Reconstruction of Heterogeneous Media by a Cross-Correlation Function and Graph Theory and Simulation of Transfor Processes Therein

Muhammad Sahimi

Mork Family Department of Chemical Engineering & Materials Science

University of Southern California, Los Angeles

Modeling of Heterogeneous Materials and Media at Multiple Scales is of Fundamental Importance

- **Ceramics**
- **Composite materials**
- **Catalysts and membranes**
- **Biological tissues**
- **Field-scale porous media (aquifers; soil; etc.)**

Depending on the Length Scale, Many Approaches Have Been Developed to Model Heterogeneous Media

- **Continuum approach (phenomenological)**
- **Pore-network models (porous media)**
- **Direct simulation with 2D or 3D images of the media**
- **Developing a high resolution model and then up-scale it**
- **Reconstruction**

What is Reconstruction?

Given a certain amount of data for a heterogeneous material, how does one construct a model for the material that not only honors the data, but also provides accurate predictions for those properties for which little or no data may be available, or are too difficult/expensive to measure, or are not used in the reconstruction?

Stochastic Approaches to Reconstruction

One attempts to generate plausible realizations of the material based on some data

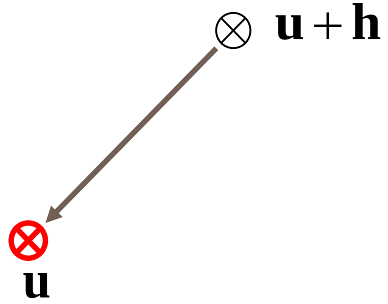
- **Stochastic approaches based on statistics (Adler; Torquato; Jiao, others)**
- **Process-based approaches that use such statistics (Roberts and Schwartz; Hilfer; Bryant and Blunt; Øren and Bakke, others)**

Stochastic approaches to reconstruction (continued)

In all cases, an "energy functional" that measures the difference between the data and the "predictions" of the model is minimized by simulated annealing, the genetic algorithm, or by another minimization/optimization method.

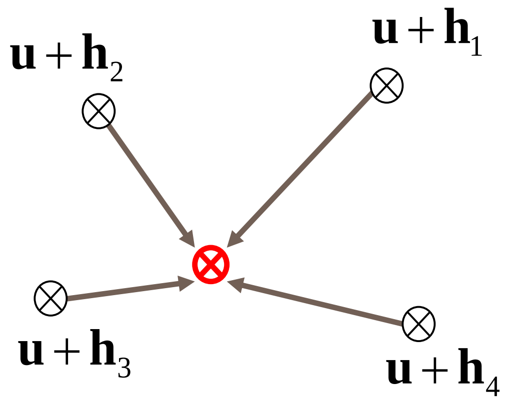
The data used are usually two-point statistics (correlation functions; cluster connectivity functions, etc.)

$$Prob\left\{Z(u) \leq z | Z(u+h) = z_1\right\}$$



Stochastic approaches to reconstruction (continued)

But, there are also methods that do not use an energy functional, but use higher other statistics



- **Multi-point statistics (developed by geoscientist; Strebelle; Arpat; Zhang)**
- **Cross-correlation function approach (Tahmasebi & Sahimi)**

Reconstruction with Cross-Correlation Function

- **As the input data, the method uses digital images, or hard data, etc., which contain the essential features of the material’s morphology**

- **But, it is capable of integrating any other type of data in the reconstruction process**

- **The image could be 2D or 3D**

- **We refer to the method as cross correlation-based simulation (CCSIM)**

Tahmasebi, Hezarkhani, and Sahimi, *Computational Geosciences* **16**, 779 (2012)
Tahmasebi and Sahimi, *Physical Review Letters* **110**, 078002 (2013)

The Approach

Given the Digital Image (DI) or the Data

- Represent the material’s realization (model) by a 2D or 3D computational grid

- Every two neighboring blocks share an overlap region, whose purpose is making transition from one block to the next seamless

- Reconstruct the material block-by-block along a 1D raster path, oriented in any desired direction

The approach (continued)

- **Sample the heterogeneity of the material and fill the first block and its overlap region with the next block. We refer to the details of the overlap region as *data event that can change during reconstruction***
- **Match the overlap region (not the block) with the entire data set or digital image**

How do make sure that the overlap region matches?

The Matching is Based on a Cross-Correlation Function

- **Euclidean distance:**

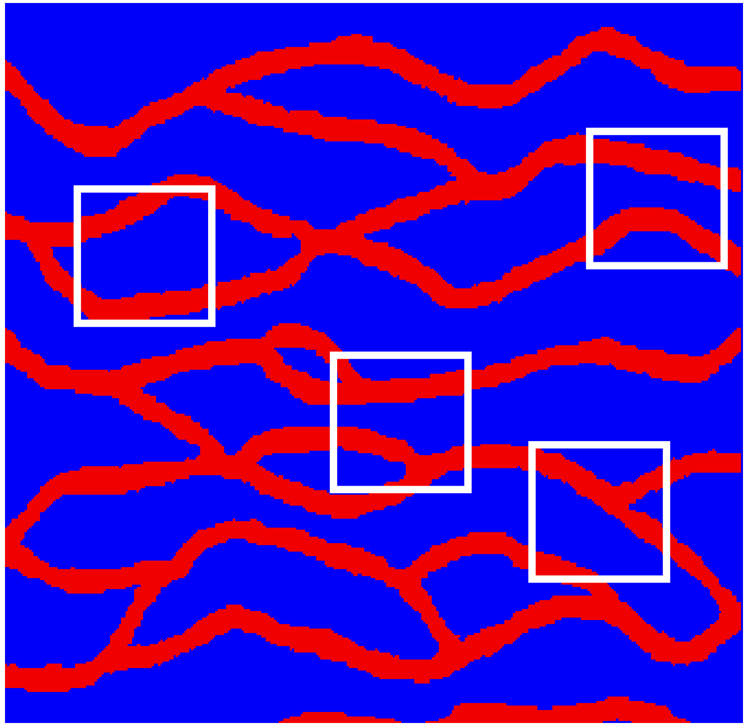
$$d^2(i,j) = \sum_{x=0}^{L_x-1} \sum_{y=0}^{L_y-1} [DI(x+i, y+j) - dev_T(x,y)]^2$$

$$d^2(i,j) = \sum_{x=0}^{L_x-1} \sum_{y=0}^{L_y-1} \left[DI^2(x+i, y+j) - 2DI(x+i, y+j)dev_T(x,y) + dev^2(x,y)\right]$$

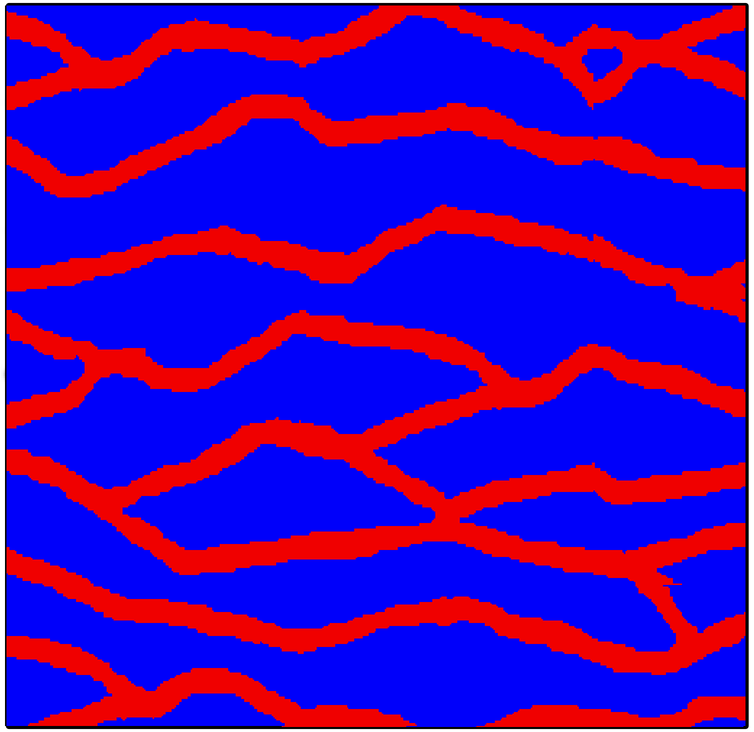
✓ **For the "distance" to be minimum, is represented the cross-correlation function (convolution of the data and the sample) should be maximum:**

$$C(i,j) = \sum_{x=0}^{L_x-1} \sum_{y=0}^{L_y-1} DI(x+i, y+j)dev_T(x,y)$$

The approach (continue)



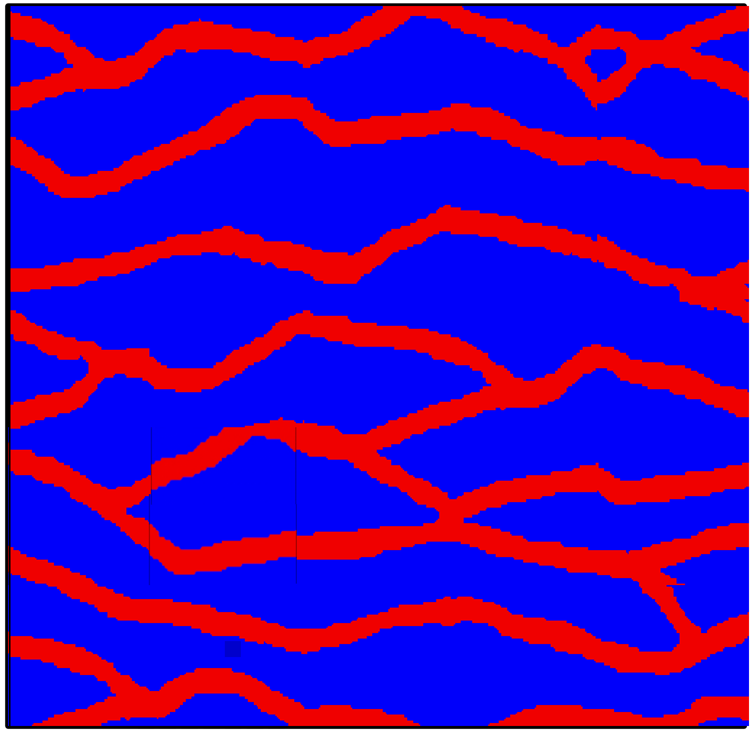
Digital Image



Simulation Grid

- Using raster path for visiting the simulation grid
- Using overlap region for seamless transition
- Using cross correlation for similarity calculations

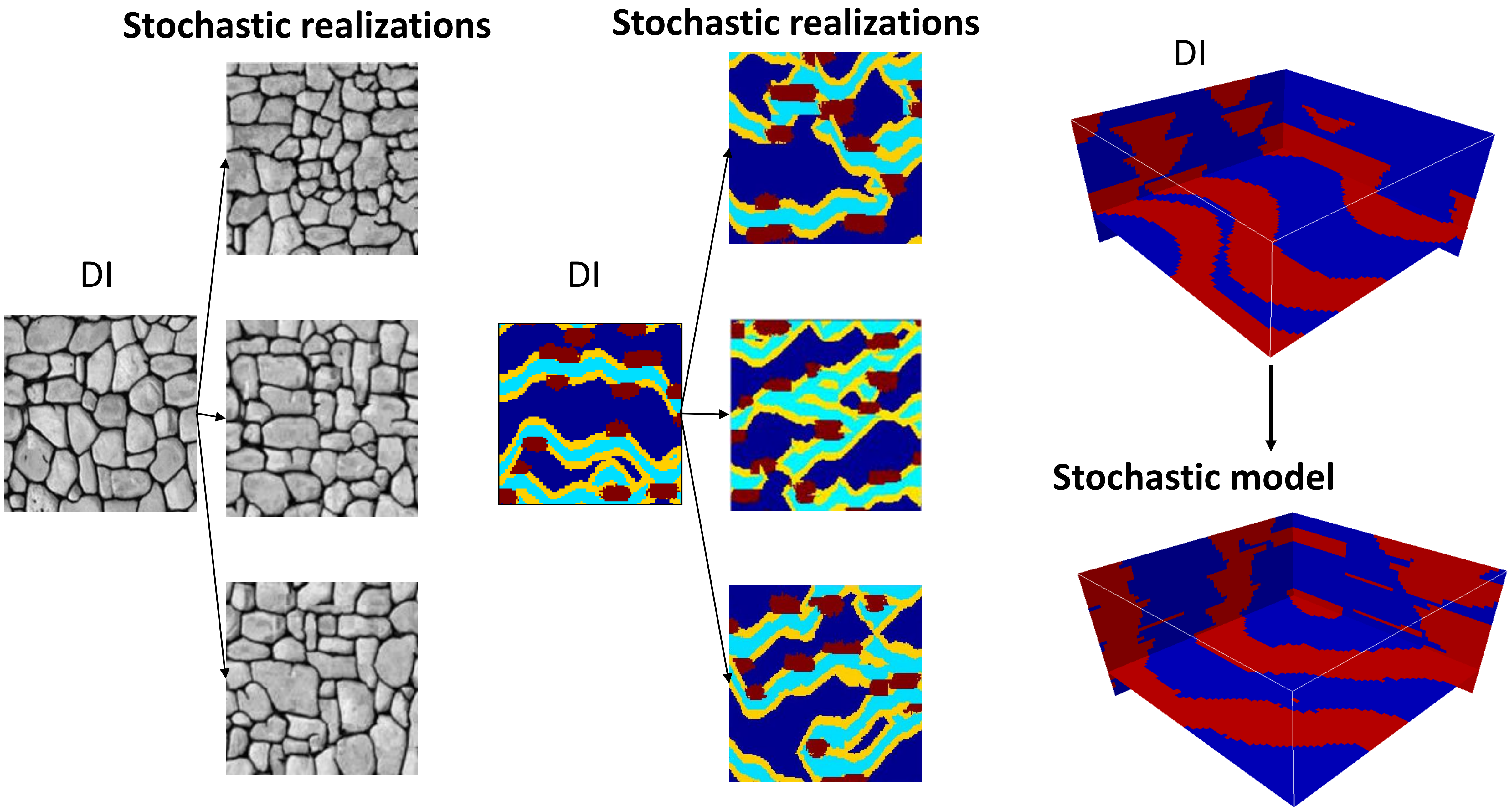
Conditional Modeling: Subject to Honoring Hard Data



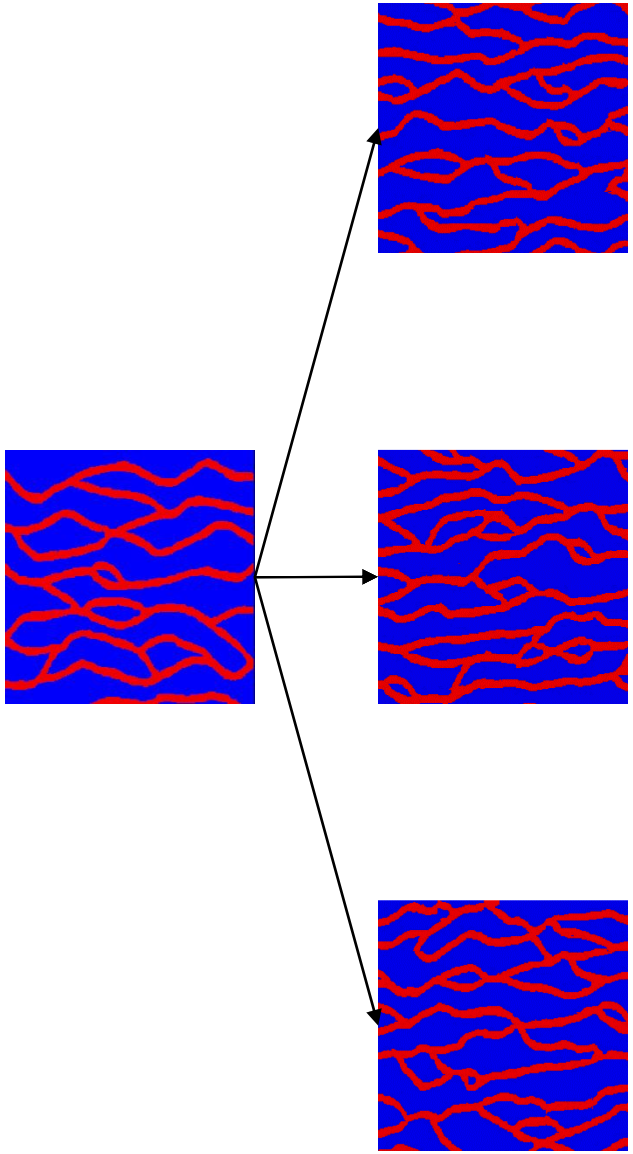
Simulation Grid with Hard Data

- ✓ First, the algorithm tries to find a pattern in the morphology that honors the hard data.
- ✓ If no hard data can be found, then the block containing the data event is divided into smaller blocks

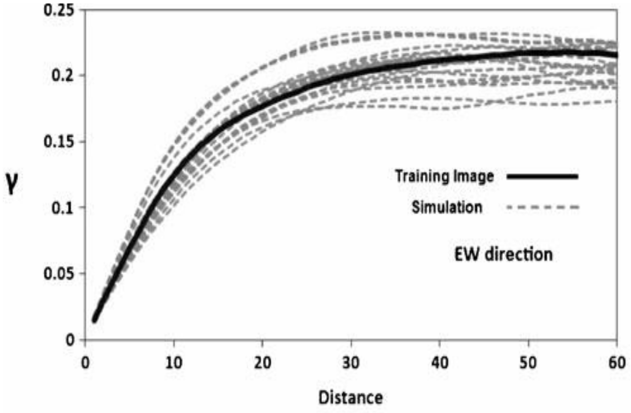
Adaptive recursive template splitting


Stochastic realizations
DI
Stochastic realizations
DI
DI
Stochastic model

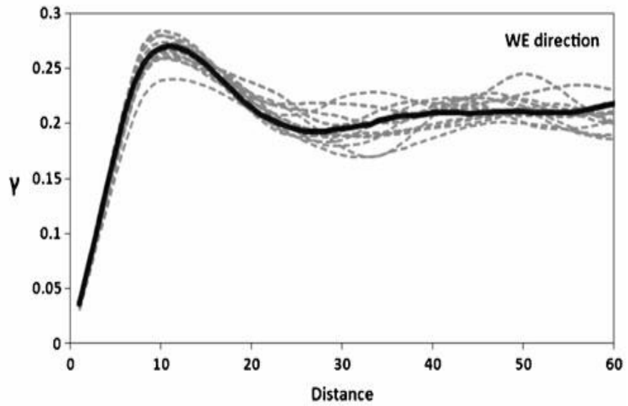
Quantitative Check of the Accuracy



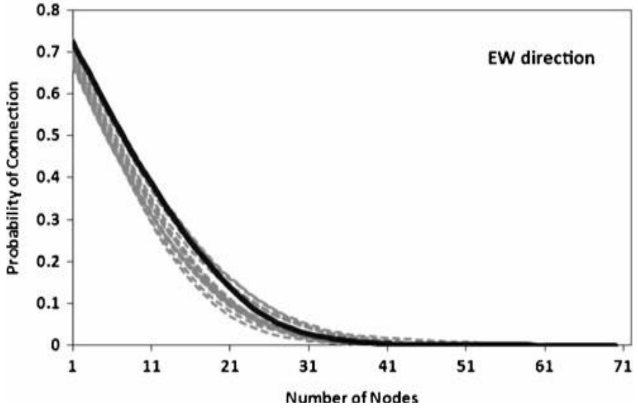
Variogram
in x direction



Variogram
in y direction



MPC



Effective permeability

	$K_e\ (x)$	$K_e\ (y)$
Maximum	32.26	0.73
Upper quartile	29.36	0.29
Median	27.64	0.19
Lower quartile	14.18	0.17
Minimum	1.52	0.15
Mean	27.85	0.32
Variance	3.23	0.07

Limitations of the Algorithm

- **CPU time**
  - **It is not fast enough for multi-million cell 3D grids and DIs**

- **Patchiness**
  - **The problem remains in the case of continuous properties**

