Reconstruction of a digital core containing clay minerals based on a clustering algorithm

Yanlong He,[1,2] Chunsheng Pu,[1,2,*] Cheng Jing,[1,2] Xiaoyu Gu,[2] Qingdong Chen,[3] Hongzhi Liu,[2] Nasir Khan,[2] and Qiaoling Dong[4]

[1]*School of Petroleum Engineering, Xian Shiyou University, Xi'an, Shanxi, 710065, China*

[2]*School of Petroleum Engineering, China University of Petroleum, Qingdao, Shandong, 266555, China*

[3]*CNOOC Energy Technology & Services Limited, Tianjin, Tianjin, 300457, China*

[4]*Daqing Oilfield Company Ltd., CNPC, Daqing, Heilongjiang, 163712, China*

(Received 6 July 2017; published 9 October 2017)

It is difficult to obtain a core sample and information for digital core reconstruction of mature sandstone reservoirs around the world, especially for an unconsolidated sandstone reservoir. Meanwhile, reconstruction and division of clay minerals play a vital role in the reconstruction of the digital cores, although the two-dimensional data-based reconstruction methods are specifically applicable as the microstructure reservoir simulation methods for the sandstone reservoir. However, reconstruction of clay minerals is still challenging from a research viewpoint for the better reconstruction of various clay minerals in the digital cores. In the present work, the content of clay minerals was considered on the basis of two-dimensional information about the reservoir. After application of the hybrid method, and compared with the model reconstructed by the process-based method, the digital core containing clay clusters without the labels of the clusters' number, size, and texture were the output. The statistics and geometry of the reconstruction model were similar to the reference model. In addition, the Hoshen-Kopelman algorithm was used to label various connected unclassified clay clusters in the initial model and then the number and size of clay clusters were recorded. At the same time, the K-means clustering algorithm was applied to divide the labeled, large connecting clusters into smaller clusters on the basis of difference in the clusters' characteristics. According to the clay minerals' characteristics, such as types, textures, and distributions, the digital core containing clay minerals was reconstructed by means of the clustering algorithm and the clay clusters' structure judgment. The distributions and textures of the clay minerals of the digital core were reasonable. The clustering algorithm improved the digital core reconstruction and provided an alternative method for the simulation of different clay minerals in the digital cores.

DOI: 10.1103/PhysRevE.96.043304

I. INTRODUCTION

Reconstruction of digital cores has been studied for several decades as the microstructure reservoir simulation method [1–4]. Generally, two kinds of methods are used to reconstruct the digital cores. The first is to reconstruct by a physical modeling technique, for instance, x-ray microcomputed tomography (CT) scanning, focused ion beam scanning electron microscopes (FIB SEMs), and the combination of a series of casting thin sections [5–7]. Due to various factors, such as the unavailability and high cost of the instrument, and the fact that the composition of rocks in the three-dimensional (3D) images cannot be divided, the application of such methods has not become widespread [8]. In the other method, numerical modeling techniques, such as the stochastic method, process-based (PB) method, multiple point statistic (MPS) method, continuum geometrical (CG) method, and simulated annealing (SA) method are frequently used [1,9–22]. The necessary data of such methods can be obtained with relative ease. It is believed that the PB method effectively simulates the processes of sedimentation, compaction, and diagenesis based on the two-dimensional (2D) information from the reservoir, particularly a sandstone reservoir. However, the model reconstructed by the PB method is much simpler [23]. The SA method considers the statistical properties of the reference model from the reservoir, such as porosity, aautocorrelation (AC) function, and linear-path (LP) function. However, connectivity of the model is poor and the convergence speed decreases drastically while updating the model [24–26]. The hybrid method combines both the PB and SA methods, improving transportation properties and accuracy of the reconstructed digital cores [27,28].

The diagenesis of a sandstone reservoir is simple and grains are sometimes cemented loosely. Subsequently, it would be a challenge to acquire a core from such an unconsolidated reservoir. Eventually, fundamental parameters that provide essential information for the reconstruction of digital cores based on the physical modeling technique can be obtained in a difficult manner. Therefore, reconstruction of digital cores using a 2D data-based method is applicable for sandstone reservoirs [13,29–31], although the reconstruction method of digital cores based on the 2D data can model rock matrix and their pores. However, partition and reconstruction of clay minerals in the digital cores are still challenging to perform [32,33]. Meanwhile, clay minerals play pivotal roles in the formation not merely in terms of plasticity with water, but also for the absorbability and exchangeability of ions. In addition, the formation damages are often caused by variant properties of clay minerals during the production phase [4,34–38]. In general, the accuracy and practicability of digital cores are influenced by the reconstruction of clay minerals.

