Research Article

Quantitative Analysis for the Reconstruction of Porous Media Using Multiple-Point Statistics

Qing Xie,1,2 Jianping Xu,1,2 Yuanda Yuan,1,2 and Cong Niu3

1College of Petroleum Engineering, Xi’an Shiyou University, Xi’an 710065, China
2MOE Engineering Research Center of Development & Management of Western Low & Ultra-Low Permeability Oilfield, Xi’an 710065, China
3College of Building Environment Engineering, Zhengzhou University of Light Industry, Zhengzhou 450000, China

Correspondence should be addressed to Cong Niu; niucong@zzuli.edu.cn

Received 8 July 2020; Revised 3 September 2020; Accepted 11 September 2020; Published 6 October 2020

Academic Editor: Andrea Brogi

Copyright © 2020 Qing Xie et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The pore structure reconstruction of the porous media is of great importance to the research of mechanisms of fluid flow in porous media. To capture the large-scale patterns in the pore space, the multiple-point statistical technique is generally adopted for porous media reconstruction. Commonly, two different schemes, i.e., the single-grid scheme and the multiple-grid scheme, can be applied for simulation realization. The selection between this two schemes and a proper data template size have thus become a new research issue, and the performance of the characteristic reproduction of the training image using this two schemes must be quantified. In this paper, a series of multiple-point statistics simulation basing on a 2D micro-CT sandstone image are proceeded using both single- and multiple-grid schemes, and different data templates are adapted for porous media reconstruction. Further, to quantify the impact of the computational schemes and setting of the data template to the simulation realizations, a number of measurements considering the pore diameter, porosity, connectivity, and permeability are implemented to fully analyze the results obtained. Results show that by using the single-point statistical method, a large template is necessary to reproduce large-scale structures. The multiple-grid template method may bring great benefits to simulation efficiency over the simple data template method, as well as the recovery of the pore long-range geometric features and seepage characteristics. With the extension of the template for the multiple-grid scheme, the simulation results show lack of variations to some extent.

1. Introduction

Pore space reconstruction of porous media is of great interest in fields of earth science and engineering, because an accurate pore-scale modeling of the microstructure can help us understand and predict flow and transport phenomena. A faithful and accurate model of the pore space is an essential prerequisite for predicting the macroscopic properties and flow behavior of the medium.

Several categories of methods have been proposed to reconstruct the porous media. The most direct method to achieve this is to use X-ray computed tomography (micro-CT) to image the rock at a resolution of a few microns [1, 2], while the resolution is insufficient to image the smaller pore spaces in many carbonate samples. Process-based techniques have been developed to generate a model of the porous medium by simulating the sedimentation process, also incorporating the grain size distribution and other lithofacies information [3–6]. However, this method is computationally demanding and is specific to the particular process it models, and for many materials, the generating process is either unknown or highly complex. Some researchers used statistical feature models to describe the complex spatial information to generate representations of the pore space [7, 8], which are based on only two-point statistics (essentially the variogram model), and may fail to reproduce the long-range connectivity in some cases.

Due to capability of curvilinear and large-scale connected structures reproduction, multiple-point statistics (MPS) method has become one of most widely used approaches for porous media reconstruction. Okabe and Blunt [9, 10] aggregated the information coming from a 2D thin section of a carbonate rock with a linear pooling formula in order
to reproduce the micropores at the submicrometer scale. Haji-
zaeder et al. [11] proposed a technique which relies on succes-
sive 2D multiple-point statistics simulations coupled to a 
multiscale conditioning data extraction procedure. Comunian 
et al. [12] developed a method which relies on merging the lists 
obtained using the IMPALA algorithm [13] from the diverse 
2D training images, creating a list of compatible data events 
that are used for the MPS simulation. Xu et al. [14] took 
avantage of the entropy method to obtain the appropriate size 
of the template and combined successive 2D MPS simulation 
and 3D MPS simulation with reconstruction methods for 3D 
pore spaces. As a direct 3D simulation is straightforward 
but a CPU and RAM intensive procedure, accompanying with
the difficulties to achieve a 3D training image, all previous 
researches focused on restoring 3D porous media microstruc-
ture based on 2D thin section reconstruction. As current two 
main implementations of multiple-point statistics method, 
SNESSIM [15] and IMPALA [13], both rely on intermediate 
data structure to store patterns from training image, the recon-
struction process suffers from the running speed and RAM 
consumption, particularly in large scale and large data 
template simulation cases.