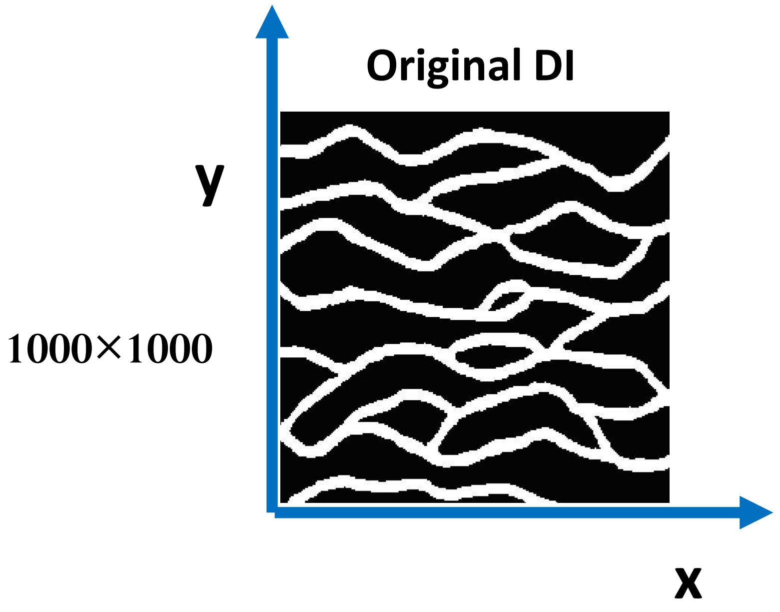
  - **The method was accelerated by carrying out most of the computations in the Fourier space**

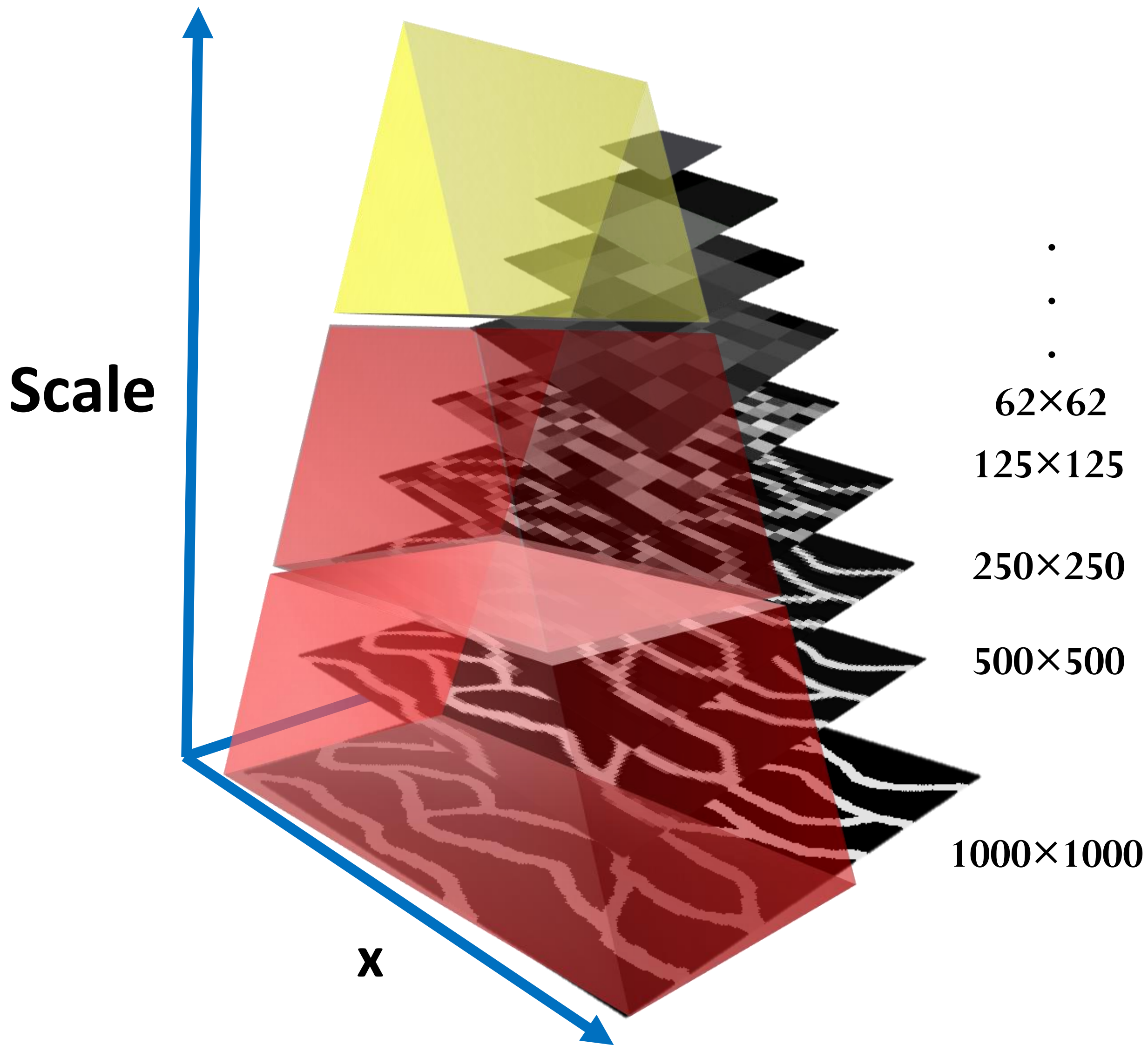
- Tahmasebi, Sahimi & Caers, *Computers & Geosciences* **67**, 75 (2014)
- Tahmasebi & Sahimi, *Water Resources Research* **52**, 2015WR017806; **52**, 2015WR017807 *(2016)*

Multiscale CC Simulation (MS-CCSIM)

- Most of the computations are for the cross-correlation function
- Cross-correlation functions are computed between the overlaps and DI
- The high-resolution DI can be transformed into a pyramid of consecutively coarsened views of the same image
- The pyramid allows for rapid search of the matching patterns

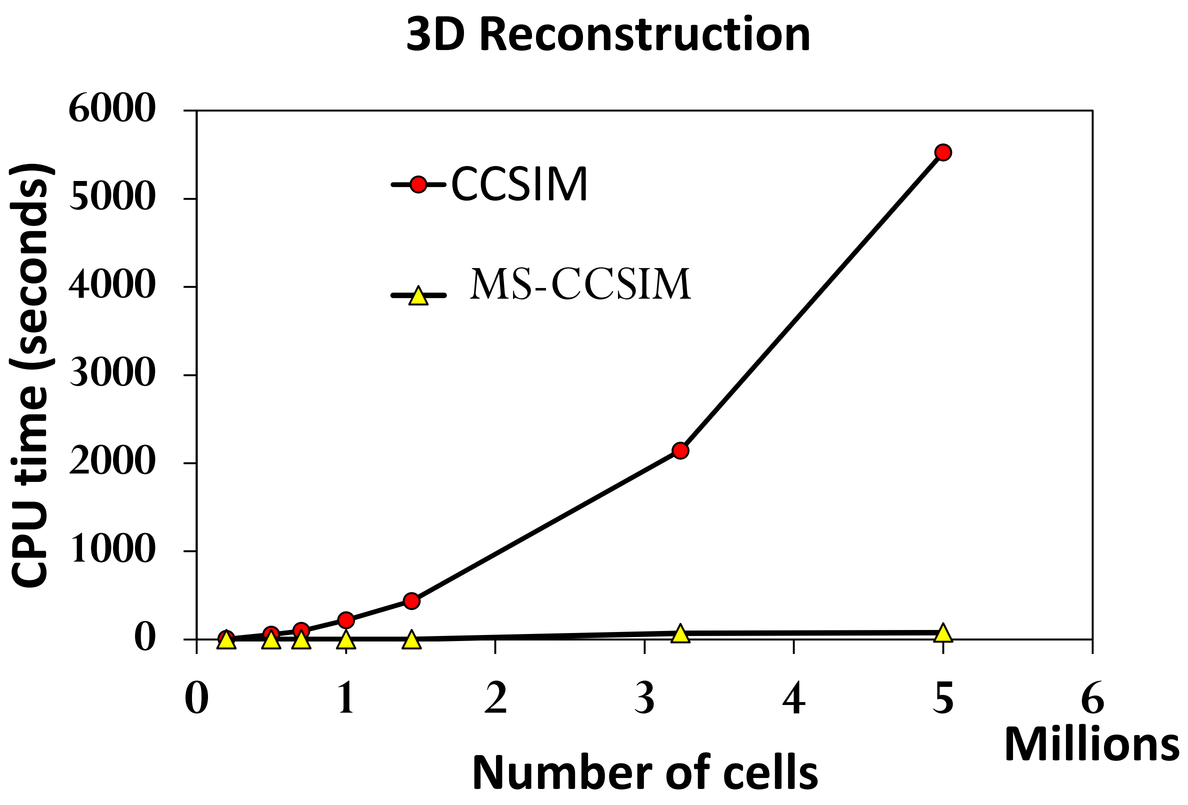
Addressing the CPU Issue: Constructing DI at Various Scales



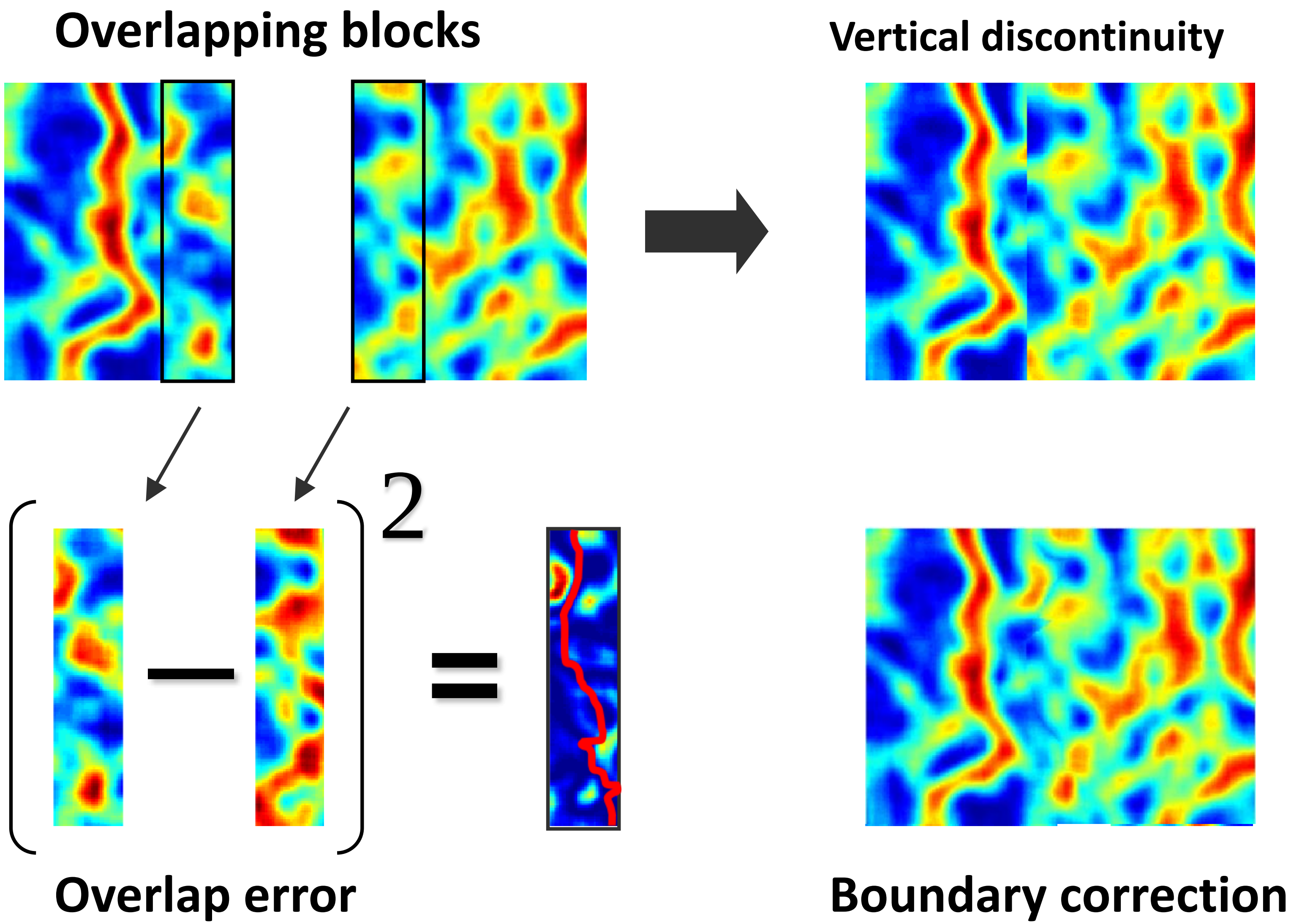


✓ The rescaled DI can be obtained by several methods

CPU Improvement

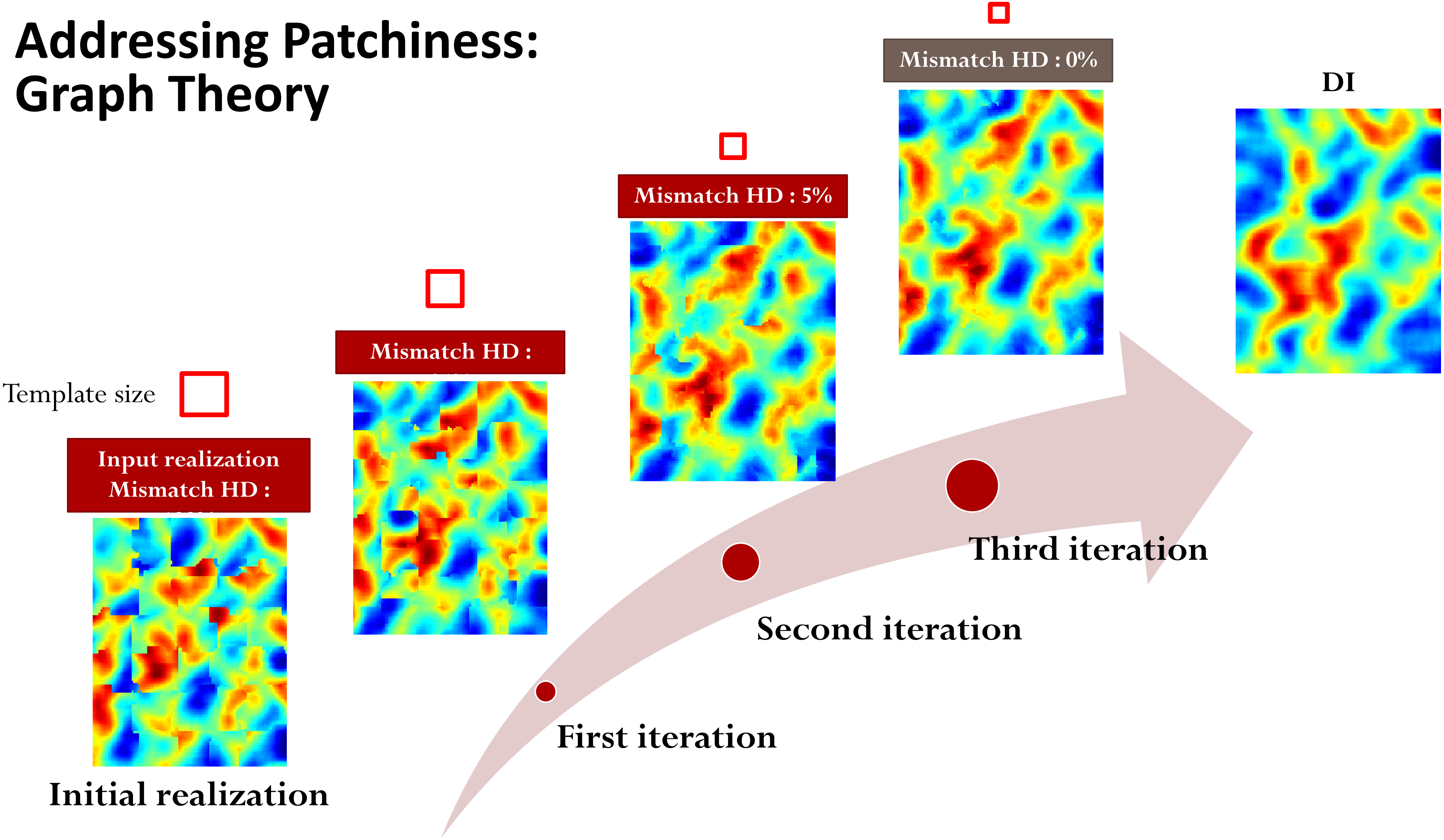


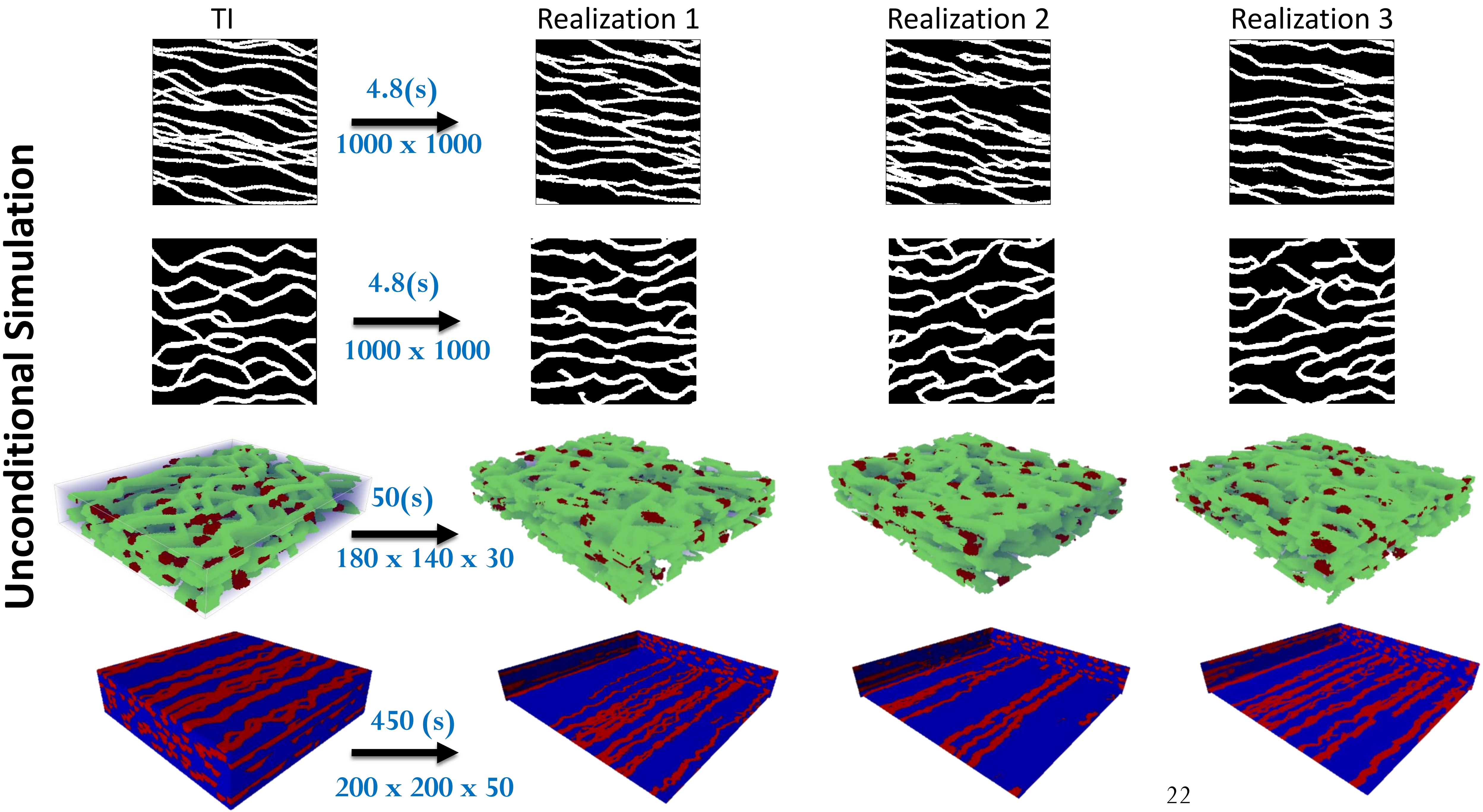
Addressing Patchiness: Graph Theory



Tahmasebi & Sahimi, *Water Resources Research* **52**, 2015WR017806; **52**, 2015WR017807 *(2016)*