As an accurate physical modeling technique, the different components in the CT images are difficult to divide and reconstruct [3,28]. Øren and Bakke [1] and Jin *et al.* [39] first reported a simple method to reconstruct the clay clusters

*228128575@qq.com

using the typical numerical modeling method (PB method). According to their research, they took a circumscribed polygon of the grains in the model as the clay minerals [please see the Supplemental Material [40] for the schematic diagram of the spherical grains and the fractal grains (Fig. 8S)]. The more serious drawback is that the content and the volume of the clay minerals in their work are only decided by the size of the grains; however, the various types and textures of clay minerals in the reservoir were not considered. In either method, the division and reconstruction of the minerals in the digital cores are crucial for improving the above-mentioned technique [28].

In this paper, the algorithms relevant to label and divide various clay clusters were used. To the best of our knowledge, the Hoshen-Kopelman (HK) algorithm was the first to propose a "cluster multiple labeling technique" for the porous media percolation, and it has a wide variety of applications [41–43]. Some examples of such applications are in nuclear fuel rod processing [44], catalysis [45], curing of epoxy resin [46], classification of radar signatures [47], and assessing habitat fragmentation [42,48]. Only after the introduction of this algorithm did the very large lattice simulations by the Monte Carlo method become possible [49–51]. The HK algorithm is based on detecting connected clusters on a lattice and labeling the involved lattice sites, so that all sites are connected and sharing the same label [48]. It is a well known fact that clustering algorithm can be applied in various fields, for instance, data mining, statistics, and mode recognition [52–54]. The most widely used and frequently studied clustering algorithm is the K-means clustering. Usually, it can divide the neighboring discrete points into different clusters based on the nature of discrete points. It starts with a random initial partition and keeps reassigning the points to clusters based on the similarity between the point and the cluster centers until a convergence criterion is met [52,55]. The K-means clustering algorithm can process a large amount of data along with theoretical reliability, simplicity in the algorithm, and fast converging speed [54,56].

In the present work, we have proposed a truly hierarchical method to reconstruct the digital cores containing various clay minerals in which all of the connected unclassified clay clusters in the initial model were labeled with size and number by the HK algorithm. After that, any unreasonable structure of large communicating clusters were divided into smaller clusters on the basis of the affiliation between voxels and centers of clay clusters by the K-means algorithm. The textures of clay clusters were typically divided by the structure judgment into three major textures, such as surface filling of the grain, intergrain filling, and metasomatic texture. Furthermore, considering the different types, content, and textures of the clay minerals in the reservoir, the clay clusters in the reconstructed model were reunified with the types, content, and textures of the clay minerals in the reservoir. Finally, all of the clay clusters were labeled with size, number, and texture. Eventually, digital cores containing various clay clusters with different structure properties were reconstructed. The method was motivated by recent observations [1,28,39,42,56] and the purpose was to divide and reconstruct various clay minerals with different types and textures in the digital cores. The results allowed us to obtain the distributions of various clay minerals and textures with the size of the clay clusters in the reconstruction model. In general, the digital cores containing clay minerals based on the clustering algorithm reflected the classification and structure of clay minerals. Moreover, clay clusters' textures and distributions in the digital core corresponded to that in the reservoir. The present work was an attempt to reconstruct the digital cores containing various clay minerals with different textures. The accuracy and the applicability of the digital cores were improved.

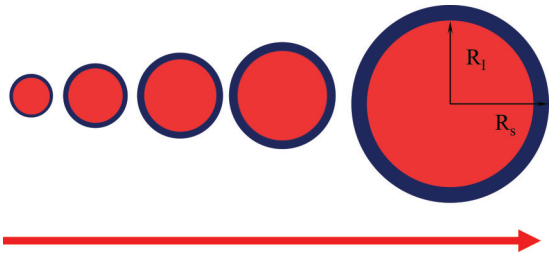


FIG. 1. Schematic diagram of the size of sediment grains. R_I is the radius of the initial grain, R_S is the radius of the sediment grain (considering the clay content), V_I is the volume of the red grain (the initial grain), V_C is the volume of the blue ring (the clay mineral), V_S is the volume of the sediment grain (considering the content of clay), where $V_S = V_I + V_C$.