Many efforts have been made to decrease the CPU and 
RAM expense during the simulation, and the most com-
monly applied method is the multiple-grid approach [16], 
by which different sized templates are used to extract training 
image features at different scales; consequently, the confi-
figuration of the multiple-grid data template may significan-
tly influence the final realizations. However, experimental 
results show that increasing the number of multiple-grid 
scheme does not necessarily improve the large-scale structure 
reproduction [17]. Therefore, it is important to understand 
precisely its reproduction capacities and its parameter sensi-
tivity during the simulation process to achieve the satisfac-
tory digital models.

To address this issue mentioned above, a series of 
multiple-point statistics simulations are proceeded basing on 
a 2D micro-CT sandstone training image using both single-
and multiple-grid schemes in which different data template 
is adopted for porous media reconstruction. Further, to assess 
the impact of the computational schemes and setting of the 
data template to the simulation realizations, a number of 
measurements considering the pore diameter, porosity, 
connectivity, and permeability are implemented to fully ana-
lyze the results obtained in this paper. Results show that 
the multiple-grid template method may bring great benefits to 
simulation efficiency over the simple data template method, 
as well as the recovery of the pore long-range geometric 
features and seepage characteristics.

2. Multiple-Point Statistics Method

Two-point statistics are not enough to fully characterize low-
entropy structures, and different structures may have the 
same two-point statistics. So complex geometries, such as 
pore space structure, cannot be modeled accurately using 
only traditional two-point statistics such as the variogram-
based method, which manages only to reproduce propor-
tions and two-point variograms. As the MPS method uses 
higher-order statistics, which considers joint categorical 
variability at more points simultaneously for the parameteri-
zation of shapes, it is superior in reproduction of such 
complex geometries.

In multiple-point statistics, the concept of the training 
image (TI) must be introduced first, which is used to provide 
storage for multiple-point geostatistical probability patterns. 
Considering a category property \( Z \) with \( N \) possible states \( \{ z(n), n = 1, \cdots, N \} \), a geometric configuration \( \tau_n \) defined by \( n \) 

\[
\tau_n = \{ h_1, h_2, \cdots, h_n \}
\]

named data template is used to infer the spatial structure of the values for the categorical variable \( Z(u) \) at location \( u \). The combination of data template \( \tau_n \) and the \( n \) data values \( \{ z(h_n), a = 1, \cdots, n \} \) in the training image is 
called data event \( d_n \). Figure 1(a) shows a horizontal section of a 
fluvial system as a training image (black corresponds to 
mudstone with category code 0, and white corresponds to 
sandstone with category code 1). Figure 1(b) shows a data 
template \( \tau_n, n = 4 \), and Figure 1(c) shows a data event \( d_n, n = 4 \), associated with the data template Figure 1(b); the 4 data 
(category) values correspondingly are 0, 0, 1, and 1.

Consequently, the conditional probability of the 
ocurrence of value \( z(u) \) is defined following Bayes’ theorem:

\[
\text{Prob}\{ Z(u) = z_n | d_n \} = \frac{\text{Prob}\{ Z(u) = z_n, d_n \}}{\text{Prob}\{ d_n \}} = \frac{c_k(d_n)}{c_i(d_n)}, \hspace{1cm} (1)
\]

where \( c_i(d_n) \) is the number of replicates of the conditioning 
data \( d_n \), and \( c_k(d_n) \) is the number of replicates inferred from 
\( c_i(d_n) \) when the central node \( Z(u) \) has a value of \( z(n) \). Thus, 
the problem of calculating the probability of the value at 
location \( u \) can be simply turned to counting the number of 
replicates of the training image.