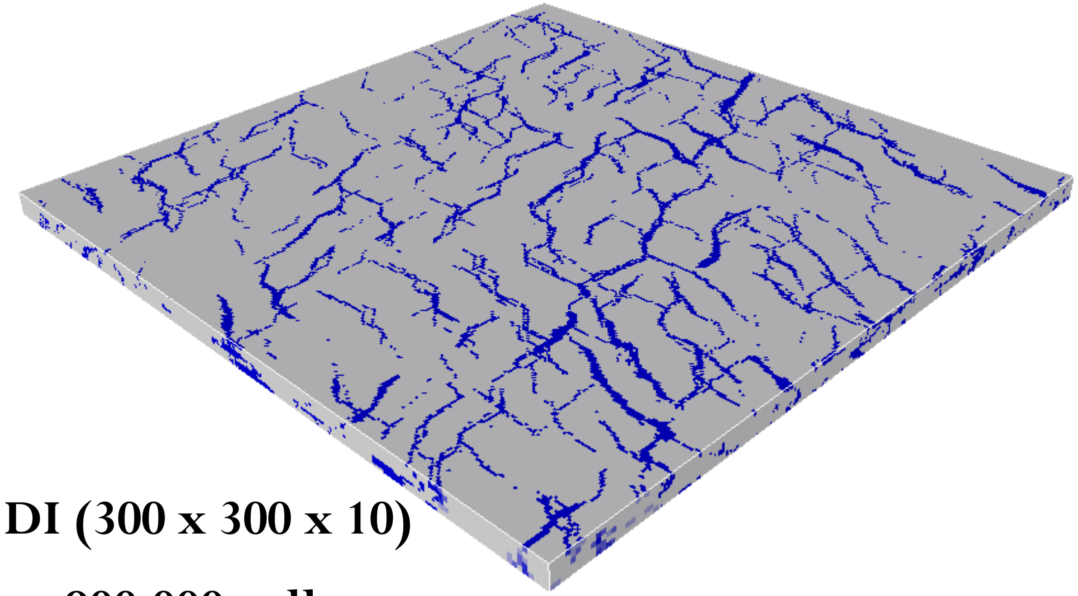
Addressing Patchiness: Graph Theory





Unconditional Simulation
TI
Realization 1
Realization 2
Realization 3
4.8(s)
1000 x 1000
4.8(s)
1000 x 1000
50(s)
180 x 140 x 30
450 (s)
200 x 200 x 50

Fracture Pattern

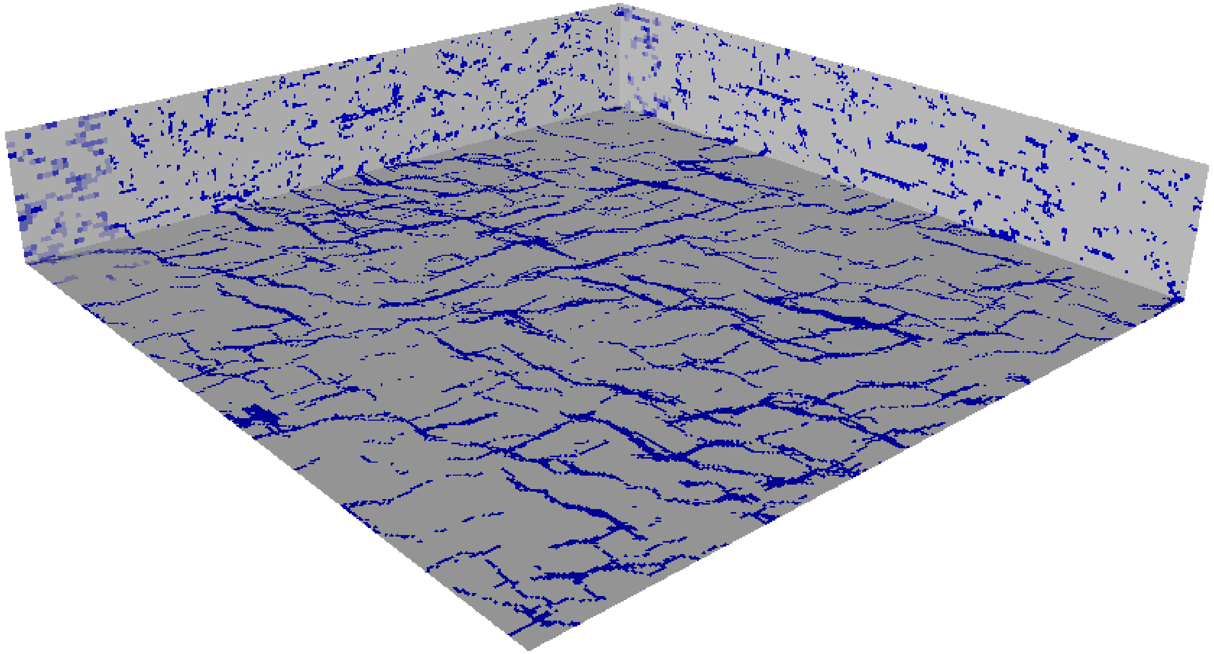


DI (300 x 300 x 10)

900,000 cells

Reconstructed (300 x 300 x 50)

4,500,000 cells

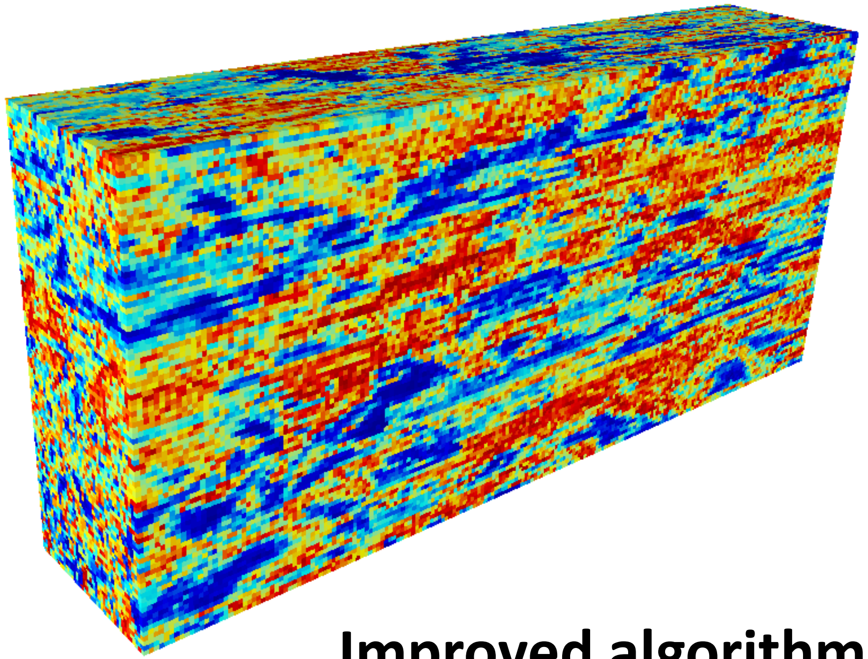


CPU time: 7 (s)

(Data from JAPEX)

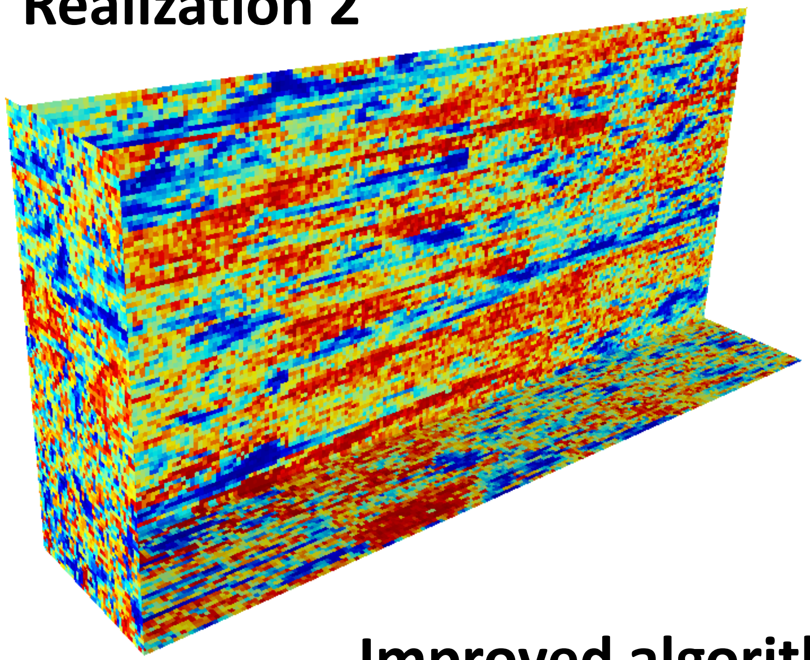
Highly Heterogeneous Material

Realization 1



Improved algorithm

Realization 2

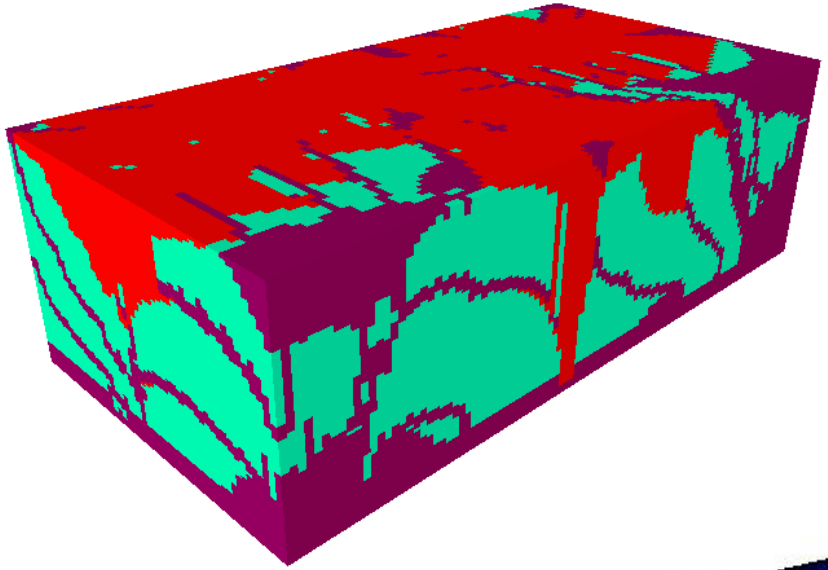


Improved algorithm

Reconstructed (200 x 100 x 40)

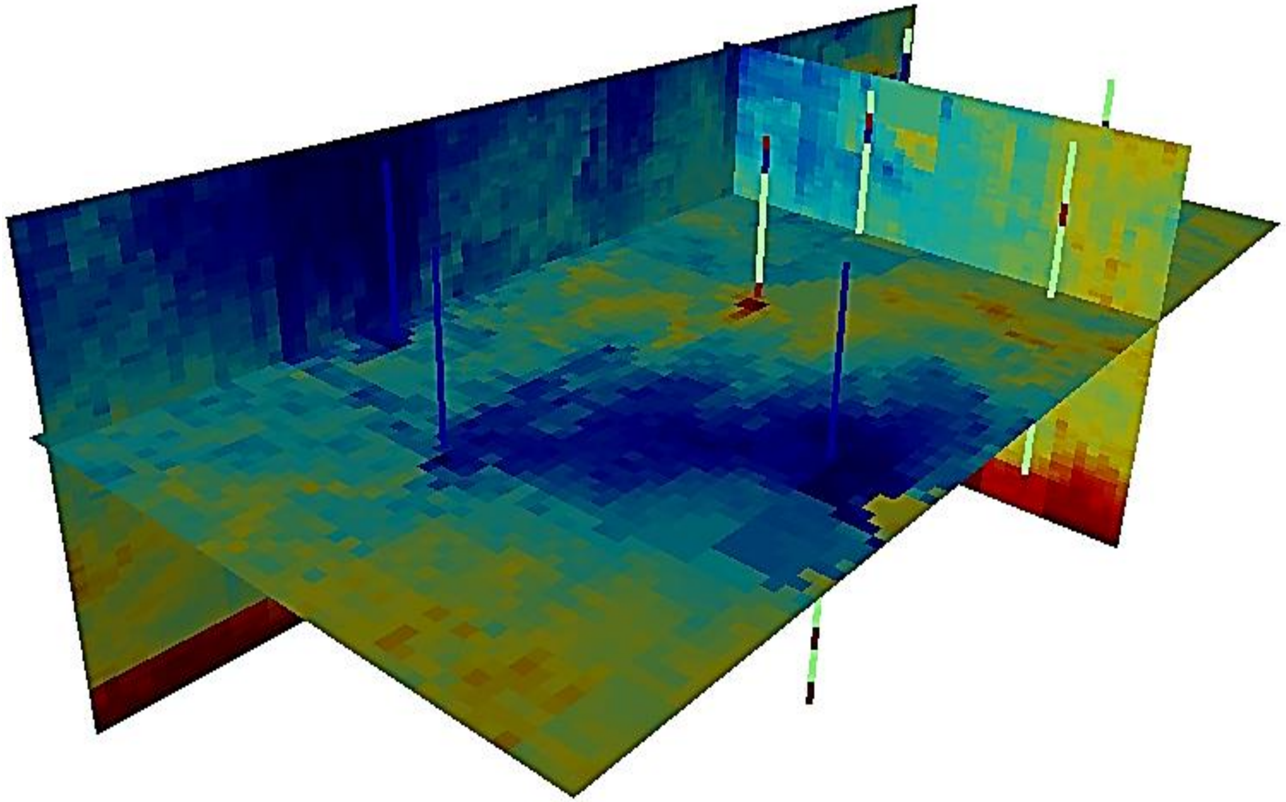
CPU time: 30 (s)

3D Conditional Simulation



DI

Ensemble average



(Data from ExxonMobil)

Integration of Several Types of Data: Each Dataset Has its Own Particle CC Function

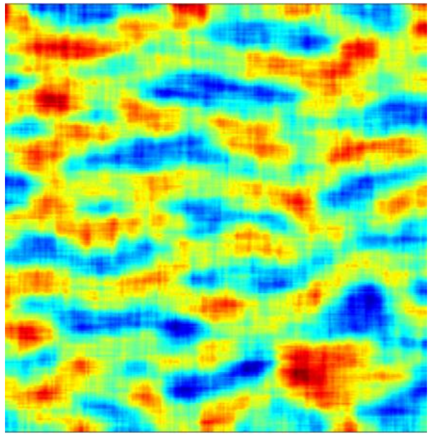
- **Data from various sources are integrated using the CCSIM algorithm**

$$\boldsymbol{C_{overall}(i,j) = C_{DI}(dev_T, DI) + \sum_{m=1}^{n} \omega_m C_{mDI}(\mathrm{m}dev_T, mDI)}$$

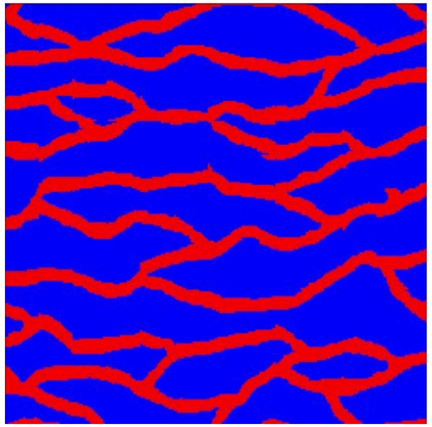
- **Data integration helps reducing the uncertainty and use the available information more effectively.**

Tahmasebi and Sahimi, *Transport in Porous Media* 107, 871-905 (2015)

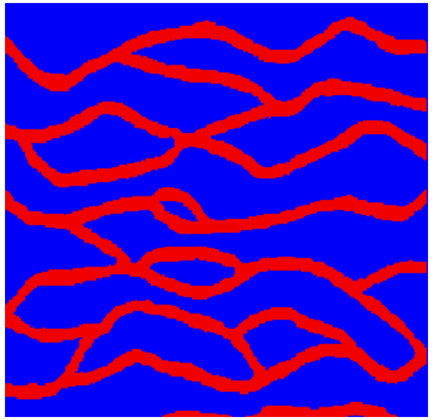
Binary Morphology



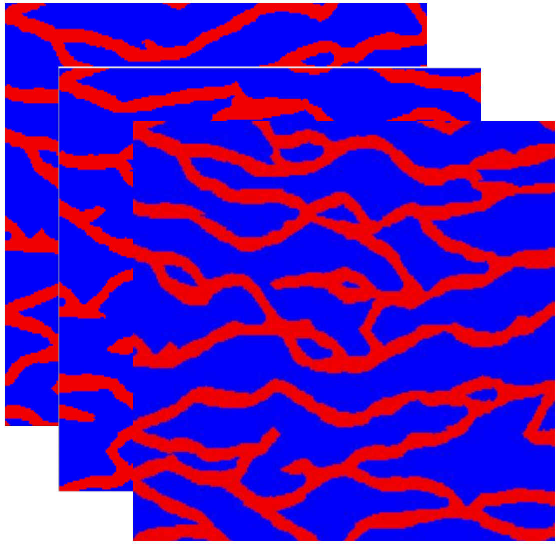
Soft data



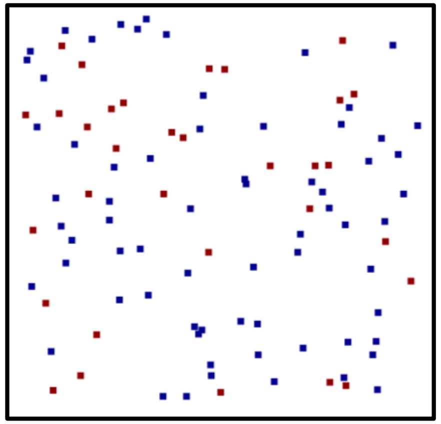
True model



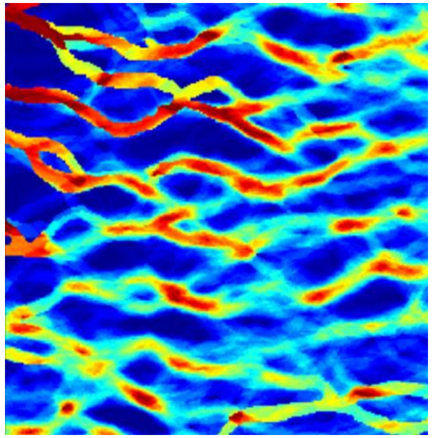
DI



Realization

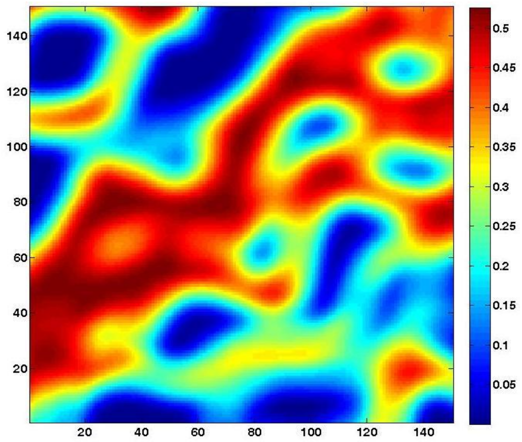


Hard data

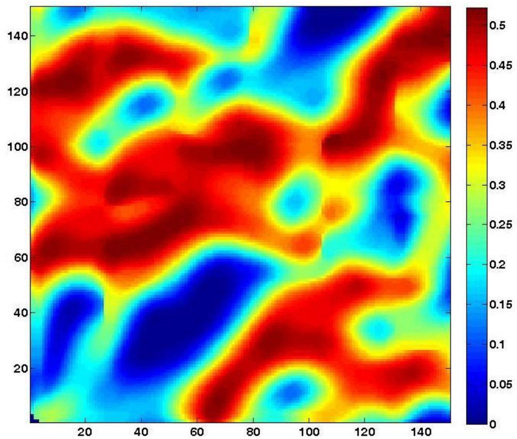


Ensemble average

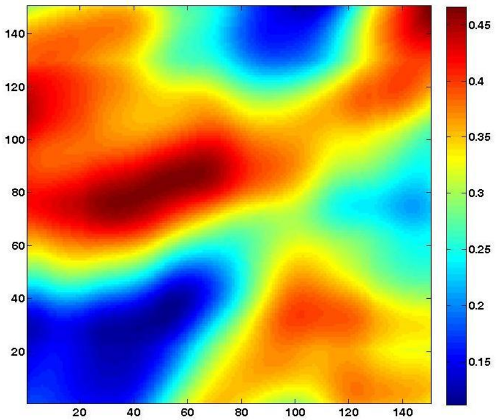
Continuous Morphology



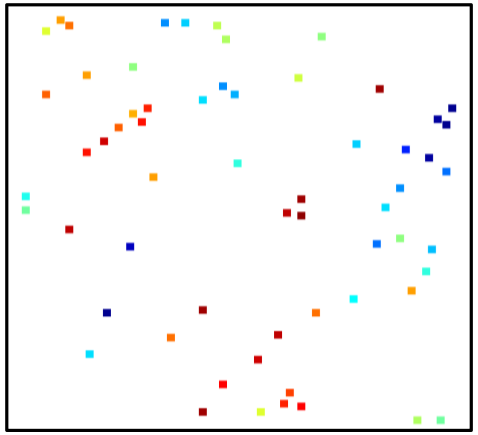
DI



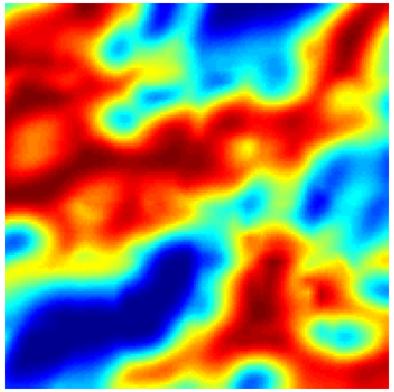
Reference Image (RI)