The rest of this paper is arranged as follows. In the next section we briefly describe the 2D data from the reservoir. The method of reconstruction of the initial model containing the rock matrix, and unclassified clay clusters and pores is given in Sec. III. In Sec. IV, the hierarchical method to divide and gather the statistics of the various clay clusters with types, sizes, number, and textures is presented. In Sec. V, the reconstruction of a 3D digital core containing various clay minerals with three typical textures is performed. Sec. VI describes the results, including the geometry and statistics of the digital core, and the distributions of various clay clusters with different sizes and textures in the model. The paper is summarized in Sec. VII.

II. 2D DATA FROM THE RESERVOIR

The GD reservoir located in the northeast of the Shengli oilfield (China) commenced oil production in 1983. It is a typical unconsolidated sandstone reservoir with a high clay mineral content. The 2D data of the reservoir are from the GD827-10 well. The schematic in Fig. 1 shows details of the sediment grains, and the grain size distribution and the casting thin section of the reference model from the reservoir are shown in Figs. 2(a) and 2(b). The porosity, permeability, and shale content of the reference model are 26.38%, 0.614 μm^2, and 12.36%, respectively. In addition, the content of montmorillonite, chlorite, kaolinite, and illite are 40.8%, 27.4%, 19.1%, and 6.3%, respectively. The major textures of montmorillonite are reported as the surface filling of the grain and intergrain filling, along with some metasomatic texture. The textures of illite include intergrain filling, film-type filling, and metasomatic texture. Surface filling of the grain, film-type filling, and metasomatic texture are also common in chlorite's textures. Last but not least, the principal texture of kaolinite is intergrain filling [57].

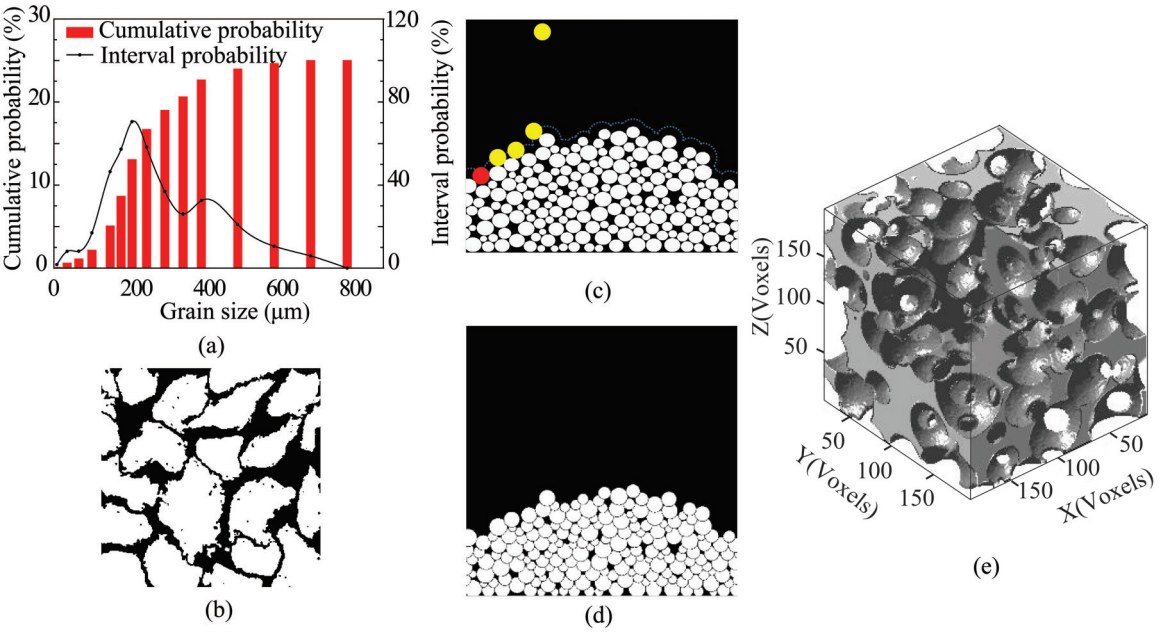


FIG. 2. Digital core reconstructed by the PB method. (a) The initial grain size distribution. (b) Casting thin section of the reference model. (c,d) The 2D slices of sedimentation and compaction process by the PB method. (e) The 3D digital core reconstructed by the PB method.