A large template would be required to reproduce large 
scale structures, at least double the size of the structure. 
However, the approach is commonly impractical because of 
memory constraints. The multiple-grid computation scheme 
is adopted to solve the problem, and large-scale structures of 
the training image can be reproduced while considering a 
data template with a reasonably small number of grid nodes.

The simulation process using multiple-grid approach 
consists of several parts which use different scale grids. For 
G-grid simulations, the \( g - \text{th} \) (\( 1 < g < G \)) grid is constituted of 
each \( 2^{g-1} - \text{th} \) node of the smallest simulation grid. At each 
grid scale, the data template is configured proportionally to 
the spacing of the nodes within the grid to be simulated. If 
any original sample data exists, they must be assigned as 
conditional data to the nearest nodes in the current simula-
tion grid. Figure 2 shows the example for multiple grids of 
three levels, including multiple-grid data templates based on 
single-grid data template (Figure 1(b)).

Algorithms 1 and 2 give a detailed description of the 
multiple-point statistics simulation of a single grid and 
multiple grids.

3. Data and Measurement

3.1. Training Image. We evaluated the performance of the 
reconstruction of sand rock by performing several simula-
tions with SGeMS [18] which implement single-grid and
multiple-grid multiple-point statistics simulation methods. Finally, we compared the results using different quantitative measurements.

The training image adopted in this research is shown in Figure 3 in which void space is shown white and the solid black. The image comes from scanning the rock sample using micro-CT, whose size 400 × 400 pixels and resolution is 5 μm/pixel. Successively, we performed several unconditional single-grid and multiple-grid simulations on a grid of 200 × 200 for comparison.

3.2. Quantitative Measurements. Several quantitative measurements are used to quantify the reproduction performance of the statistics of the sand rock sample. Different measurements highlight different features of the simulation results. These measurements include univariate, bivariate, multivariate, and flow characteristic, which are summarized in Table 1 and briefly introduced below. Note that the computation methods and the correlation of these parameters are not emphasized in this research. We are simply interested in how the consistency of characteristics between the rock sample and the simulation realizations is influenced by the simulation methods and parameters.

The univariate statistics adopted describe the morphology of a porous medium, which consist the information of the geometry that describes the pores’ shapes and sizes and the topology that quantifies the way pores and throats are connected together. As the exterior behaviors of the porous medium such as flow characteristic largely depend on the medium’s microscope form, the morphology of a porous medium plays a fundamental role in all porous media researches. However, although the univariate statistics provide valuable information and insight into the morphology of the porous media, it is not by itself nearly enough for

![Figure 1: Example of a training image, a data template, and a data event.](image1)

![Figure 2: Example for multiple grids of three levels.](image2)

**Algorithm 1: Multiple-point statistics simulation of a single grid**

1. Define a data template \( \tau_n \)
2. Pre-scan training image to store multiple-point patterns specific to \( \tau_n \)
3. Relocate hard data to the nearest simulation grid node and freeze them
4. Define a random path visiting all locations to be simulated
5. for each location \( u \) along the path do
6. Find the conditioning data event \( d_n \) defined by template \( \tau_n \)
7. Retrieve the conditional cumulative distribution function (CCDF) \( \text{Prob}(Z(u) = s_k | d_n) \) from patterns got in step 2
8. Draw a simulated value \( z(u) \) from that conditional probability and add it to the data set
9 end for
developing a realistic model of porous media due to the heterogeneous and anisotropy characteristics in larger scales.

The bivariate statistics through the use of variogram analysis established the pore’s directional dependence and correlation at two nearby locations and is used to interpret the spatial dependence and the anisotropy extent of the pore structure along different directions:

\[
\gamma(h) = \frac{1}{2} E \left\{ \left[ I(u + h) - I(u) \right]^2 \right\},
\]

where \(\gamma(h)\) is the variogram between two paired data values, \(h\) is the correlation function which represents the location vector between paired points, \(E \left\{ \left[ I(u + h) - I(u) \right]^2 \right\}\) represents the expectation of \( \left[ I(u + h) - I(u) \right]^2 \), and \(I(u)\) means the data value at location \(u\).