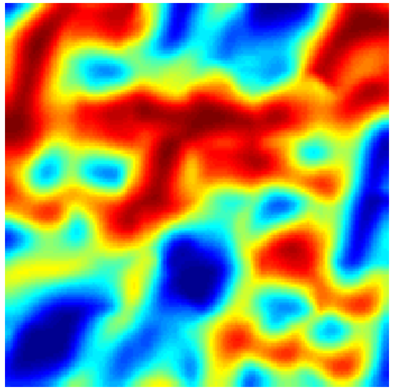
Soft data



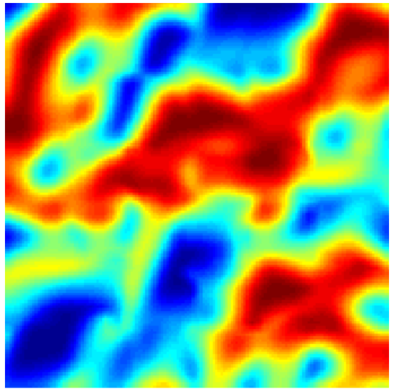
Hard data



Realization #1

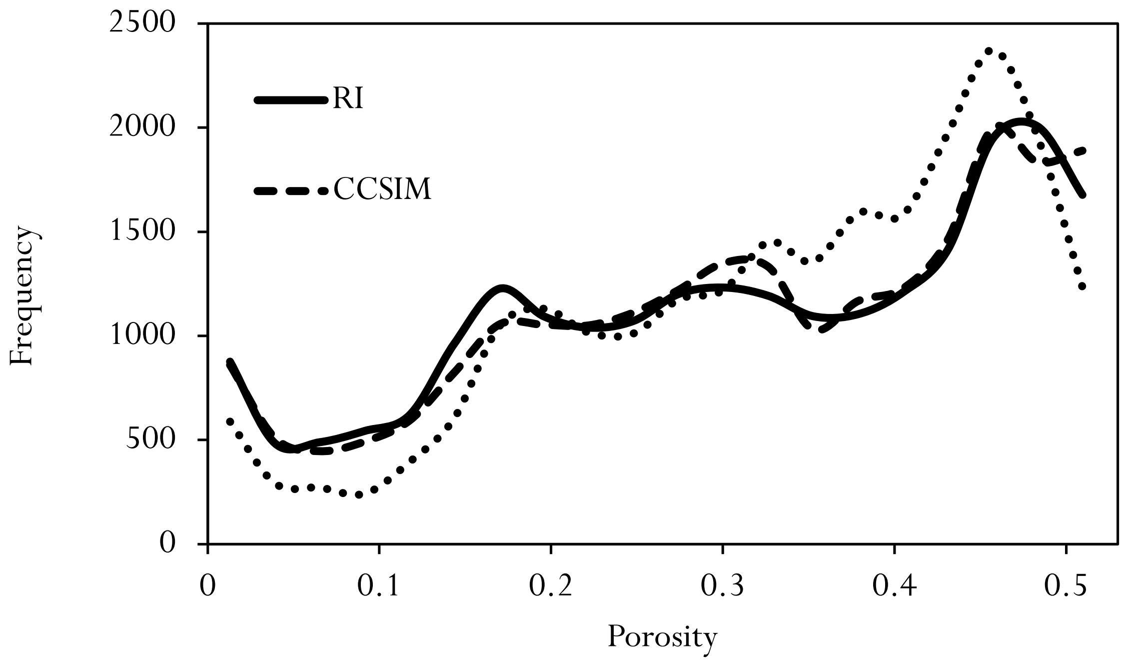


Realization #2

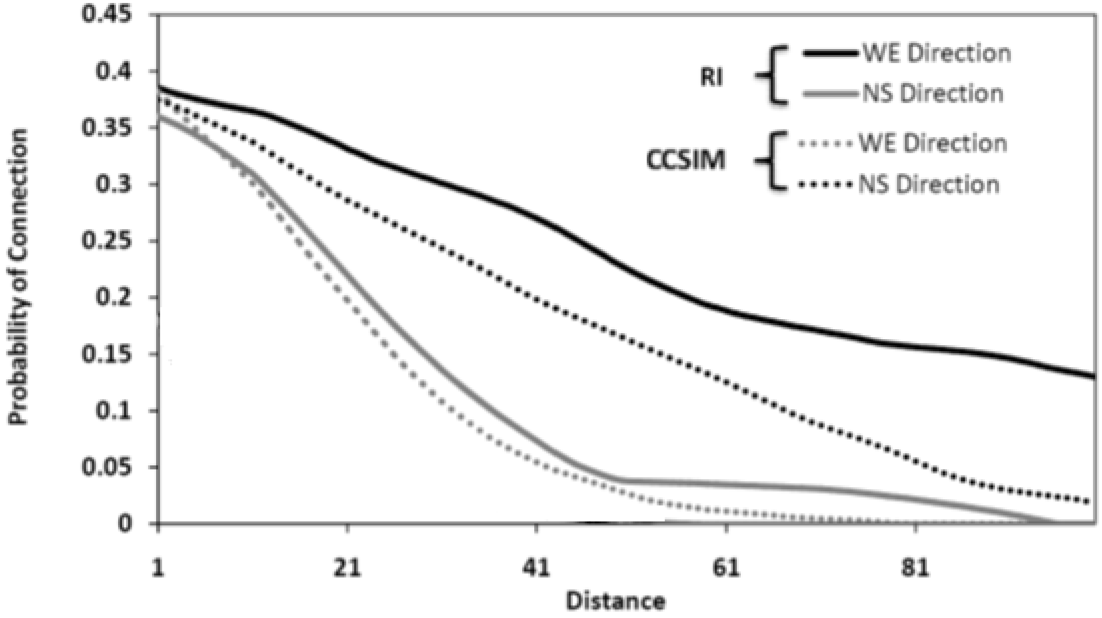


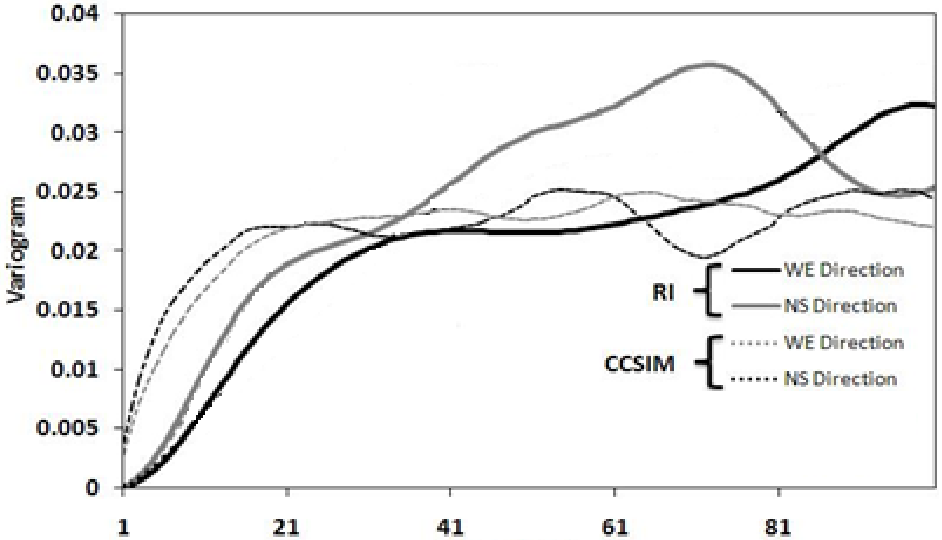
Realization #3

Porosity distribution



Multiple-point Connectivity

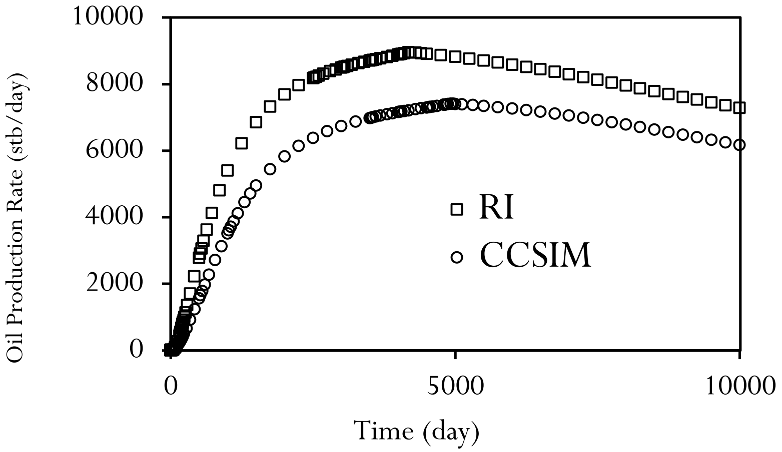


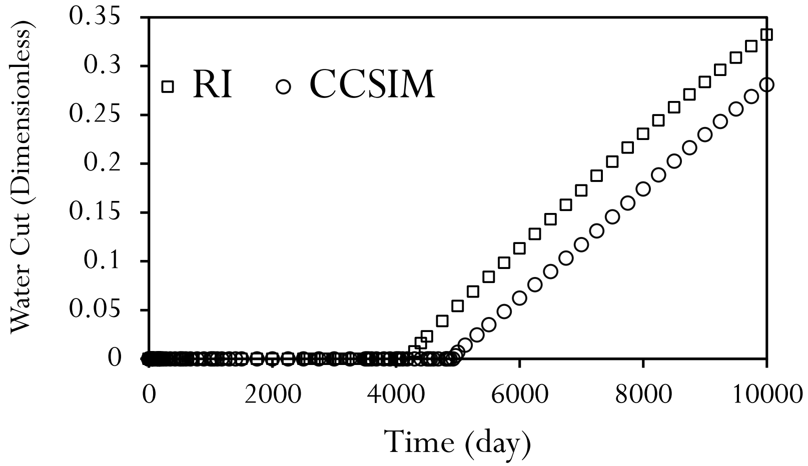


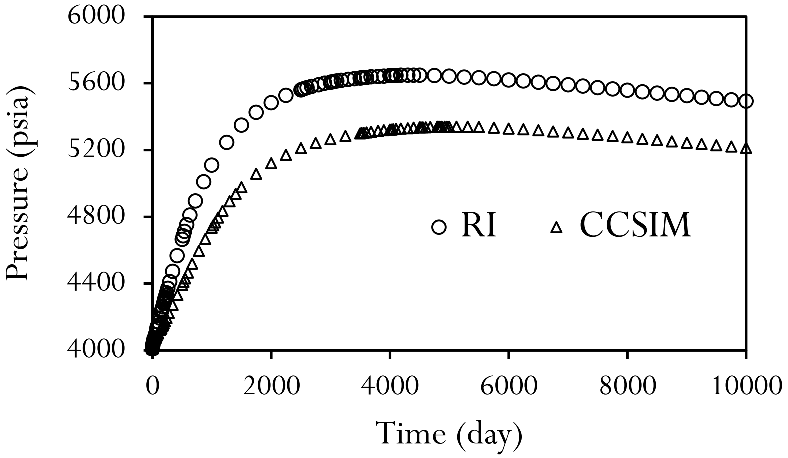
Statistical summery

Model	Mean	Variance	Maximum	Median
RI	0.412	0.024	0.842	0.416
CCSIM	0.406	0.022	0.84	0.403

Two-Phase Flow

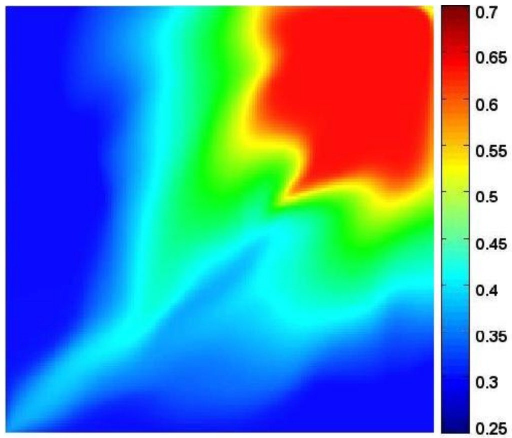


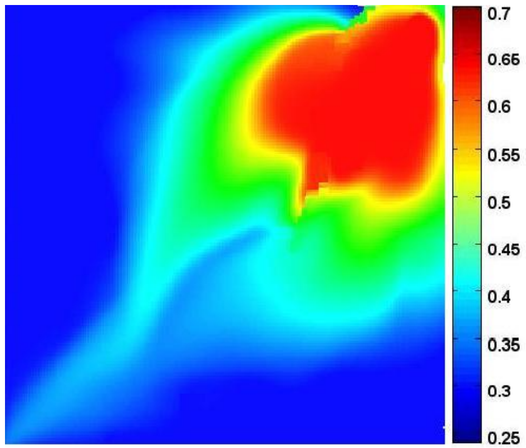




Distribution of water saturation

RI





CCSIM

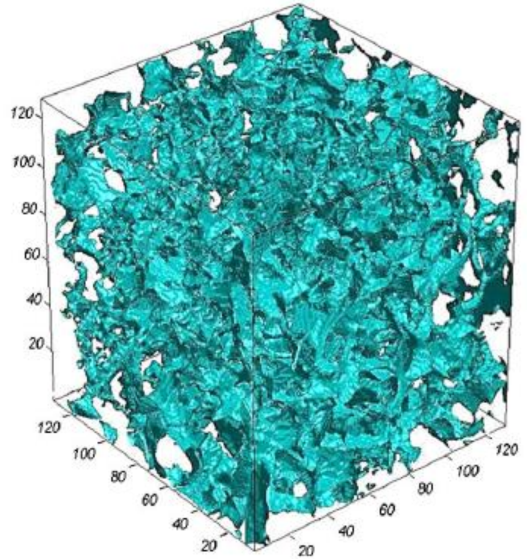
Long-Standing Problem: 2D to 3D Reconstruction

Generate a high quality 3D sample from a single 2D thin section

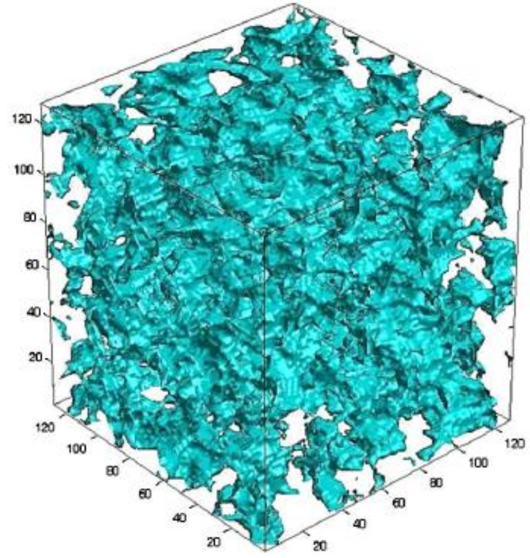
- **CCSIM is well suited for a sample with high entropy (heterogeneity)**
- **First, the external surface is reconstructed (conditional CCSIM)**
- **Next, the 3D medium is reconstructed layer-by-layer (plane-by-plane)**

Tahmasebi and Sahimi, *Physical Review E* **85**, 066709 (2012)
Tahmasebi and Sahimi, *Physical Review Letters* **110**, 078002 (2013)