III. RECONSTRUCTION OF INITIAL MODEL BY HYBRID METHOD

A. Hybrid method

For the chronicity and complicacy of the geological and hydrodynamical process, natural rock, such as sandstone, has a large number of discontinuous, multiscale, geometry-irregular pores, which form a complex porous structure with a fractal feature [58]. The simulation results were obtained by applying the hybrid method in the work of Liu *et al.* [28] by combining the PB and SA methods which are in agreement with the experimental measurements. Compared with the typical SA method, the hybrid method improves the transport properties of the digital cores [28]. Moreover, by comparing with the PB method, the fractal feature is improved [1]. In the present study, the PB method reconstructed the initial digital core [Fig. 2(e); please see the Supplemental Material [40] for the flow chart of the PB method (Fig. 1S)] using the 2D data of the reservoir [Figs. 2(a) and Fig. 2(b)]. Principally, it contained three processes, for instance, sedimentation, compaction, and diagenesis [Figs. 2(c) and Fig. 2(d)]. The sedimentation process was the successive deposition of individual grains in a gravitational field. Each grain fell under the action of its own weight until it attained a stable position at a local or global minimum in potential energy. Grains were relaxed into a local minimum if the sedimentation process occurred in a low energy environment and into a global minimum if it occurred in a high energy environment. The compaction process was modeled as a linear process according to the formula

$$Z = Z_0(1 - \beta_z), \tag{1}$$

where Z is the new vertical position and Z_0 is the original vertical position. The degree of compaction and the porosity of the digital core was controlled by the compaction factor β_z [59]. Further details of the PB method are given in the study by Øren and Bakke [1].

However, the purpose of our work was to reconstruct a digital core containing clay minerals. So in the process of sedimentation, the main difference was that the content of clay minerals was considered during random choosing of the grains. The size of the sediment grains was not only decided by the initial grain size distribution, but also the volume ratio of rock and clay minerals (Fig. 1).

The hybrid method combined the PB and SA methods to reconstruct the digital core (Fig. 3; please see the Supplemental Material [40] for the flow chart and schematic diagram of the

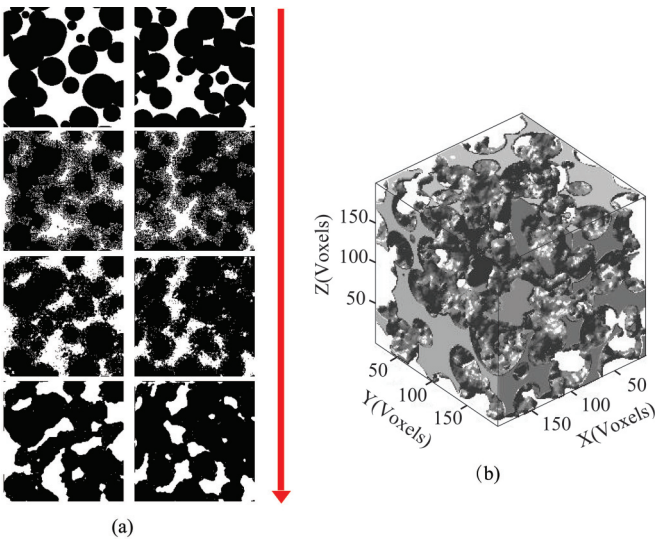


FIG. 3. Digital core based on the hybrid method. (a) The process of the reconstructed model by the hybrid method (2D slices). (b) The 3D digital core reconstructed by the hybrid method.

