predict the noise $\epsilon_\theta(x_t, t)$
    Compute the previous step: $x_{t-1} = \frac{x_t - \sqrt{1-\alpha_t}\epsilon_\theta(x_t, t)}{\sqrt{\alpha_t}}$
  end for
  Add x_0 to the data set Y
end for
return Y

3 Additional results

Figures S1 show two-point correlation and cross-correlation functions of real multiphase pore-scale Bentheimer sandstone images, Table S1 reveals the differences between the generated images and the real images after parameterization, including the parameter ranges, medians, and means.

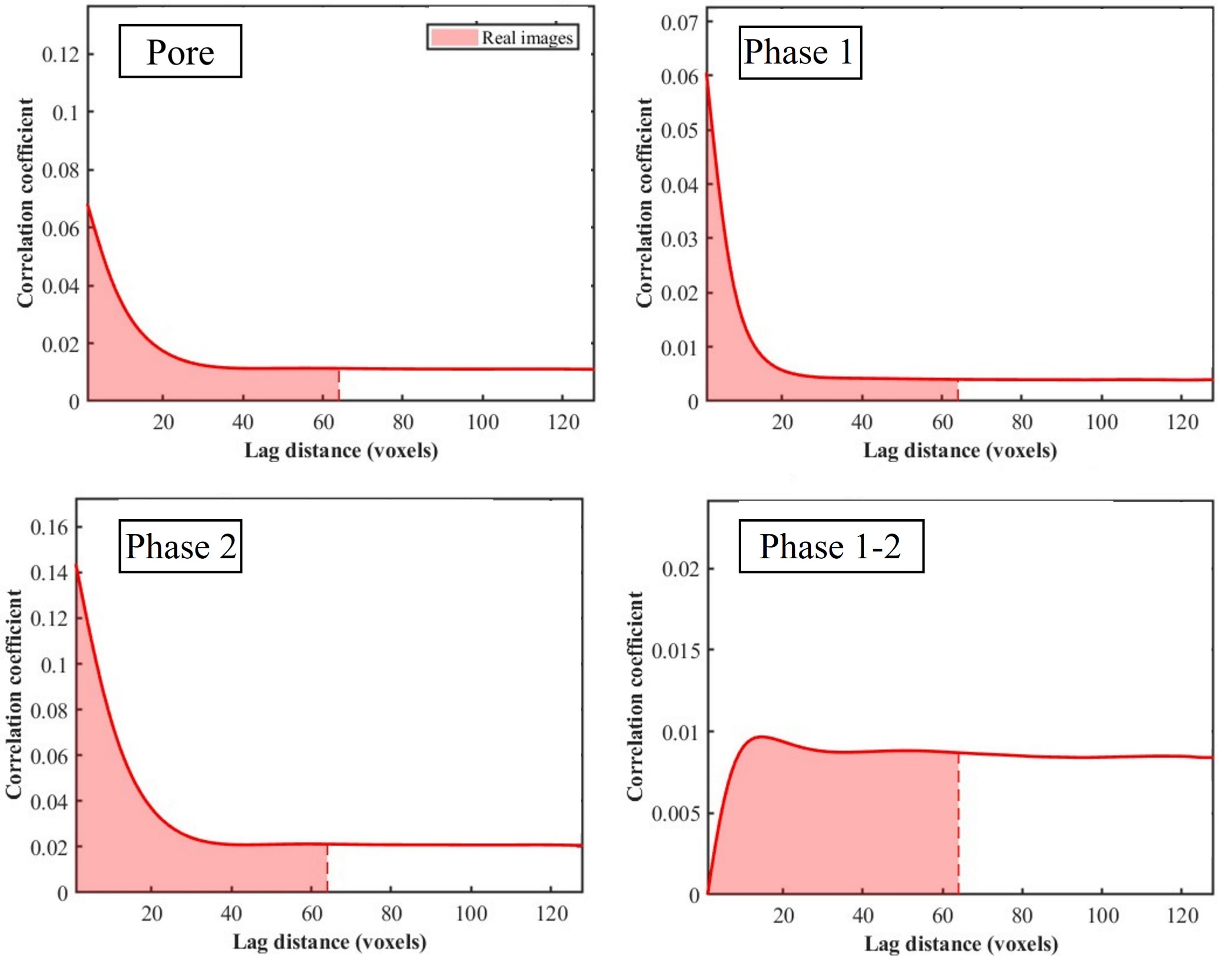

Fig. S1: The two-point correlation function and two-point cross-correlation function of the multiphase pore-scale Bentheimer sandstone images were calculated by dividing the entire image into 1470 images of 128^3 cubes and using all of them for calculation. The figure is obtained by averaging the values calculated from all the images. It can be seen that both the two-point correlation function and the two-point cross-correlation function achieve stable function values at a distance of 64 voxels, which reveals the heterogeneity of the rock and also proves that using 64^3 voxels as the training image size in this paper is appropriate.

Table 1: Comparison of parameters in real images and DDIM images. AP: Absolute permeability, Curv: Curvature, SSA: Specific surface area, ECD: Euler characteristic density, SIA: Specific interfacial area

Real images					
Porosity / -	Statistics	Pore			
		AP $/mm^2$	Curv $/mm^{-1}$	SSA $/mm^{-1}$	ECD $/mm^{-3}$
Standard deviation: 3.7×10^{-2}	Standard deviation	0.98×10^{-6}	3.79	12.23	2.22×10^2
Mean: 0.19	Mean	1.61×10^{-6}	10.4	90.8	-1.44×10^2
Median: 0.18	Median	1.58×10^{-6}	10.7	91.7	-1.26×10^2
		Oil			
		Saturation / -	Curv $/mm^{-1}$	SIA $/mm^{-1}$	ECD $/mm^{-3}$
	Standard deviation	0.14	10.8	23.57	2.59×10^2
	Mean	0.65	35.1	1.55×10^2	3.47×10^2
	Median	0.65	33.3	1.52×10^2	2.95×10^2
		Brine			
		Saturation / -	Curv $/mm^{-1}$	SIA $/mm^{-1}$	ECD $/mm^{-3}$
	Standard deviation	0.14	4.6	8.98	4.37×10^2
	Mean	0.35	13.9	90.8	7.92×10^2
	Median	0.34	14.9	91.66	8.08×10^2

DDIM images					
Porosity / -	Statistics	Pore			
		AP $/mm^2$	Curv $/mm^{-1}$	SSA $/mm^{-1}$	ECD $/mm^{-3}$
Standard deviation: 5.5×10^{-2}	Standard deviation	1.52×10^{-6}	4.3	15.58	2.289×10^2
Mean: 0.20	Mean	2.31×10^{-6}	10.1	84.2	-1.40×10^2
Median: 0.22	Median	2.16×10^{-6}	9.9	82.5	-1.39×10^2
		Oil			
		Saturation / -	Curv $/mm^{-1}$	SIA $/mm^{-1}$	ECD $/mm^{-3}$
	Standard deviation	0.074	9.1	41.1	2.994×10^2
	Mean	0.63	36	1.72×10^2	3.65×10^2
	Median	0.64	34.6	1.67×10^2	3.35×10^2
		Brine			
		Saturation / -	Curv $/mm^{-1}$	SIA $/mm^{-1}$	ECD $/mm^{-3}$
	Standard deviation	0.074	4.24	15.85	4.798×10^2
	Mean	0.37	13.3	84.75	8.74×10^2
	Median	0.36	13.7	84.19	8.70×10^2