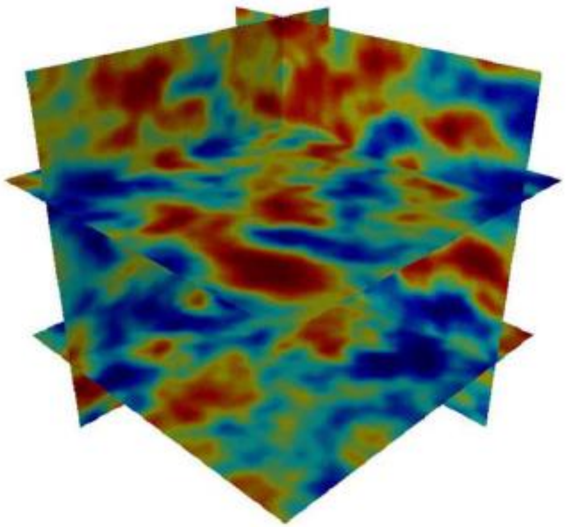
Original



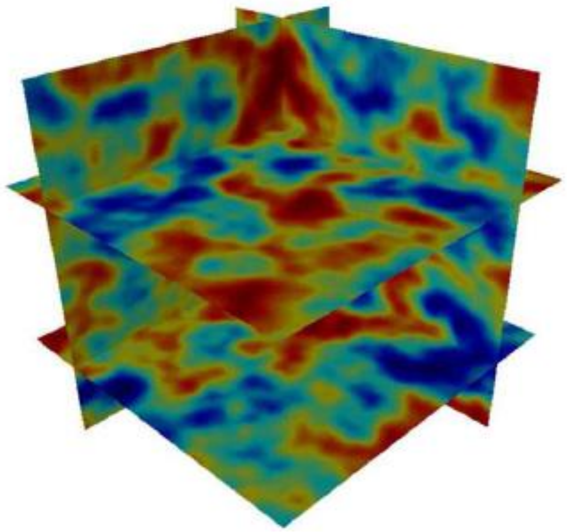
Reconstructed



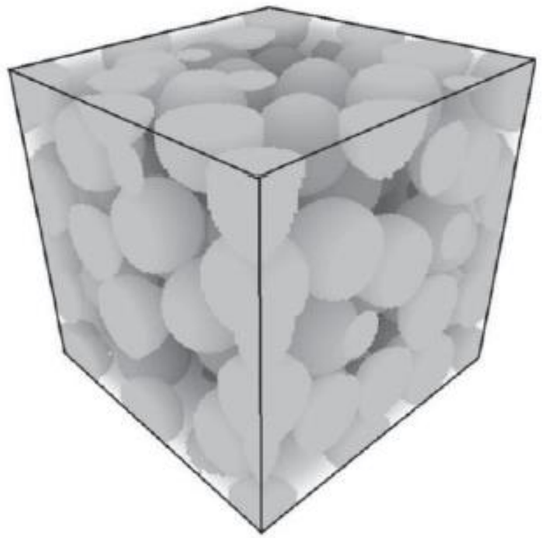
Original



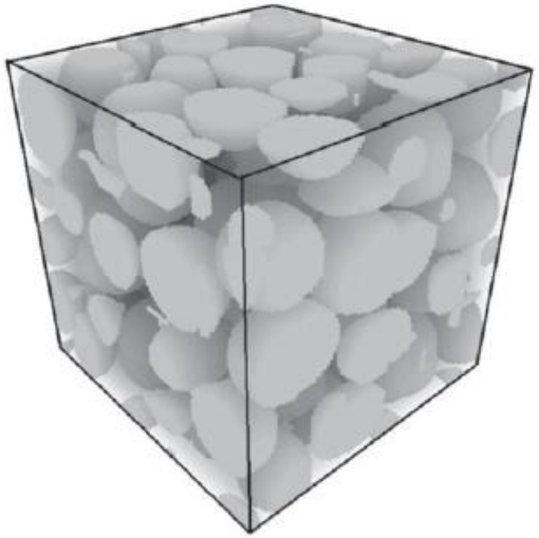
Reconstructed



Original

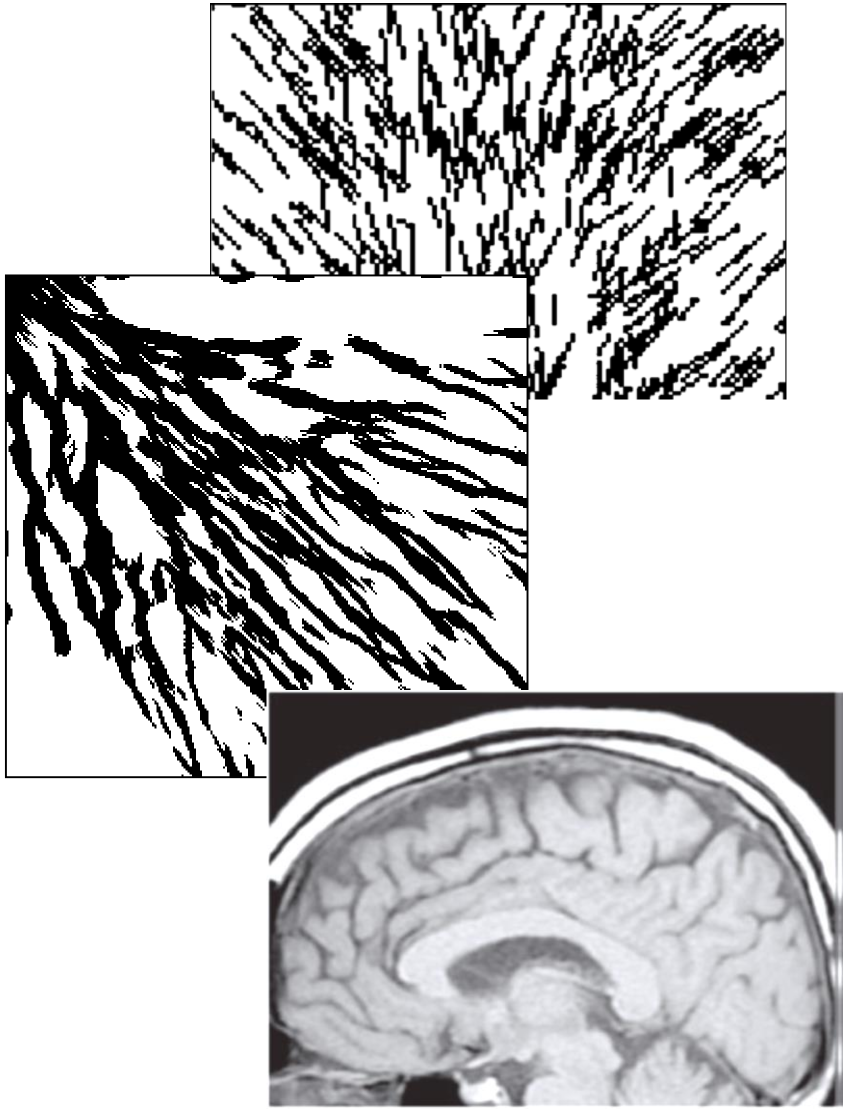


Reconstructed



Non-Stationary Media

- **Practically every large-scale porous medium is non-stationary**
- **By non-stationary we mean that the probability distribution functions of various properties vary spatially**
- **How do we reconstruct such porous media?**



Tahmasebi and Sahimi, *Physical Review E* **91**, 032401 (2015)

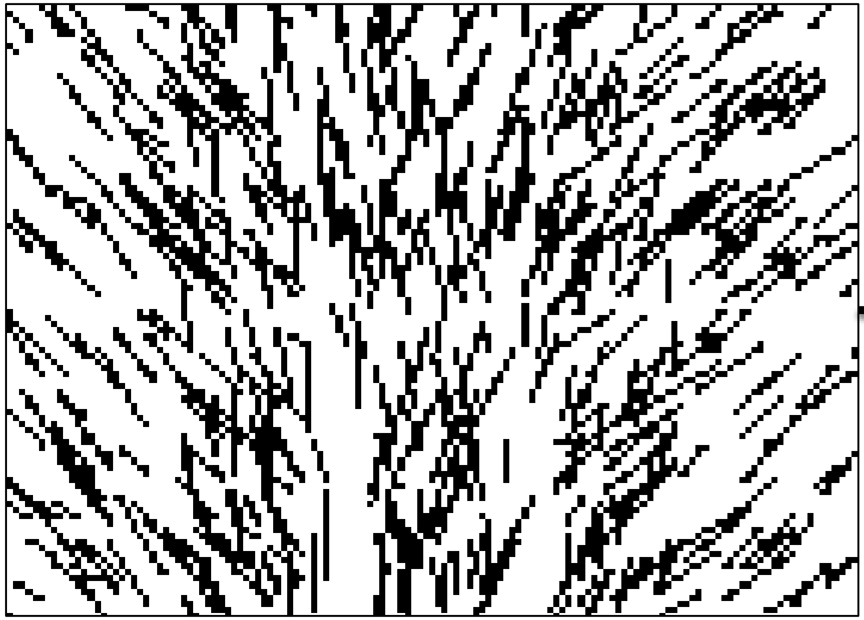
Approach 1: Watershed Transform

- **Watershed transforms construct a gradient image (instead of working with the image or datasets)**

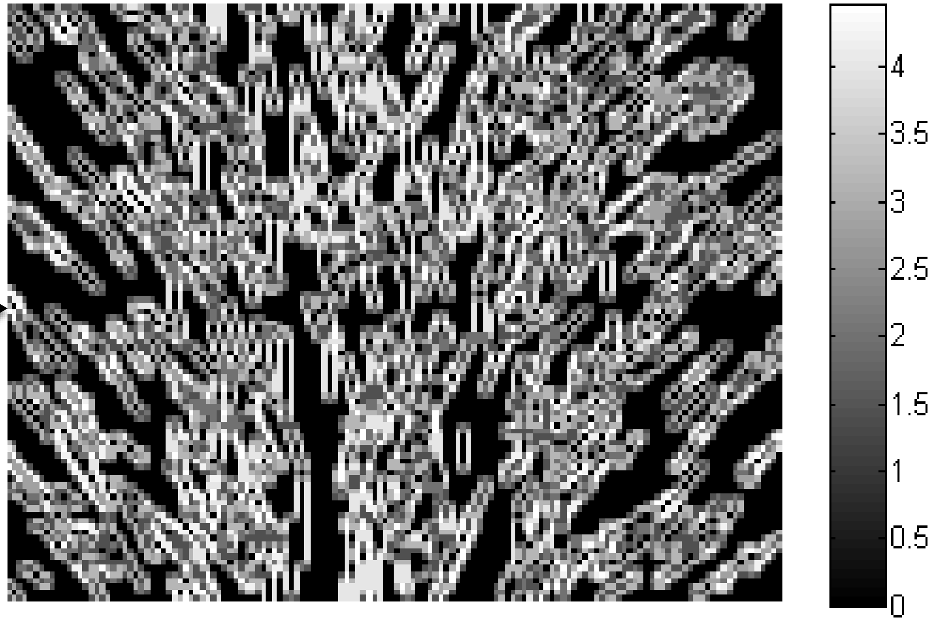
- **That is, a new image is constructed based on the local gradients between neighboring points in the original image or data**

Tahmasebi and Sahimi, *Physical Review E* **91**, 032401 (2015)

Input image

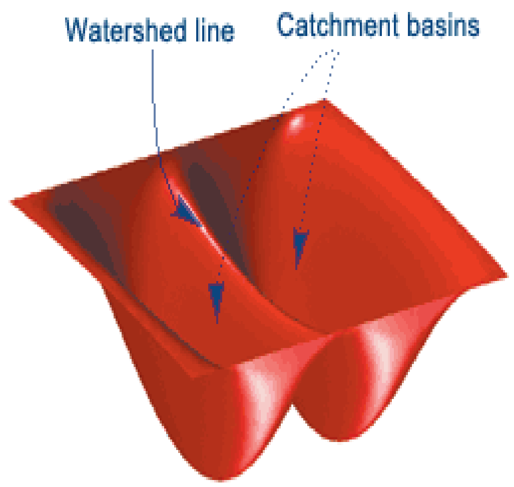


Gradient image

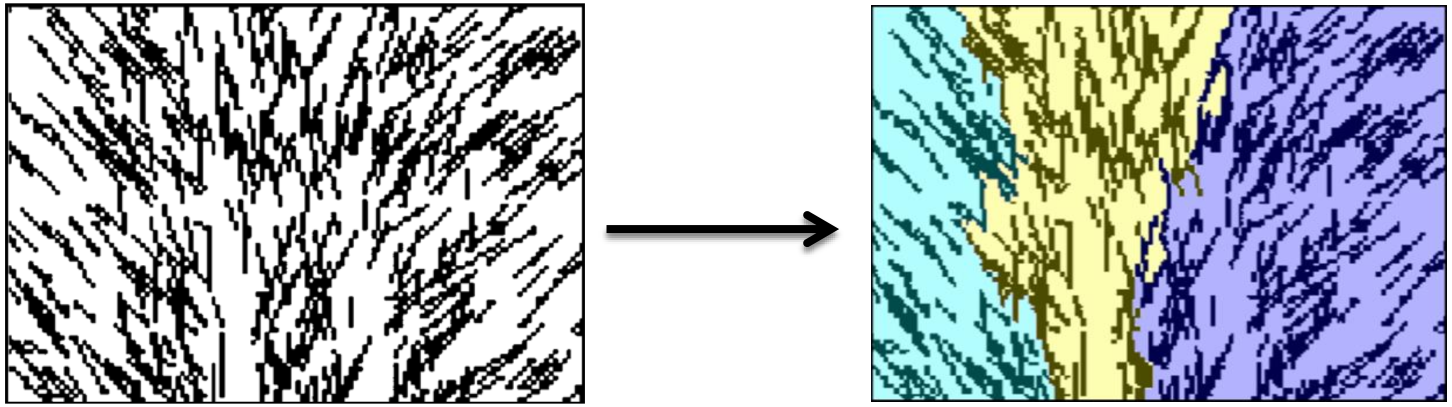


Watershed Transform

- **Three types of points**
  - Points belonging to a regional minimum
  - Catchment basin / watershed ***of a regional minimum***
    - Points at which a drop of water will certainly fall to a single minimum
  - Divide lines / Watershed lines
    - Points at which a drop of water will be equally likely to fall to more than one minimum
    - Crest lines on the topographic surface



Watershed Transforms

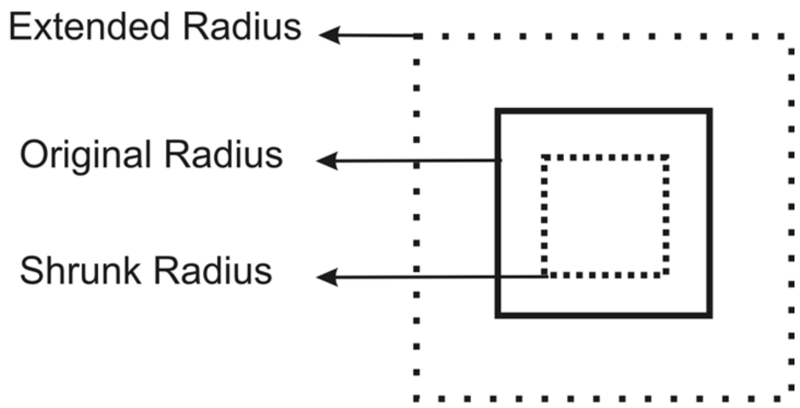


Non-stationary surface

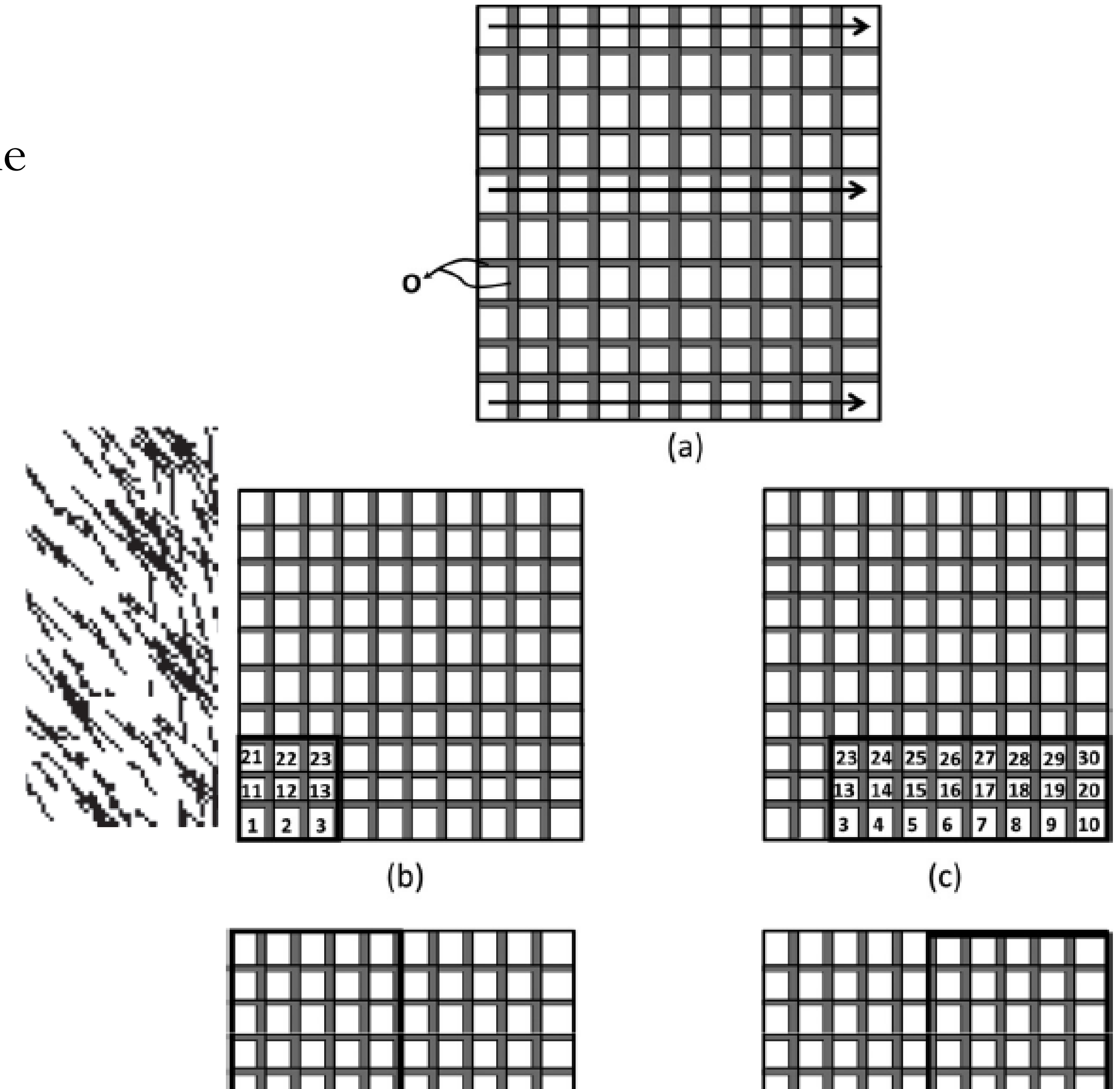
Approach 2: Shannon Entropy

- Start with a radius
- The radius can be extended or shrunk based on the Shannon entropy
- $S = -\sum_{i=1}^{n} p_i \ln p_i$

p_{i} = (histogram of sample *i*)/(length of the sample)