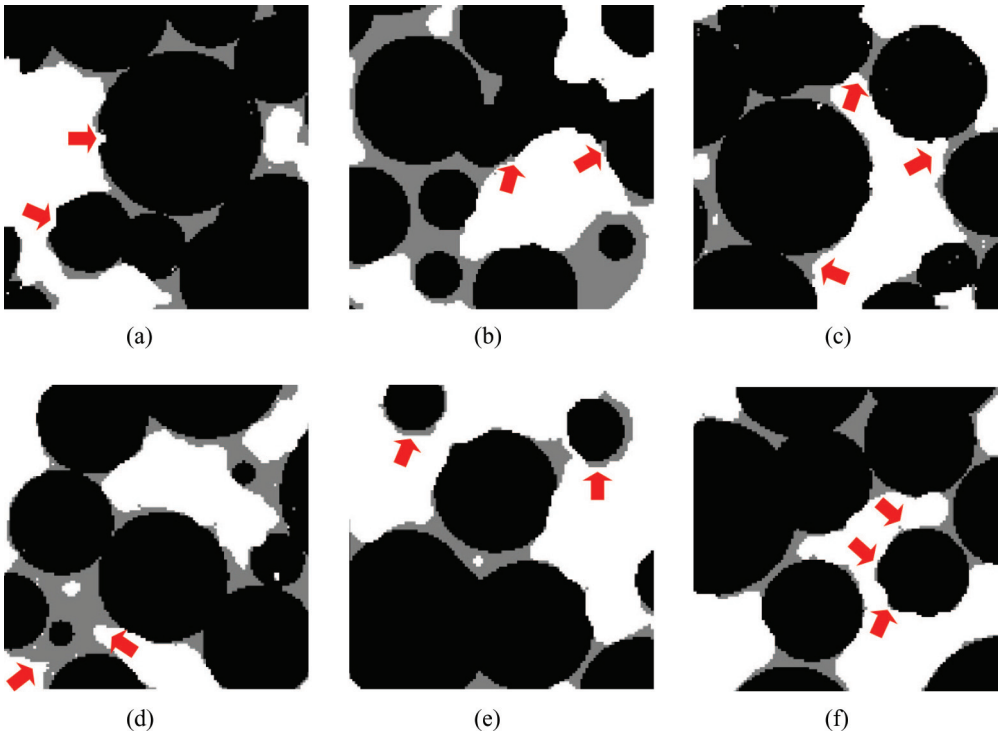


FIG. 4. Distribution patterns of clay clusters on the surface of the grain after division of rock matrix and unclassified clay minerals. (a) Surface indentation of the grain. (b) Interactive filling between clay cluster and grains. (c) Intergrain filling between two grains. (d) Intergrain filling among multiple grains. (e) Surface filling of strati form cluster on the grain. (f) Scattered distribution on the grain. The black parts represent the rock matrix and the grayish parts represent the unclassified clay minerals.

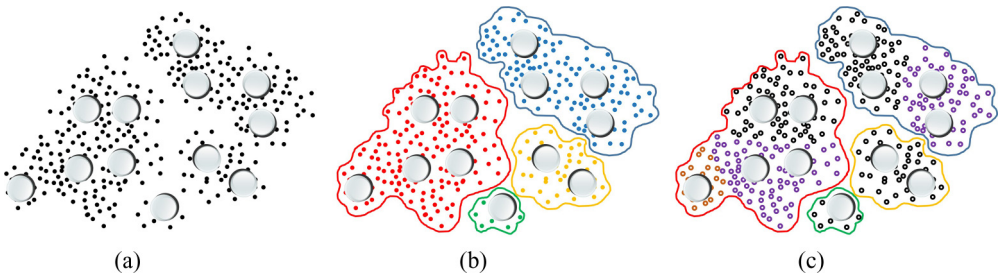
hybrid method [Figs. 2S and 6S)] based on the 2D data of the reservoir [28]. The casting thin section was taken as the reference model [Fig. 2(b)]. The procedures of the hybrid method are as follows: (i) Initialization: The digital core reconstructed by the PB method was used as the input for the hybrid method. Meanwhile, the initial temperature was set; the temperature played an important role in the cooling schedule. The initial temperature should be high enough to explore the whole solution space; the parameters of the initial system, such as AC function, LP function, fractal dimension, and energy (E) were calculated. The energy of the system was defined as the sum of the squares of the statistical properties' differences between the reference model and the reconstructed model.

$$E = \sum_i \sigma_i [S_s(r) - S_0(r)]^2, \tag{2}$$

where $S_0(r)$ and $S_s(r)$ are the statistical properties of the reference model and the reconstructed model. σ_i is the weight of different statistical properties. (ii) Interchanging the arbitrary discrete voxels: The arbitrary discrete voxels were in both matrix and pores. When arbitrary voxels were surrounded by the voxels belonging to other part, the energy of the system would subsequently be higher than the voxels surrounded by the same portion. On the basis of the randomness of the SA method, the discrete voxels on the interface of one part were selected to interchange with the voxels in the other part. An alternative system was obtained by interchanging two arbitrarily selected voxels of the two parts. (iii) Updating the system: The parameters of this system, including AC function, LP function, fractal dimension, and energy were calculated at the same time. The energy difference (ΔE) between this system and the previously unchanged system was also obtained. If $\Delta E \leqslant 0$, the changed system would be updated unconditionally. Otherwise, the Metropolis criterion [60,61] was used to update the system. The system could be updated with a probability (P); if the system could not be updated, then it would go back to step (ii).

$$P = \begin{cases} 1 & \Delta E \leqslant 0 \\ \exp\left(-\frac{\Delta E}{T}\right) & \Delta E > 0. \end{cases} \tag{3}$$

FIG. 5. Cluster identification by the HK algorithm. (a) 2D slice after division of rock matrix and unclassified clay minerals. (b) Cluster identification by the HK algorithm (different colors represent the labeled clay clusters).