As the rectilinear variogram does not include curvilinear information of complex pore structures, such bivariate statistics is largely insufficient to characterize the complex shape and spatial continuity of the pore structure. In order to make up for it, the measurement of curvilinearity and multiple-point connectivity is adopted as:

\[
K(h) = E(I(u)I(u + h)) = \Pr \left\{ I(u) = 1, I(u + h) = 1, \text{conn}(u, u + h) = 1 \right\},
\]

where \(K(h)\) is the correlation function and \(\text{conn}(u, u + h)\) is the connective function, which shows \(u\) and \(u + h\) are connected.

Besides, the statistical measurements mentioned above, the permeability is considered to better characterize the fluid flow throughout the porous media. The permeability is a parameter that measures the resistance that fluid encounters when flowing through a porous medium. Many efforts have been devoted in the determination of flow and transport in porous media [25–27]; we in this paper adopted the finite element Stokes simulation proposed by Chapuis and Aubertin [28] to calculate the permeability of various simulation realizations.

Typically, it is known that according to the Kozeny-Carman equation [28], the absolute permeability \(k\) is related to porosity \(\phi\) and grain size \(D\) as:

\[
k \sim D^2 \phi^3. \tag{4}\]

Comparing permeability calculated using finite element Stokes simulation with this fitting curve, we can determine which of the various realizations has the best permeability reproduction.

These quantitative measurements are calculated for each simulation realizations, using both single and multiple grid and different data template size.

4. Results and Discussion

In this section, we evaluated the performance of reconstruction samples by performing simulations and comparing the results using different measurements, including single-grid and multiple-grid schemes. Due to the stochastic feature of the multiple-point geostatistical simulation, different random seeds are used, and dozens of the simulation are completed on a 200 × 200 grid with data template of 3 × 3, 5 × 5, 7 × 7, ..., 19 × 19, respectively. The results calculated using different measurements are then averaged and compared with the statistics of the training image to quantitate the effect of simulation methods and parameters. As the training image and the simulation grid are relatively small, we use just two grid levels (\(G = 2\)) for multiple-grid scheme.

Figure 4 shows the typical simulation results with data template size of 3 × 3, 11 × 11, and 19 × 19 using single- and multiple-grid scheme, respectively. Firstly, it can be found that as the size of data template increases, the restoration of curvilinear structure for sand channel using the single-grid method becomes better, while the improvements
using the multiple-grid method are relatively small. Secondly, when the data templates have the same size, the reproduction of curvilinear pore structure in the realization using the multiple-grid method is much better compared with the realization using the single-grid method. It shows that as the data template becomes larger, the characteristics of the pores can be better captured, and the multiple-grid method can search a larger range with the same template size, which is in accordance with the definition of the data template.

To further clarify the relations between the effects of the pore shape reconstruction and the simulation methods and parameters, the information about pore shape and interconnection are counted and shown in Figures 5–7, respectively.

For the measurement of average pore diameter shown in Figure 5, a similar regularity of single-grid scheme is shown compared to Figure 4 that as the size of the data template increases, the average pore diameter of the realizations becomes closer to the value of the image. Unfortunately, the average pore diameter may even exceed the value of the training image when using relatively large data template, which will bring a certain degree of unreality for the simulation results. Therefore, finding the appropriate data template size