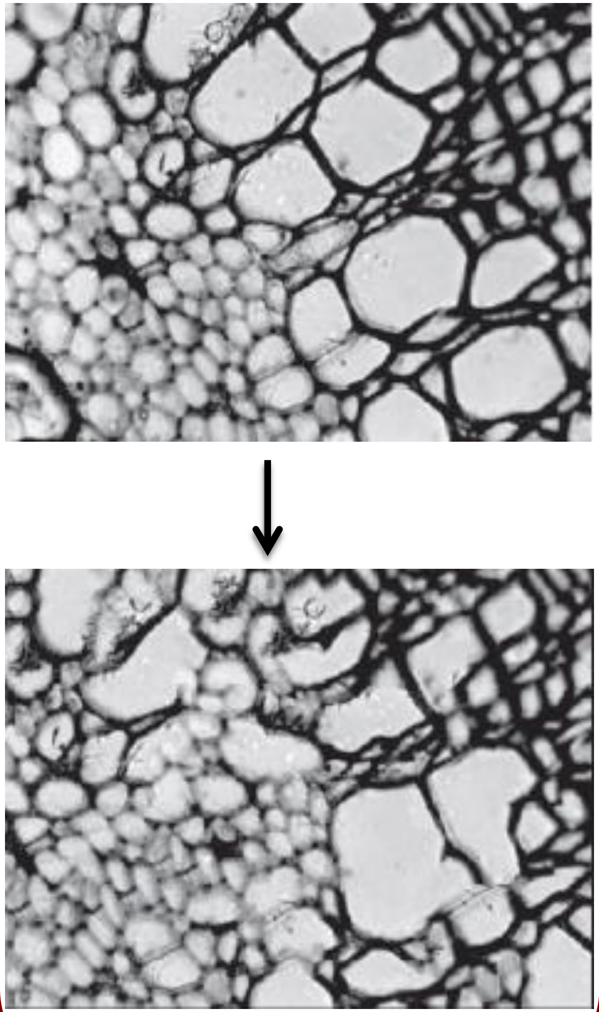


Non-stationary surface

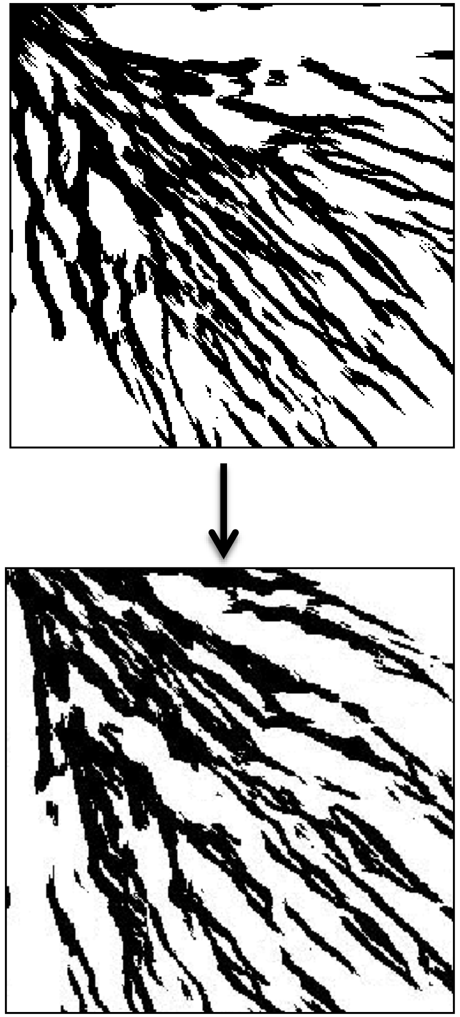


Some Non-stationary Examples

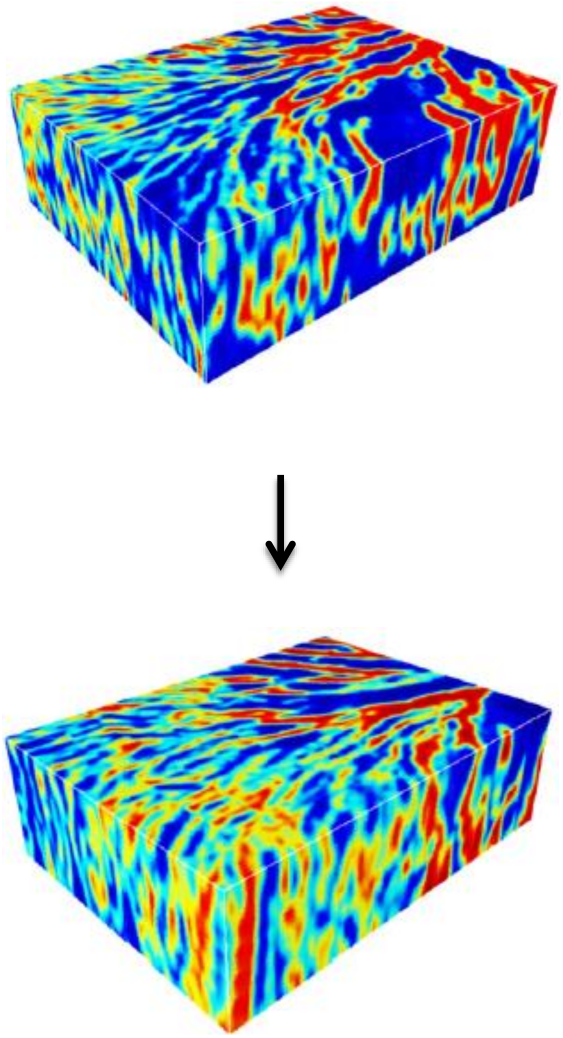
Dicot wooden stem



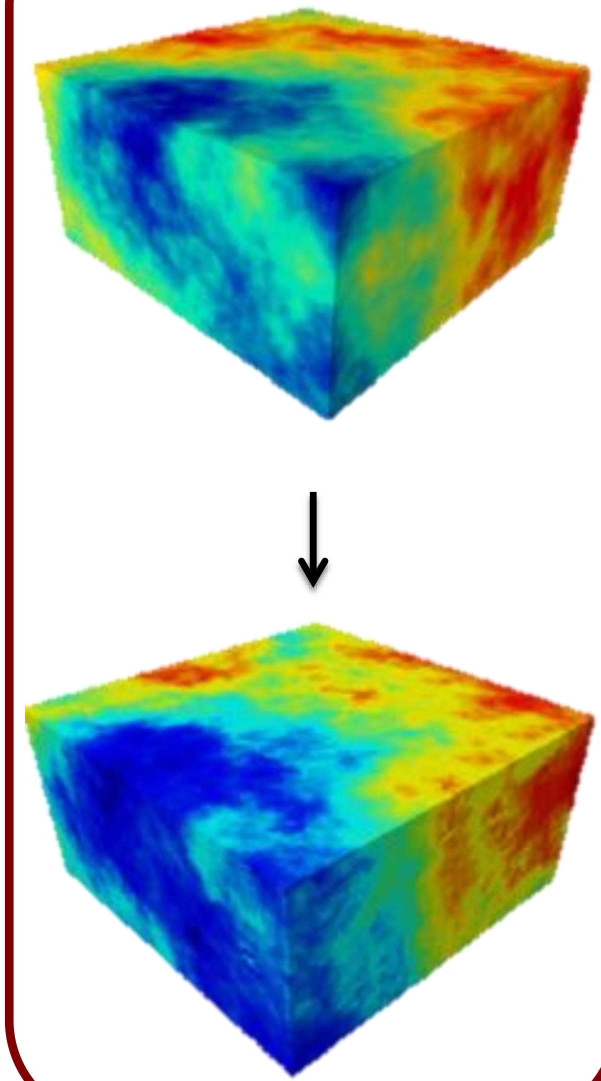
Delta System



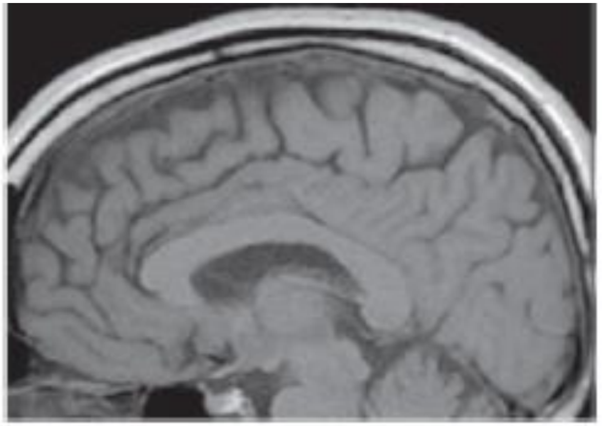
3D Delta system



3D FBM

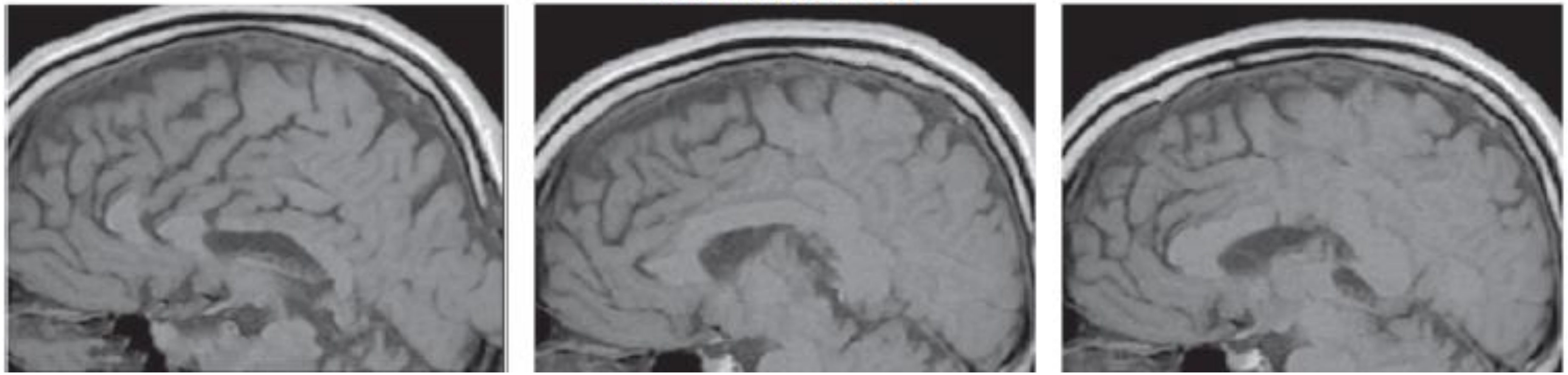


Application to Medical Imaging



Brain MRI

Reconstructed

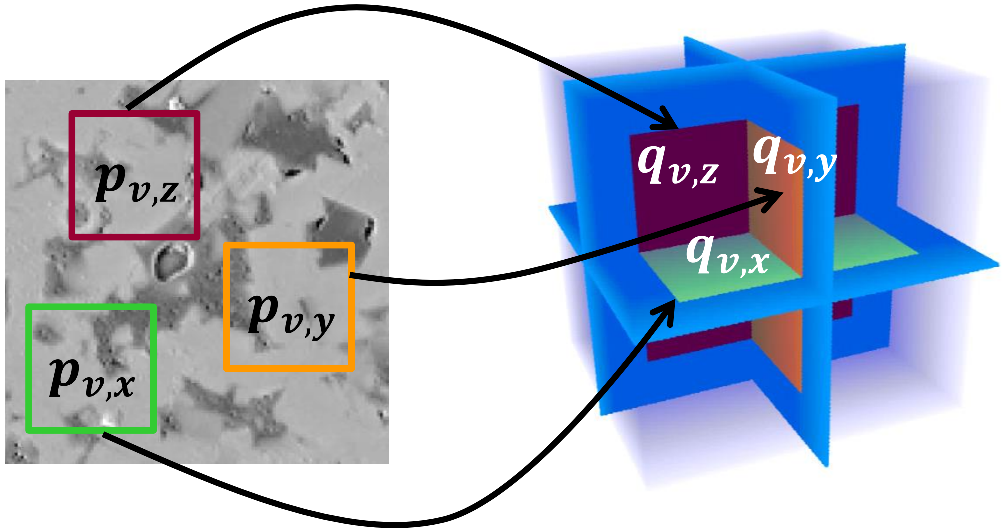


Application to Modeling of Shale Formations

- **The main framework of the new method is the CCSIM**
- **CCSIM is well suited for a sample with high entropy (heterogeneity)**
- **But, shales manifest low entropy (low disorder)**
- **To deal with a low entropy image, a histogram matching is included to honor the one moments statistical properties**

Iterative Algorithm

- The neighborhoods in X, Y and Z directions are used to find their corresponding matches in the input image.



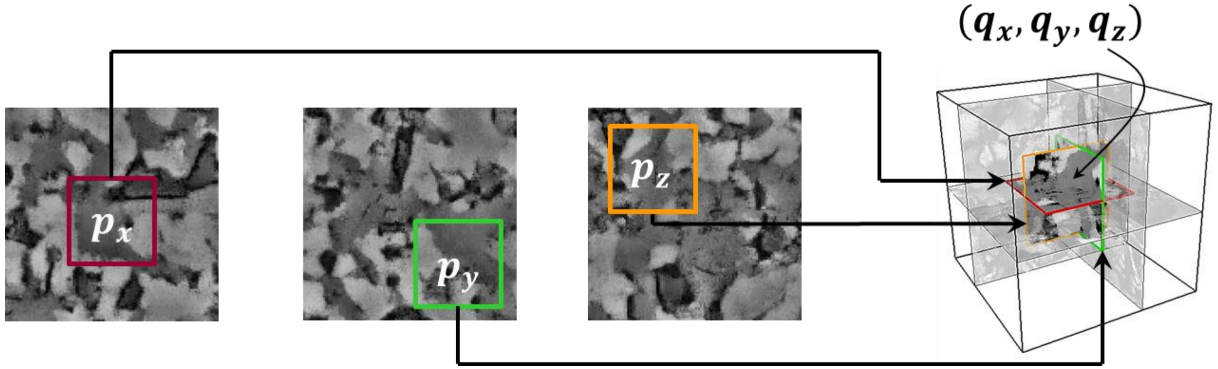
- Difference between the patterns in the generated model and DI is calculated.

- The final selected 3D pattern should minimizes the energy function of final model.

$$E(q,\{p\}) = \sum_{v} \sum_{i\in\{x,y,z\}} \left\| q_{v,i} - p_{v,i} \right\|^{\gamma}$$

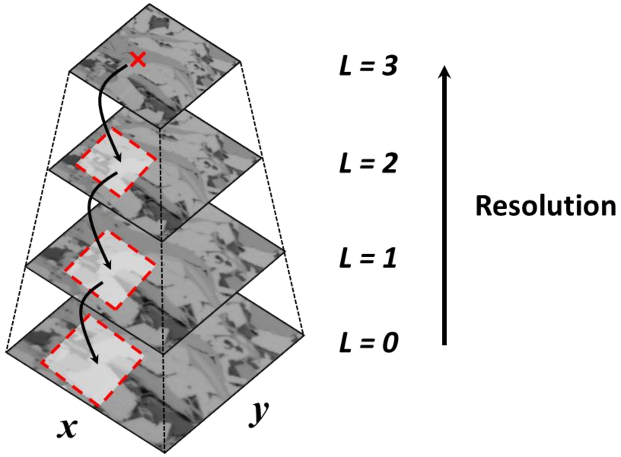
$$= \sum_{v} \sum_{i\in\{x,y,z\}} \sum_{u\in N_i(v)} \left\| q_{v,i,u} - p_{v,i,u} \right\|^{2}$$

Using different images for each direction

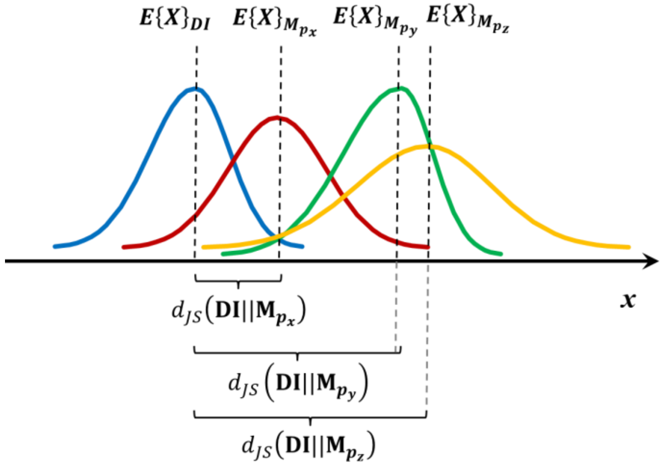


✓ **Use three different images for representing the directional heterogeneities**

Multi-scale approach



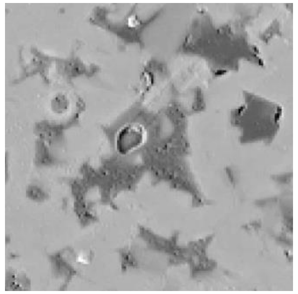
Calculating histogram distances for pattern selection



✓ **For capturing more structural features, a multi-scale algorithm in three levels is used**

✓ **Histogram matching helps reproducing the multi-modal distributions**

Iterative reconstruction

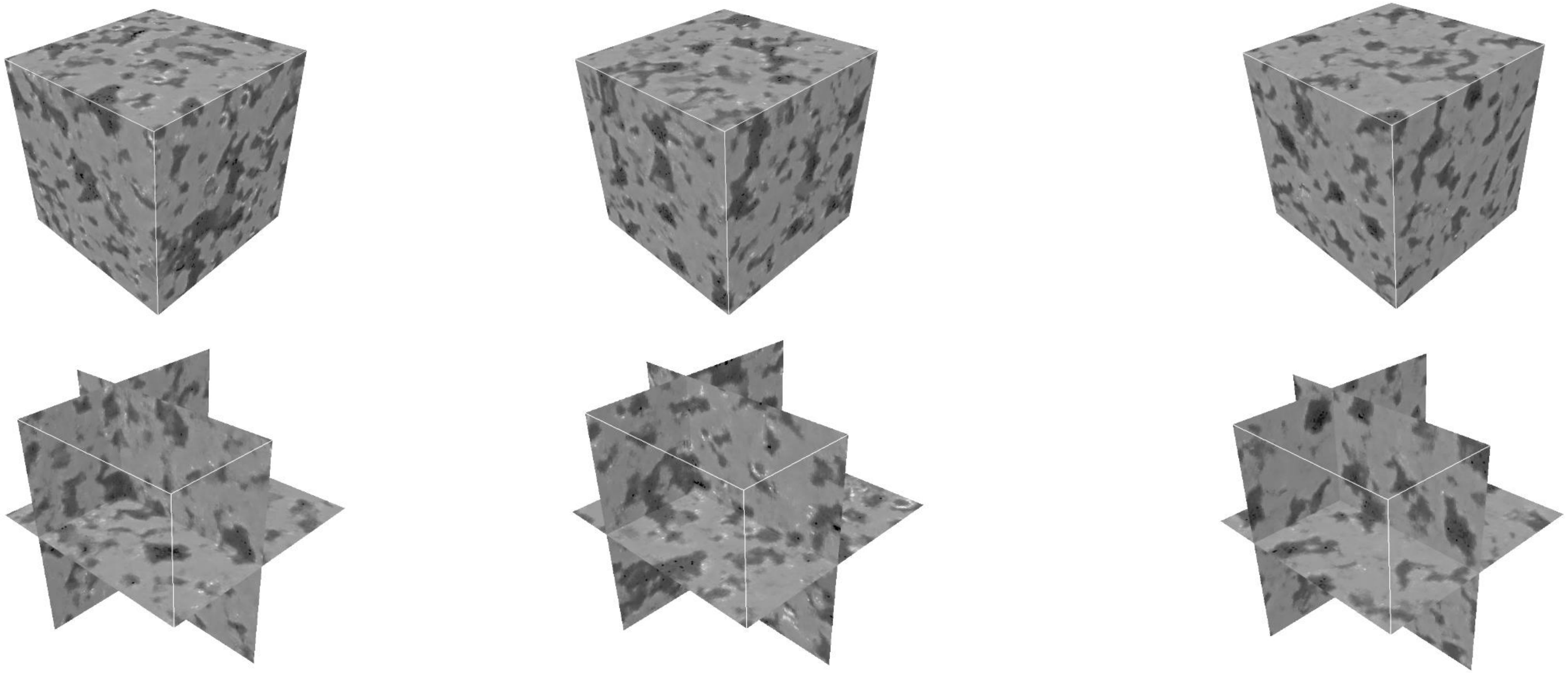


Input 2D image

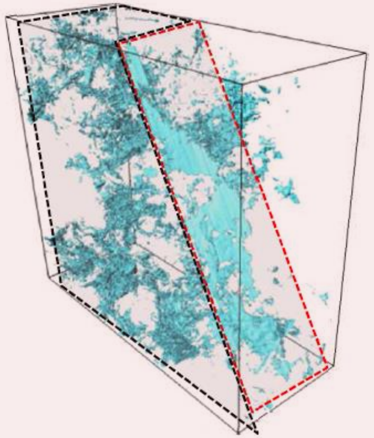
Realization 1 | Realization 2 | Realization 3