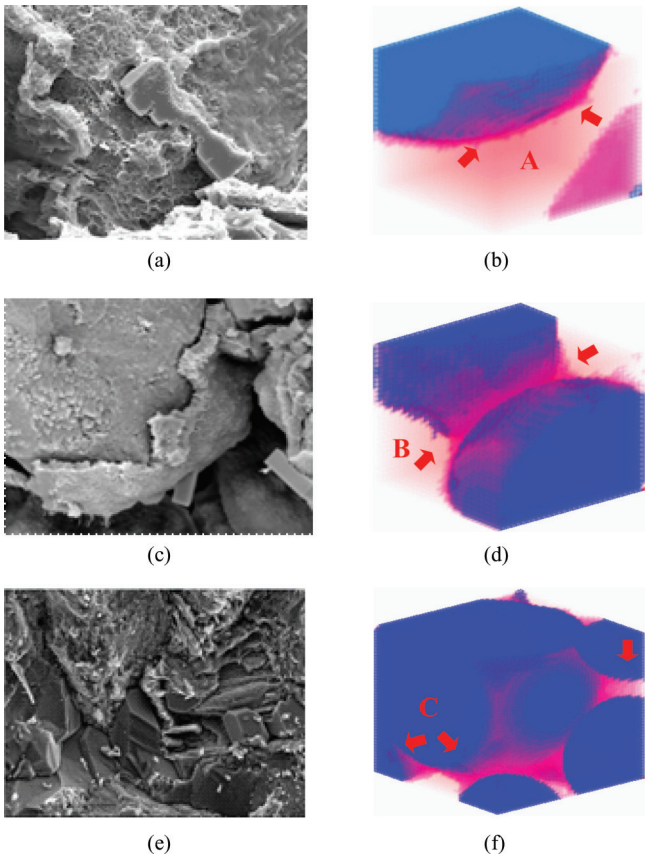
(a) (b) (c)

FIG. 6. Schematic of clustering algorithm. (a) Initial system. (b) System labeled by the HK algorithm. (c) System labeled and divided by the HK and K-means algorithms.

(iv) Cooling: If ΔE was less than the minimum set point, the temperature could be decreased. The temperature was updated by

$$T_{k+1} = \omega T_k, \quad 0 < \beta < 1, \tag{4}$$

where k is the number of generations [62,63]. As a matter of fact, the cooling schedule could be adjusted by modifying parameter ω, which was set between 0.5 and 0.999 in the simulations to ensure the optimization of the system. Meanwhile, the failure rate (f_f) was set up in order to avoid the unnecessary cooling caused by the immediate increase of energy. The failure rate (f_f) is shown by $f_f = \frac{N_f}{N}$. Whereas N_f is the number of times of failure update caused by the increase of energy, N is the total number of update times of the update. If f_f was more than the set value, the temperature could be decreased by an equal interval. (v) Terminate: The process could be terminated when the temperature dropped to the set value or ΔE less than the specified value.

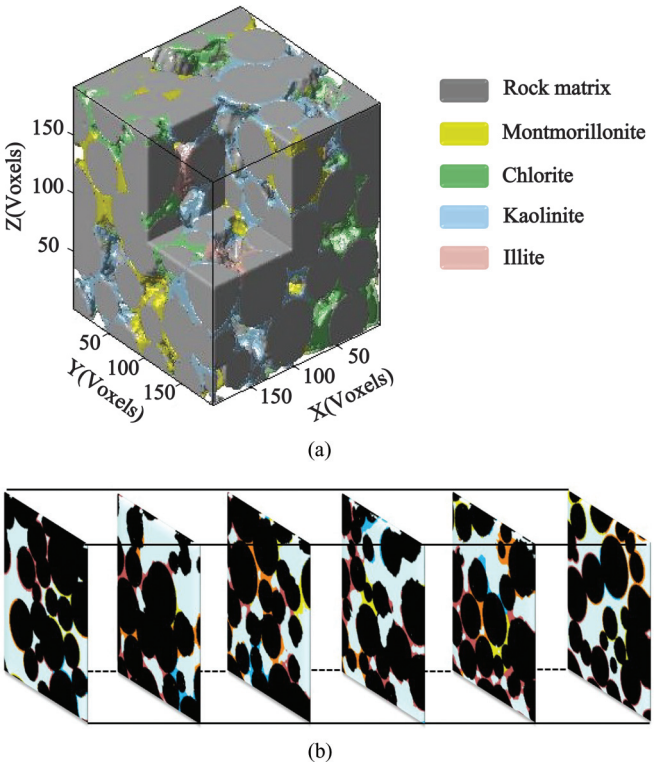


(a) (b) (c) (d) (e) (f)

FIG. 7. Structural features of clay clusters in the digital cores and scanning electron microscope (SEM). (a,c,e) SEM of surface filling of the grain, intergrain filling, and metasomatic texture. (b,d,f) textures of surface filling of the grain, intergrain filling, and metasomatism in the digital core.