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
<th>Measurement</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Univariate statistics</td>
<td></td>
<td>Porosity</td>
<td>Rabbani et al. [19]</td>
</tr>
<tr>
<td></td>
<td>Morphology</td>
<td>Pore (grain) size distribution</td>
<td>Rabbani et al. [20]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Coordination number</td>
<td>Rabbani et al. [21]</td>
</tr>
<tr>
<td>Bivariate statistics</td>
<td>Correlation</td>
<td>Variogram</td>
<td>Journel et al. [22]</td>
</tr>
<tr>
<td>Multivariate statistics</td>
<td>Continuity</td>
<td>Connectivity</td>
<td>Journel et al. [23]</td>
</tr>
<tr>
<td>Flow characteristic</td>
<td>Percolation</td>
<td>Permeability</td>
<td>Karim et al. [24]</td>
</tr>
</tbody>
</table>

![Figure 4: Simulation results using data template $3 \times 3$, $11 \times 11$, and $19 \times 19$ with single- and multiple-grid schemes. (a–c) Single-grid results. (d–f) Multiple-grid results.](image)

![Figure 5: Average pore diameter of the simulation realizations: (a) single grid; (b) multiple grid.](image)
is an important issue for single-grid scheme. For the multiple-grid scheme, average pore diameters for all data template sizes are bigger than the value of the training image. However, a clear correlation cannot yet be easily found between the data template size and the statistical data. All average pore diameters of simulations using the multiple-grid scheme are higher than the training image, and larger size of data templates may even bring greater errors than smaller one.

The data for the measurement of pore diameter distributions shown in Figure 6 better reveal the information about the pore shapes. It can be seen that for both single-grid scheme and the multiple-grid scheme, the lower outliers are substantially equal, and the higher outliers increase with the sizes of data template increase, and the medians have similar trends with the average value shown in Figure 5 for the training image and all data template sizes. For single-grid method, the interquartile range increases, whereas the ones for multiple-grid method are roughly equal. This shows that the limitation of template size has a greater impact on the simulation results. However, from another point of view, the extension of the template for the multiple-grid scheme makes the simulation loss of variations to some extent.

Figure 7 shows the results of the measurement of coordinate numbers for both single-grid and multiple-grid schemes, which describe the number of ways which interconnect pore with the pore network structure. Compared with Figure 5, it is shown that the coordinate number and average pore diameter have similar trends, and there is a certain correlation between these two datasets. Although large pores may contribute nothing to the pore network unless they connect to other pores, as the pore size increases, the likelihood of more throat connections to the pore also increases. Therefore, the coordinate numbers increase with the size of pore and with increasing number of pore throats surrounding the pore.

The simulation results for measurement of porosity, which describes the pore volume fraction in the rock samples, are also shown in Figure 8. As the results of other measurements mentioned above, the data of single-grid scheme basically shows a monotonous increasing trend, but the data of multiple-grid scheme shows a more chaotic pattern. On the other hand, for all data template sizes, the data distribution interval of multiple-grid scheme (0.185, 0.21) is more concentrated than the single-grid scheme (0.175, 0.215). As the measurement of porosity is a one-dimensional data, the magnitude of porosity still roughly indicates complexity of structure, being greater for greater complexity. Furthermore, the spatial variability of porosity is also important, greater variability correlating with greater heterogeneity of the rock. Due to the fact that heterogeneity in the porosity distribution is evident even
by visual inspection, the correlation between the porosity and the measurement of rock morphology need to be further researched.

In Figure 9, the experimental variograms are used to quantify the spatial variability in the simulation results. The nugget values are roughly the same in each figure, which shows that...
the reproductions for the discontinuity characteristics at the origin corresponding to short scale variability are kept consistent. The training image and all simulation results show the isotropic variation of the pore structure, as the ranges in the \( x \) and \( y \) directions are almost equal. Except when the template is relatively small for single-grid method, the curves of all realization are kept close to curve of the training image. It proves that as long as the template range is large enough to capture adequate patterns, the correlation of the pore space can be well recovered from the training image.

The connectivity in the space of the pore network structures is an important measuring for the interconnected pore curvilinear patterns, which is shown in Figure 10. The data for both single-grid and multiple-grid schemes shows a similar pattern with the measurements of average pore diameter, coordinate number, and variograms, which are all factors that affect the pore network. Commonly, the distance at which the model first flattens out the connectivity curve is twice the distance of the variogram curves mainly because of the difference between the bivariate and multivariate statistics. The measurement of multiple-point connectivity allows a longer relevant spatial extent than the traditional two-point variogram model.