Exterior view

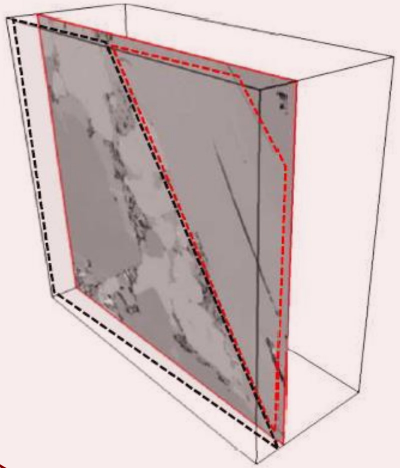
Cross sections



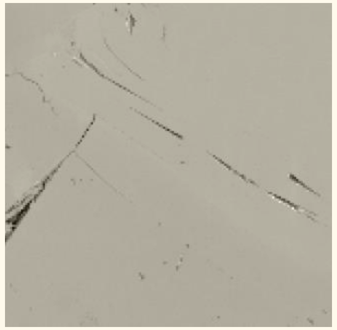
Multiscale Reconstruction of Shale



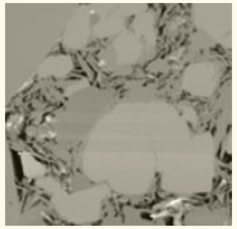
3D view



Large Scale

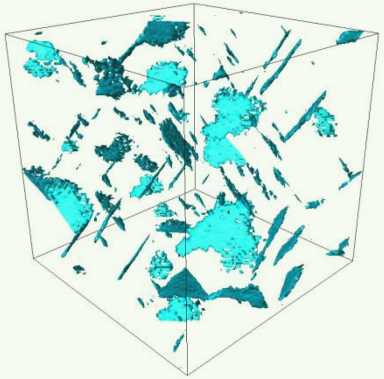


2D DIs

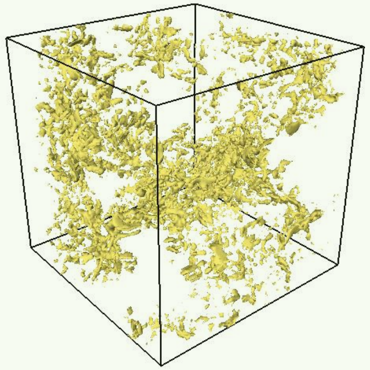


Small Scale

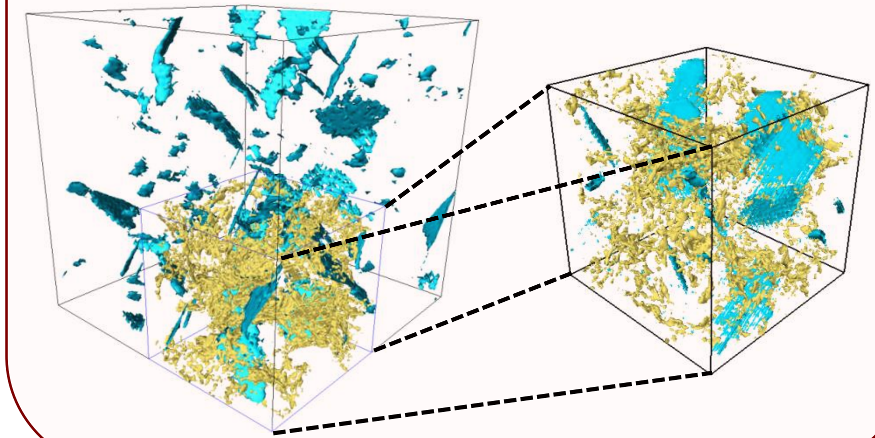
Large-Scale Reconstruction



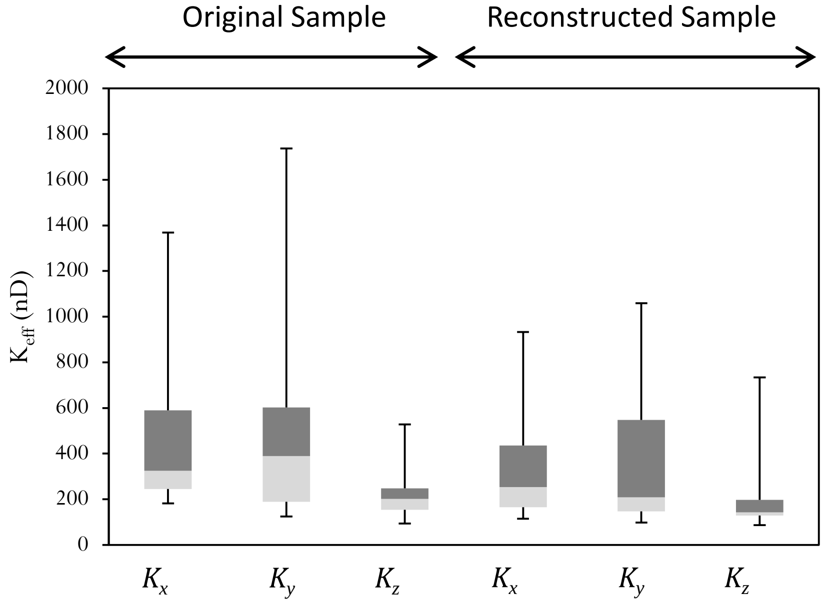
Small-Scale Reconstruction



Integrated small- and large-scale models

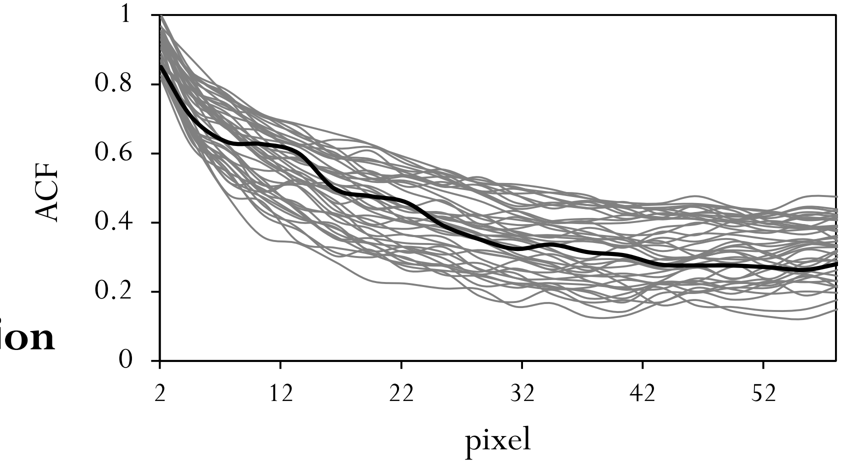


Statistical Comparison

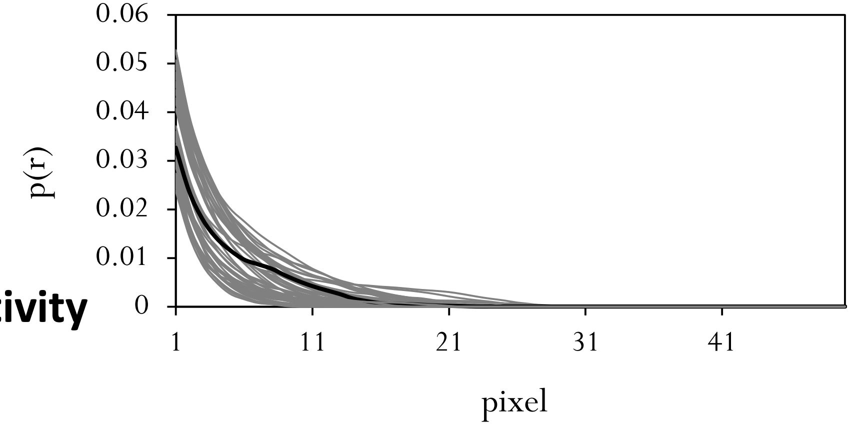


Effective permeability

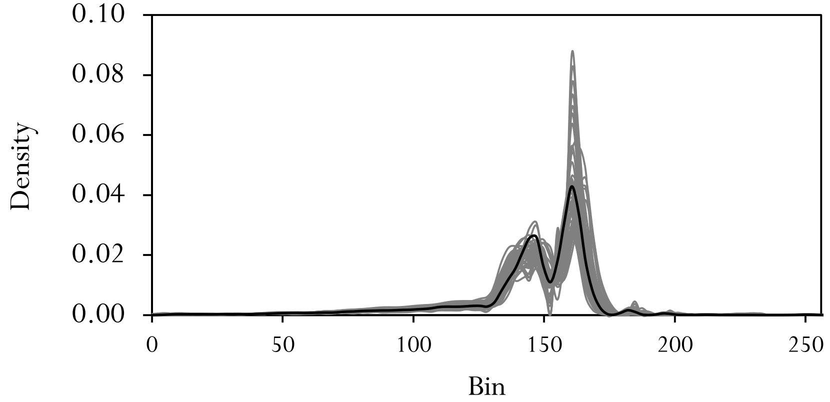
Auto-Correlation Function



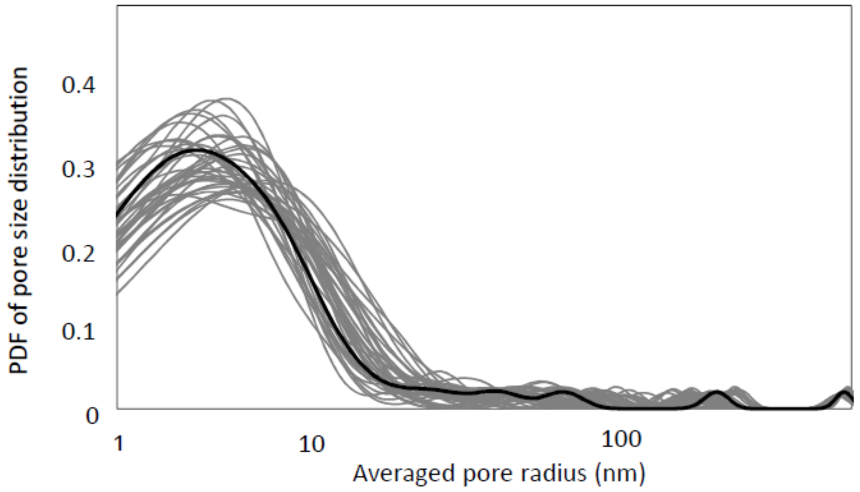
Multiple-point connectivity



Density distribution

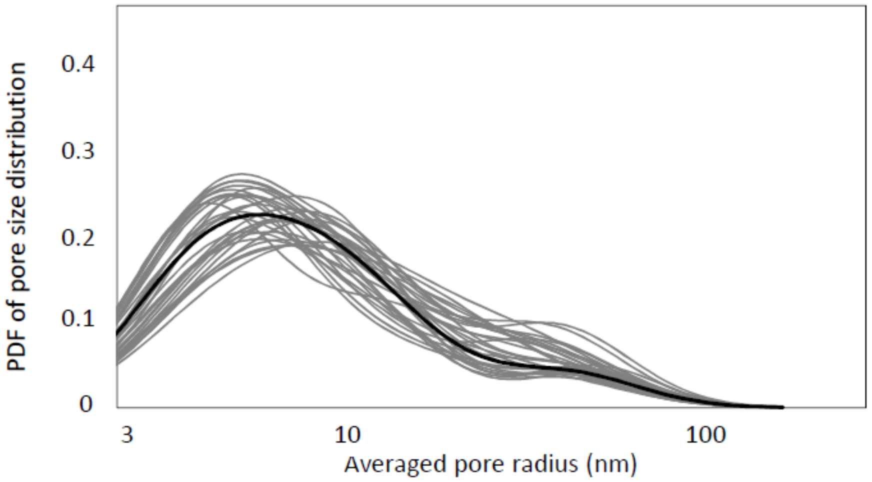


Comparison of Pore-Size Distribution



(a)

Sample #1



(b)

Sample #2

Simulation of Transport and Deformation with the Realization is Extremely Time Consuming

- ✓ **Not every aspect of the morphology of the realization (model) of the heterogeneous material makes significant contribution to its macroscopic properties**

- ✓ **Thus, one should find a way to “simplify” the realizations**

Curvelet Transforms

- **Consider a 2D image with spatial variable x, Fourier variable ω, and polar coordinates r and θ**

- **Define a pair of windows by**

$$\sum_{j=-\infty}^{\infty} w^2(2^j) = 1, \qquad r \in \left(\frac{3}{4}, \frac{3}{2}\right)$$

$$\sum_{l=-\infty}^{\infty} V^2(t-l) = 1, \qquad r \in \left(-\frac{1}{2}, \frac{1}{2}\right)$$

- **For each $j \geq j_0$ introduce the frequency window $U_j(r,\theta)$**

$$U_j(r,\theta) = 2^{-3j/4} W(2^{-j} r) V\left(\frac{2^{|j/2|}}{2\pi}\right)$$

- **Define the waveform $\varphi_i(x)$ such that**

$$\widehat{\varphi_i}(\omega) = U_j(\omega)$$

- **$\varphi_i(x)$ is the "mother curvelet". All other curvelets at scale 2^{-j} are obtained by rotation and translation of $\varphi_i(x)$**

- **Computationally, curvelets are more efficient than wavelets as they use much fewer coefficients to represent edges or wave fronts for a given accuracy**

Speeding up the Computations by Using Curvelet Transforms

We define the curvelet coefficients by

$$c(j,l,k) := \langle f, \varphi_{j,l,k} \rangle = \int f(x)\overline{\varphi_{j,l,k}(x)}dx$$

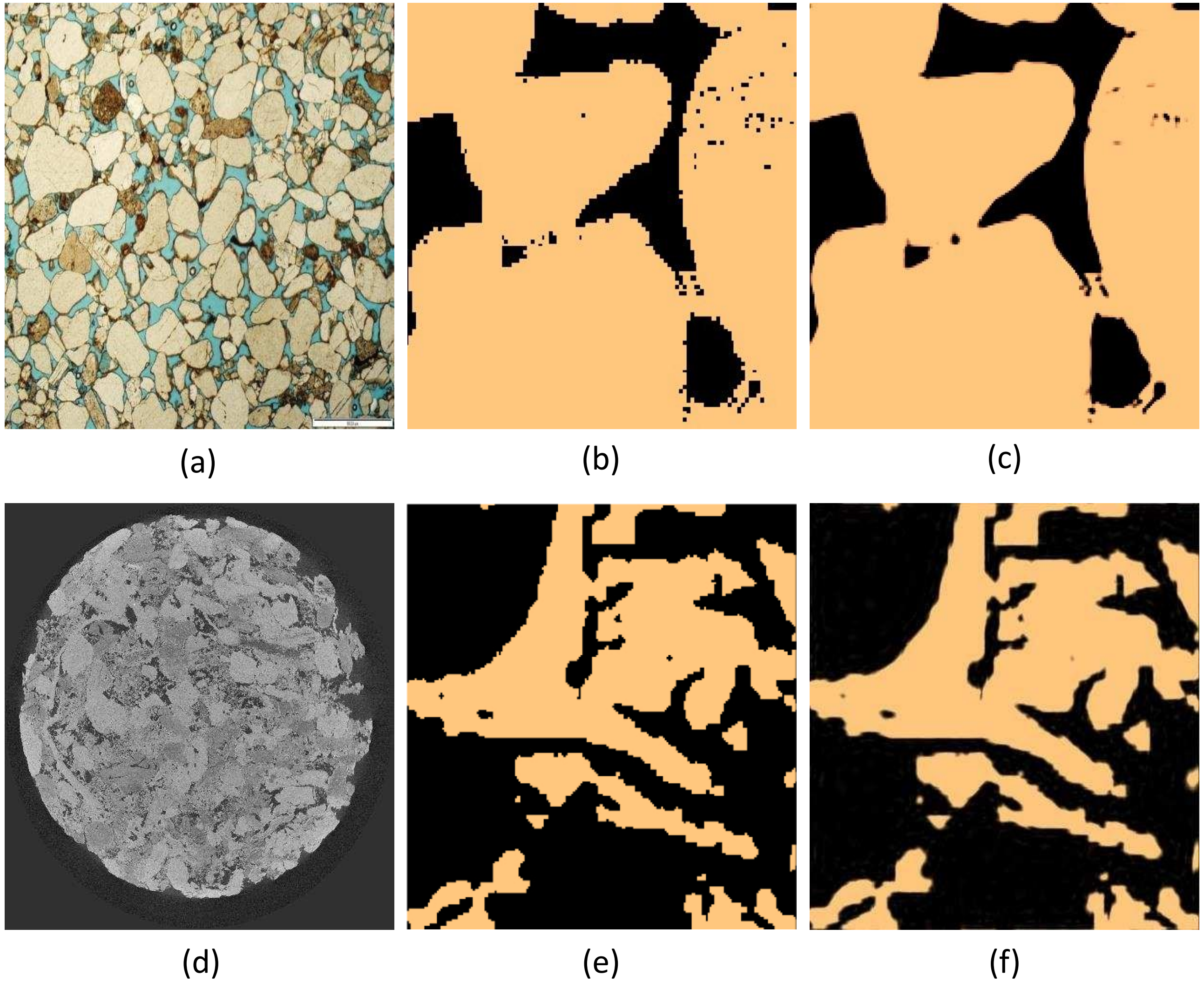
Compute the curvelet transform of the realization or model

Set a threshold for the curvelet coefficients

Set to zero all the coefficients that are smaller than the thresholod

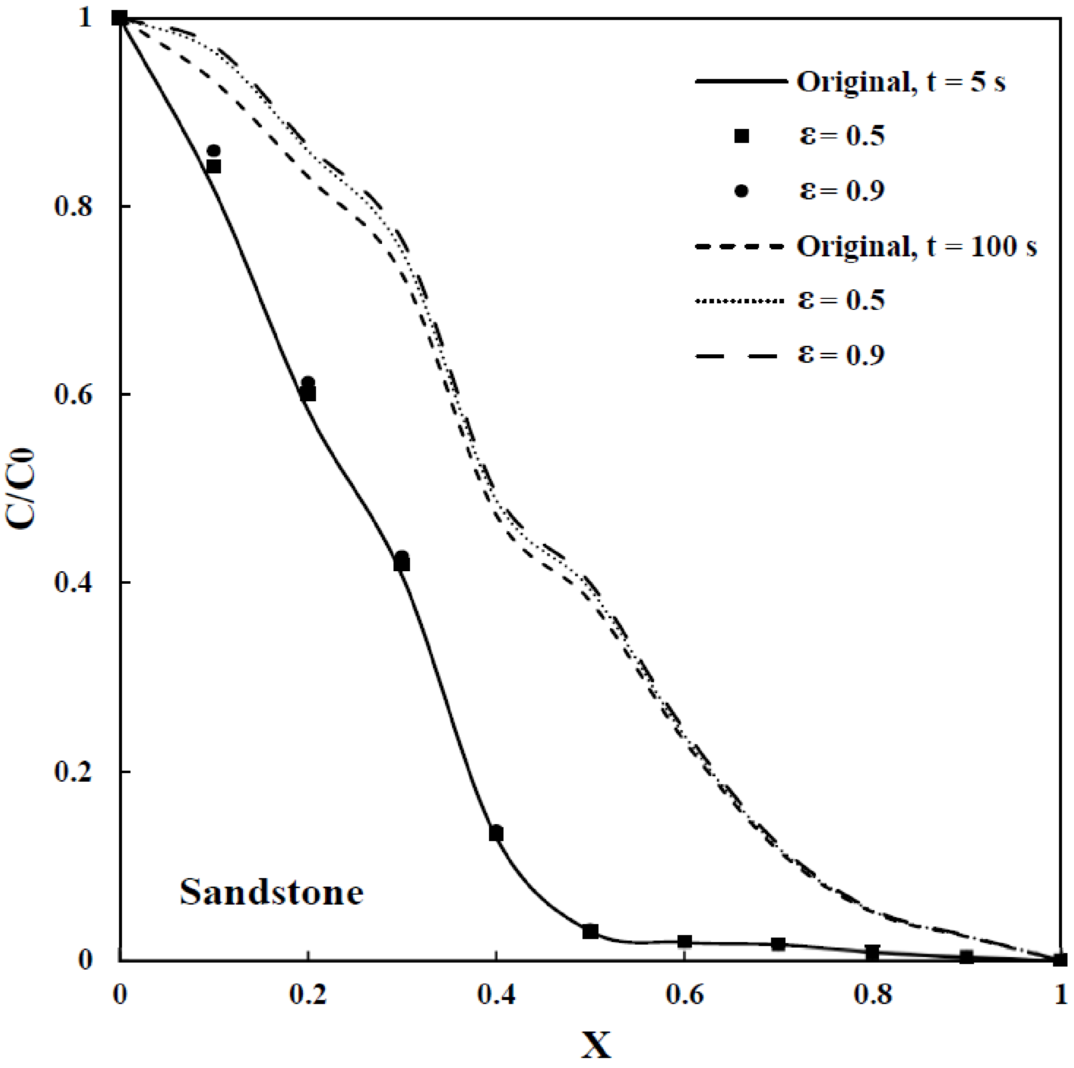
Bring back the realization to the real space