B. Division of rock matrix and unclassified clay clusters

The digital cores with the same porosity were successively reconstructed by the PB method and the hybrid method in this paper. The clay content was considered in the sedimentation process. The results in the work of Coelho *et al.* [2] showed that the transport properties remain unchanged with the shape of the grains, except for oblate ellipsoids, which yield highly ordered structures; subsequently all the beds with identical



(a)

(b)

FIG. 8. Digital core containing different clay minerals. (a) The 3D digital core containing clay minerals. (b) 2D slices of digital core containing clay minerals.

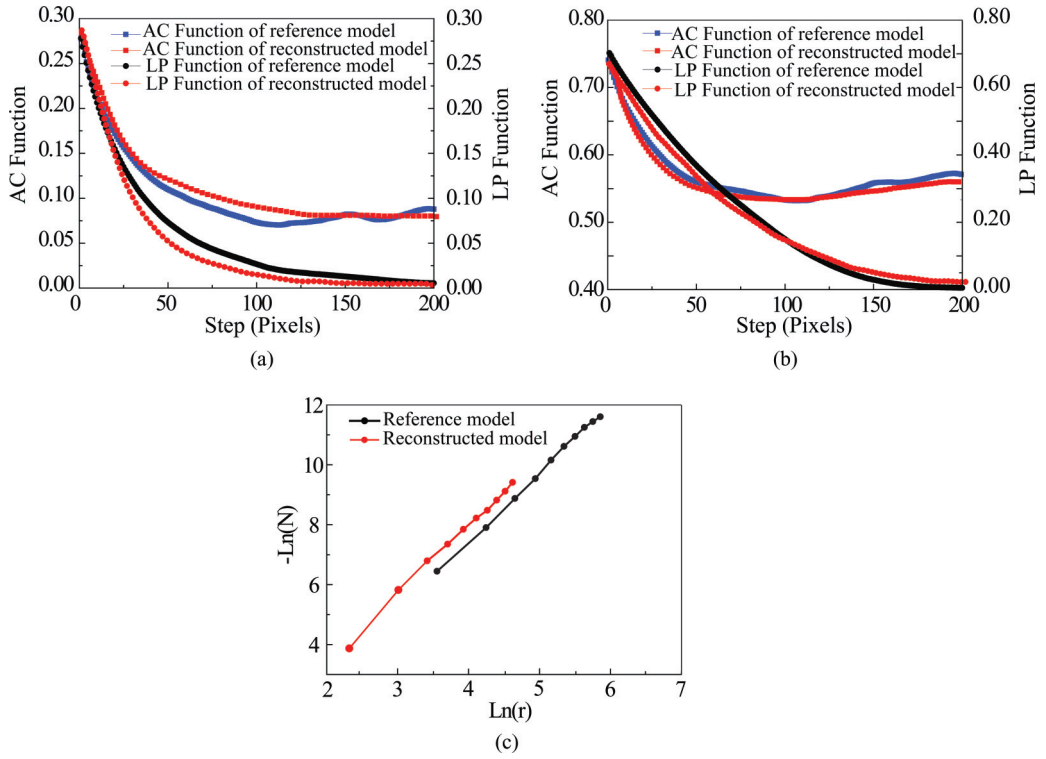


FIG. 9. Comparison of statistical functions in the digital core and the reference model in both pores and matrix. (a) AC and LP function of pores; (b) AC and LP function of matrix; (c) fractal dimension.

porosities share almost identical conductivities, permeability, and dispersion coefficients, regardless of grain shape or construction mode. The transport coefficients depend only upon porosity and an equivalent grain size [1–3,64]. It is a well known fact that, as reported by Sundari *et al.* [64], the volume fraction of grains may affect the porosity of the model, but in the present work, the initial model reconstructed by the PB method was taken as the input for the hybrid method. Therefore, the volume fraction of grains remained unchanged.

Jin *et al.* [39] proposed the area between the surface of the grain and the circumscribed polygon of a grain as the clay minerals. However, the cementing forms of the clay minerals are simple and the species of the clay minerals are unclassified. Inspired by the work performed by Jin *et al.* [29], in the present work, the uncharacterized spherical-like grains in the digital core by the hybrid method were compared with the initial spherical grains in the digital core by the PB method. The digital core was divided into three parts: pores (P), rock matrix (R), and unclassified clay minerals (C).