Predicting fluid transport capability is the final target of reconstruction a structural model of porous media, and permeability is the key parameter that measures the resistance that fluid faces when flowing through a porous medium. Stoke flows have been simulated in the simulation realization using finite element method, and the results are shown in Figure 11. It can be concluded that the training image shows an isotropic percolation capability; however, the permeability for both the single-grid and multiple-grid schemes shows more or less anisotropic characteristics. As the measurement of porosity shown in Figure 8, the permeability results also distribute messy, whereas the trend of the two kinds of data are similar.

To quantify the performance of the permeability reproduction, the constant coefficient of the simplified Kozeny-Carman relationship for the training image has been calculated, and the scatter plot (Figure 12) is depicted to show the relationship between the permeability and the sandstone morphology parameters, porosity and grain size, and the derivation from the theoretical curve of the training image. The data points of the multiple-grid schemes are more scattered than the points of the single-grid scheme, and the latter has better curve fitting effect with the theoretical curve of the training image, which also been proved by the curve
fitting statistics of mean square of residual and coefficient of determination in the Table 2. So it can be concluded that the simulation results of single-grid scheme are more consistent with the Kozeny-Carman formalism. This may be partly due to that the large search range of the multiple-grid scheme reduces the variation of the simulation to a certain degree, although the measurements in the simulation results remain the same level, the distribution of data is relatively disordered, and it brings a relatively large transmission error.

5. Summary and Conclusions

In this paper, the multiple-point statistical simulation method is adopted to reconstruct two-dimensional sandstone images with different size of square data template using both single-grid and multiple-grid scheme, and different measurements are used to quantify the performance of pattern reproduction from the original training image. The results of univariate, bivariate, multivariate, and percolation
measurements have been compared and discussed, which represent different aspects of the sandstone. According to the calculation results in this article, the following conclusions can be drawn for 2D MPS simulation:

(1) For single-grid scheme of the simulation, as the size of the data template increases, all simulation results gradually approach to the training image, and the reproduction measurements are stably improved which indicates that a large template is necessary to reproduce large-scale structures. Although the data template size can be theoretically expanded to capture the longest structure in the training image, the template size is limited by the memory limitations and computational efficiency in the practical applications.

(2) The multiple-grid scheme is commonly used to overcome the shortcomings of the single-grid scheme, in which a hierarchical grid structure is used. The multiple-grid approach captures the large-scale multiple-point statistics effectively while requiring relatively small data template. Note that increasing the number of multigrids may not improve the large-scale structures; this feature brings loss of variations of the simulation results to some extent.

(3) In multiple-grid scheme simulation, finding the appropriate data template size becomes an important issue as the single-grid scheme. The regularity between the simulation results and the data template size of the multiple-grid scheme requires further research.

Nomenclature

c: Number of events
conn: Connectivity function
D: Diameter of grain
d: Data event
E: Expectation
ϕ: Porosity
γ: Variogram function
G: Grid
g: Grid number
h: Vector
I: Data value
K: Correlation function
k: Permeability
N: Number of grids
n: Counter
Prob: Probability
τ: Data template
u: Location vector
Z: Random variables
z: Sample event.

Table 2: The K-C theoretical curve fitting statistics.

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Mean square of residual</th>
<th>Coefficient of determination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-grid scheme</td>
<td>7.98975</td>
<td>-0.46433</td>
</tr>
<tr>
<td>Multiple-grid scheme</td>
<td>18.00063</td>
<td>-8.96497</td>
</tr>
</tbody>
</table>

Data Availability

All data generated or analyzed during this study are included in this article.

Conflicts of Interest

None of the authors have any conflicts of interest.

Acknowledgments

This work was supported financially by the National Nature Science Foundation of China (grant numbers 51974247, 51874241, 41502311, and 41672114).

References