Sandstone & Carbonates



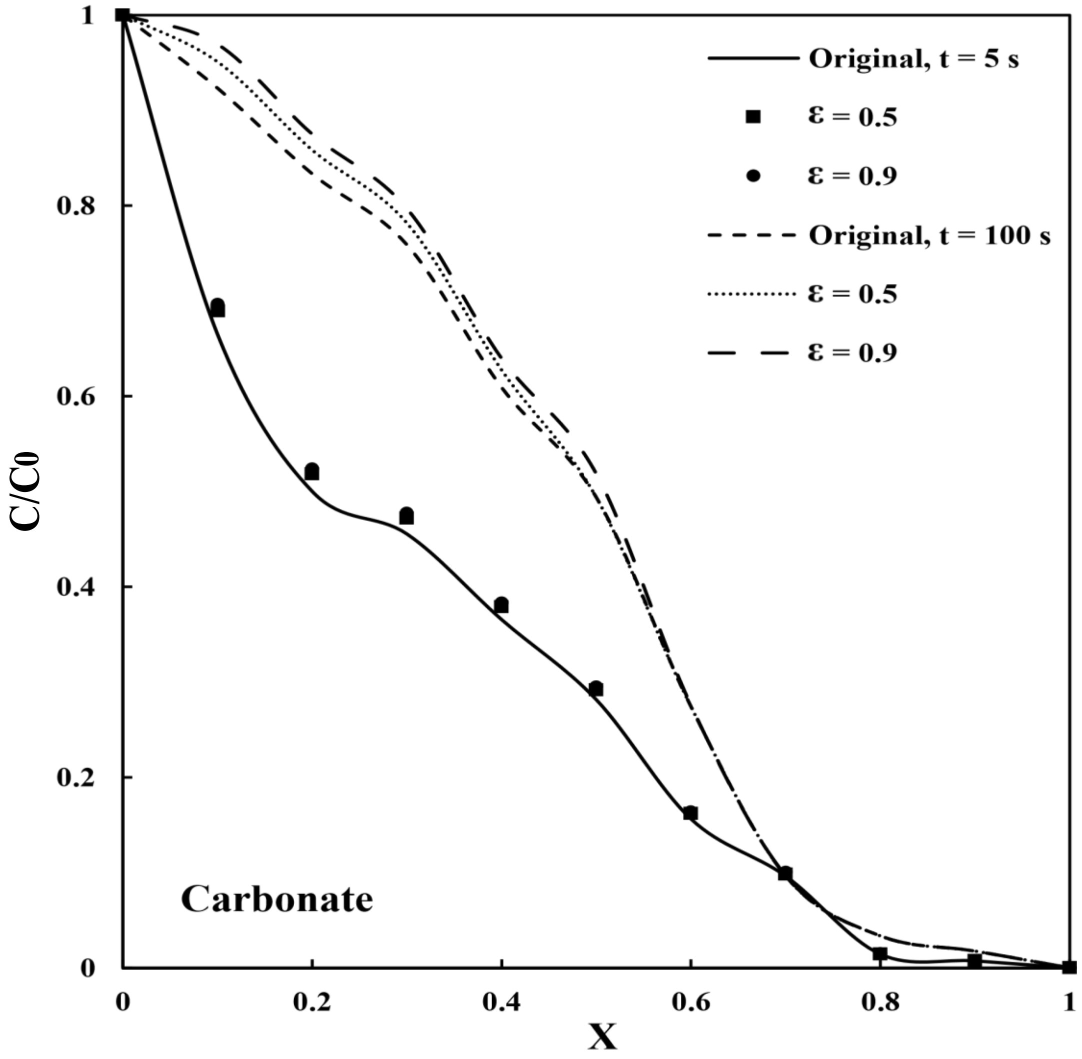
(a) Original image. (b) A small zoomed-in portion. (c) After curvelet transformation

Example: Solving Diffusion Equation in the Original Image and in its Curvelet-Denoised Version



Very close agreement

Carbonate



Very close agreement

Increase in the Computation's Speed

Sandstone	N	Threshold ε	$D_e \times 10^7$ (cm^2/s)	Time (CPU sec)
Original image	238941		23.72	602
Original image in CT space	54181	0	23.95	134
Curvelet-transformed image	3659	0.5	23.97	10
Curvelet-transformed image	2182	0.7	24.08	7
Curvelet-transformed image	1896	0.9	24.24	5
Carbonate				
Original image	543069		2.02	1556
Original image in CT space	129052	0	2.07	391
Curvelet-transformed image	8641	0.5	2.07	26
Curvelet-transformed image	5174	0.7	2.08	15
Curvelet-transformed image	4366	0.9	2.08	13

Summary

Reconstruction based on the cross-correlation function provides an accurate approach for generating realizations (models) of all types of heterogeneous media and materials, stationary or not non-stationary

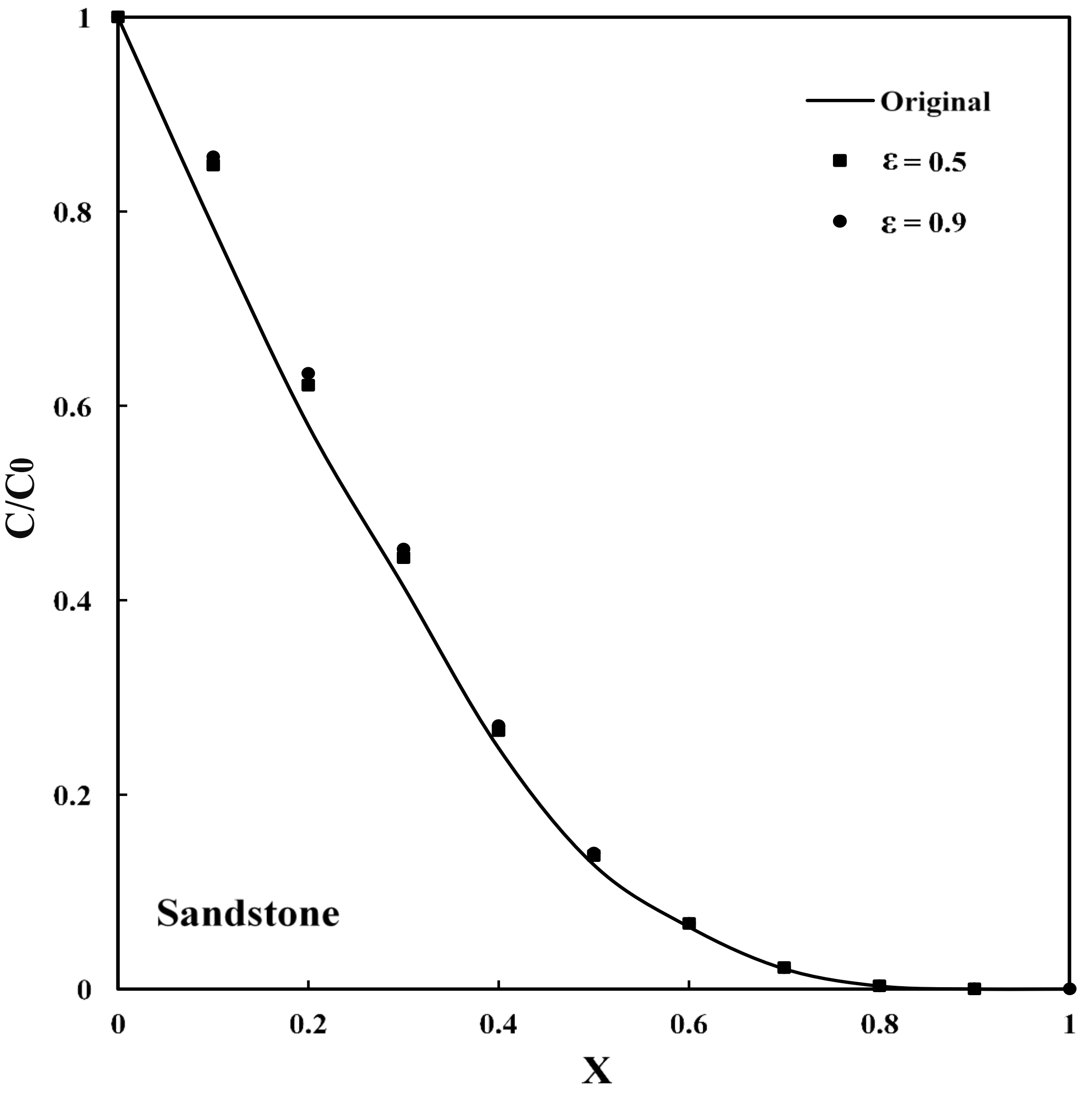
- ✓ **It is capable of integrating various types of data**
- ✓ **It can be used with hard (quantitative) data**
- ✓ **It can reconstruct multiscale, multiresolution media and materials**
- ✓ **When combined with curvelet transforms, the result is a powerful tool for modeling of complex materials and simulation of all types of phenomena in them**

Acknowledgments

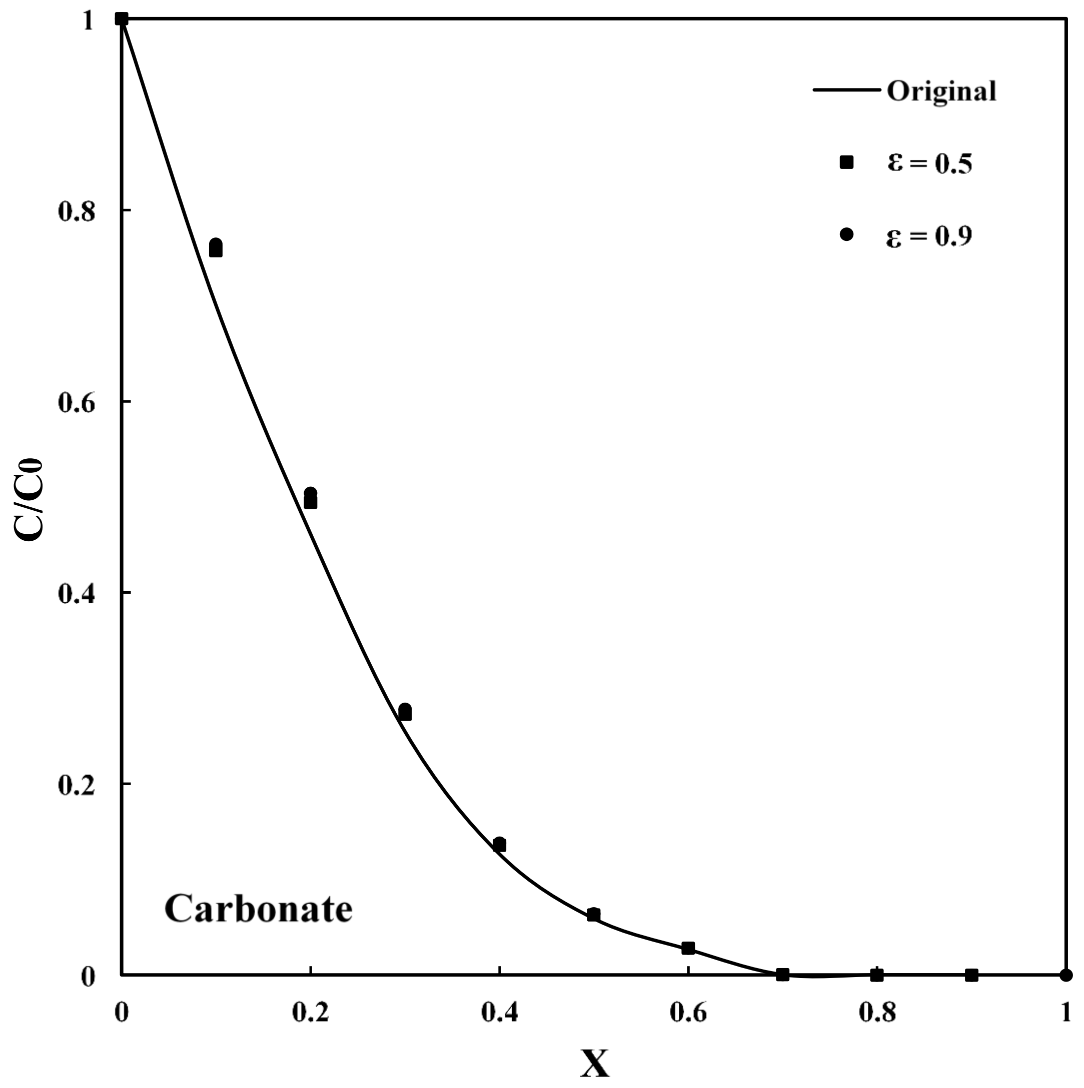
- ✓ The original algorithm based on the cross-correlation function and some of its refinements were developed in the Ph.D. thesis of my student, Dr. Pejman Tahmasebi, and as a research associate afterwards

- ✓ The work on curvelet is on-going with Ph.D. Student Abdullah Aljasmi

- ✓ The work was supported by the RPSEA Consortium

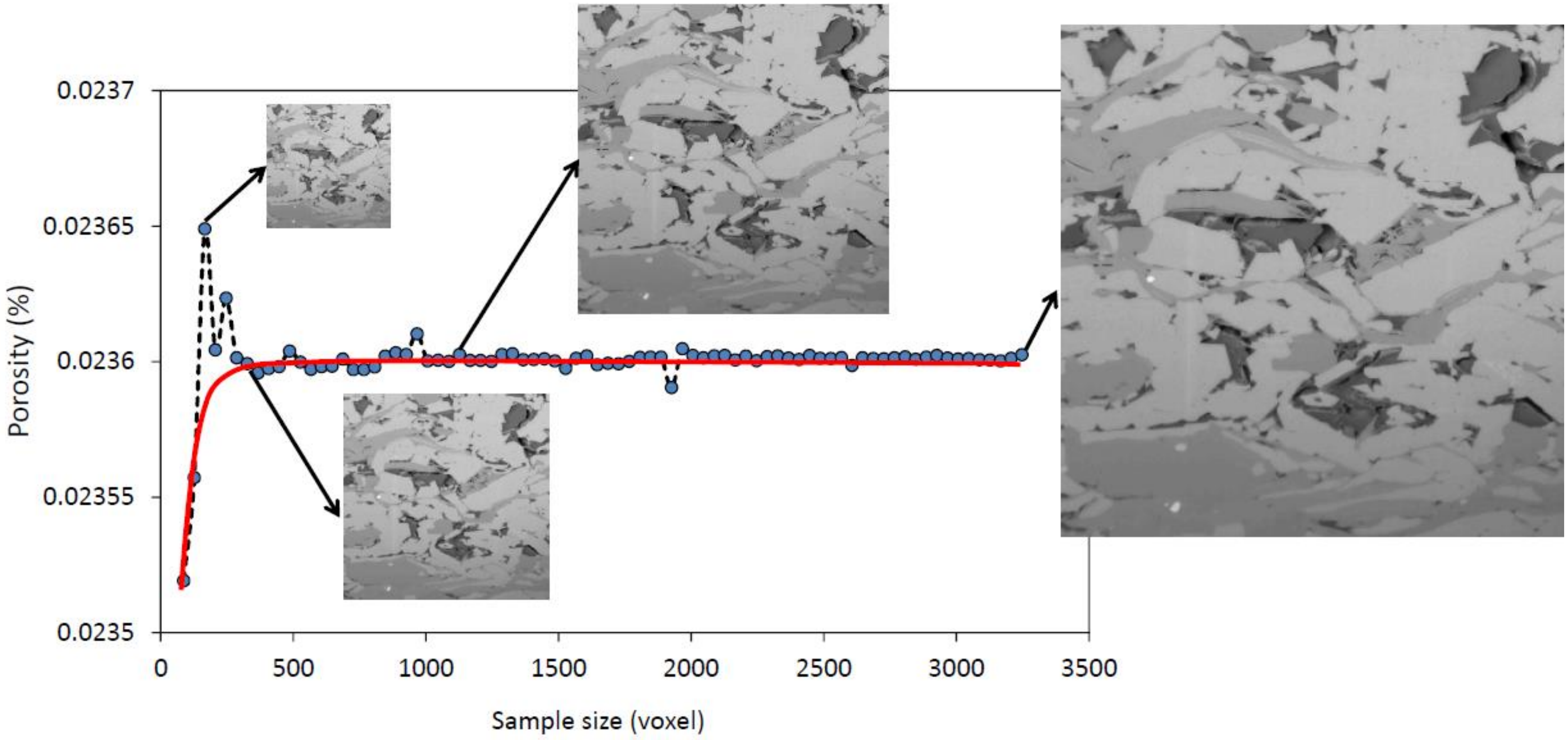


Original
ε = 0.5
ε = 0.9
Sandstone
C/C0
X
0
0.2
0.4
0.6
0.8
1

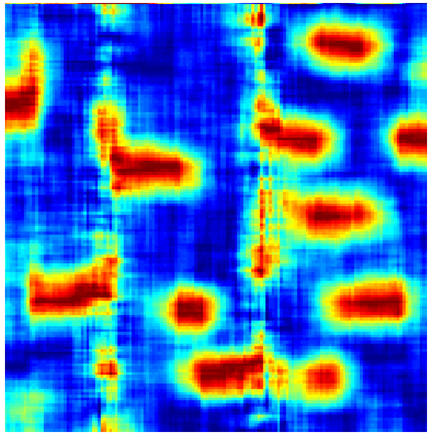


Original
ε = 0.5
ε = 0.9
Carbonate
C/C0
X
0
0.2
0.4
0.6
0.8
1

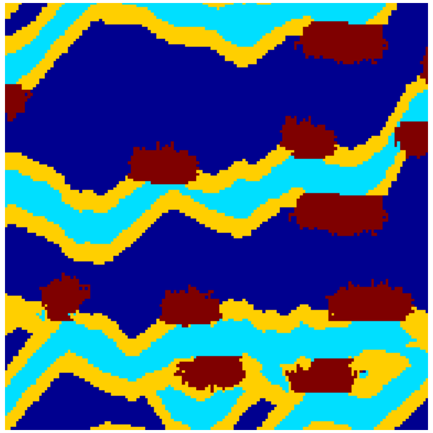
Determining Representative Elementary Volume using SEM Images



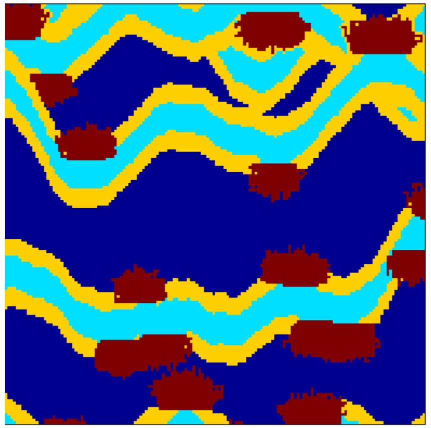
Multi-Facies (2D)



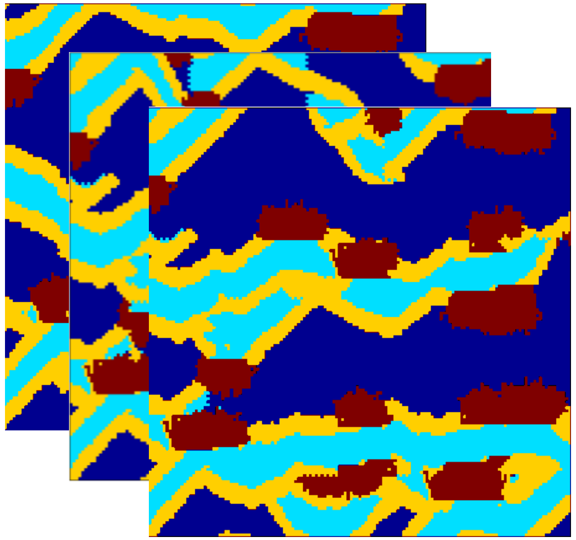
Soft data



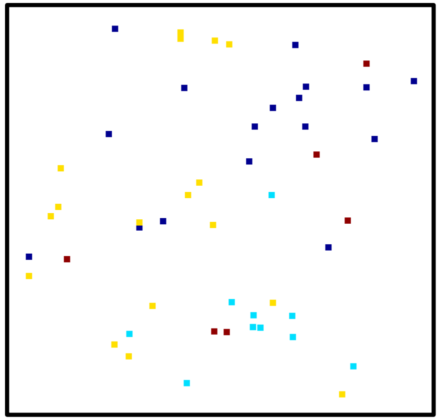
True model



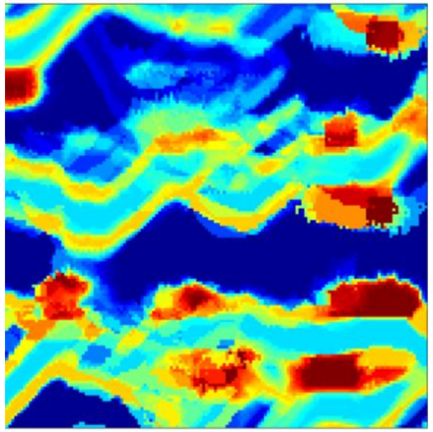
DI



Realization



Hard data



Ensemble average