The clay minerals undergo various processes, mainly sedimentation and diagenesis. Various textures of clay minerals are lining, centroclinal, fan shaped, lepidoblastic, metasomatic, and others [57]. Generally, they can be regarded as clay clusters with variant shapes and sizes on the spherical-like grains. In fact, the uneven fractal structures on the spherical-like grains in the digital core reconstructed by the hybrid method could be considered as the results of the geological process. The structures of the clay clusters on the spherical-like grains in the digital core were similar to the textures in the reservoir as shown in Fig. 4. The most common forms were surface filling of grain; intergrain filling, film type, and metasomatic texture.

IV. DIVISION AND STATISTICS OF CLAY CLUSTERS

A. Statistics of clay clusters based on the HK algorithm

The HK algorithm is applied [please see the Supplemental Material [40] for the flow chart of the HK algorithm (Fig. 3S)] to quantities on 2D or 3D lattices [48]. The objective of the HK algorithm is to deal with infinite random binary lattices containing two classes of sites denoted as occupied and unoccupied sites [42]. In the present work, part C and other parts were taken as a binary lattice to implement the algorithm; we assumed that the concentration of part C was p, which was also the probability of a voxel to be occupied by C; when the voxel in the binary system was occupied by C, then it could be labeled as m_t^α, where α is a cluster identifier, t is the number of times of using the cluster identifier. The algorithm allowed multiple label assignment to a cluster α [65]. The labels were a set of natural numbers:

$$\left\{m_1^\alpha, m_2^\alpha, \ldots, m_s^\alpha, \ldots, m_t^\alpha\right\}. \tag{5}$$

In this set of natural numbers, one number was considered as the proper cluster label. The smallest number m_s^α was termed as proper cluster label. In fact, it was not mandatory to choose the smallest cluster number as the proper label. Therefore, choosing any other label in Eq. (5) could be appropriate. The following set of integers defines the relationship between the cluster labels:

$$\left\{N\left(m_1^\alpha\right), N\left(m_2^\alpha\right), \ldots, N\left(m_s^\alpha\right), \ldots, N\left(m_t^\alpha\right)\right\}. \tag{6}$$

In Eq. (6), $N(m_s^\alpha)$ is the only positive member of the set. It denotes the number of part C which belonged to a cluster. When the time was labeled by t and the voxels of the C part

TABLE I. Geometry parameters of the reconstructed and reference models.

	Reference model	Reconstructed model	Deviation
Porosity (%)	26.38	26.38	0.00
Effective porosity (%)	26.38	26.38	0.00
Average size of pore radius (m)	2.20×10^{-4}	2.18×10^{-4}	0.90%
Average size of throat radius (m)	8.50×10^{-5}	8.40×10^{-5}	1.17%

belonging to cluster α were less than the last label, then the difference should be the number of the C part belonging to a cluster α. The rest of the numbers in Eq. (6) were negative integers. They correlated the labels m_t^α to the proper label m_s^α. These labels were related to m_s^α via the following set of equations:

$$m_r^\alpha = -N\left(m_t^\alpha\right),\ m_q^\alpha = -N\left(m_r^\alpha\right), \ldots,\ m_s^\alpha = -N\left(m_p^\alpha\right). \quad (7)$$

Equation (7) represents a tree diagram where the root of the tree corresponded to the proper label m_s^α. All other labels were the nodes of the tree pointing directly or indirectly to other nodes of the root. As shown in Fig. 5, after the application of the HK algorithm, the connected clay clusters were labeled.

In the current study, the voxels of part C were labeled from the initial voxel of the digital core using the HK algorithm and the whole digital core was swept. Parts P and R were both tagged as 0, while the voxels of part C belonging to the same cluster were represented by the same tag. If the neighboring voxels of part C either belonged to the P or R part, then the current voxel should be given a new cluster label. If one of the neighboring voxels of C part had been labeled, then the current voxel should be given the same cluster label with that voxel (the smaller label). In the third condition, if more than one neighboring voxel of part C had been labeled and the labels were very different, the voxel should be given the same cluster label (the smallest label).

B. Division of clay clusters based on the K-means algorithm

The K-means algorithm is one of the most widely used clustering algorithms to appropriately process a large amount of data based on the K-means potential function [54,56]. The data of the K-means algorithm in the reconstructed digital

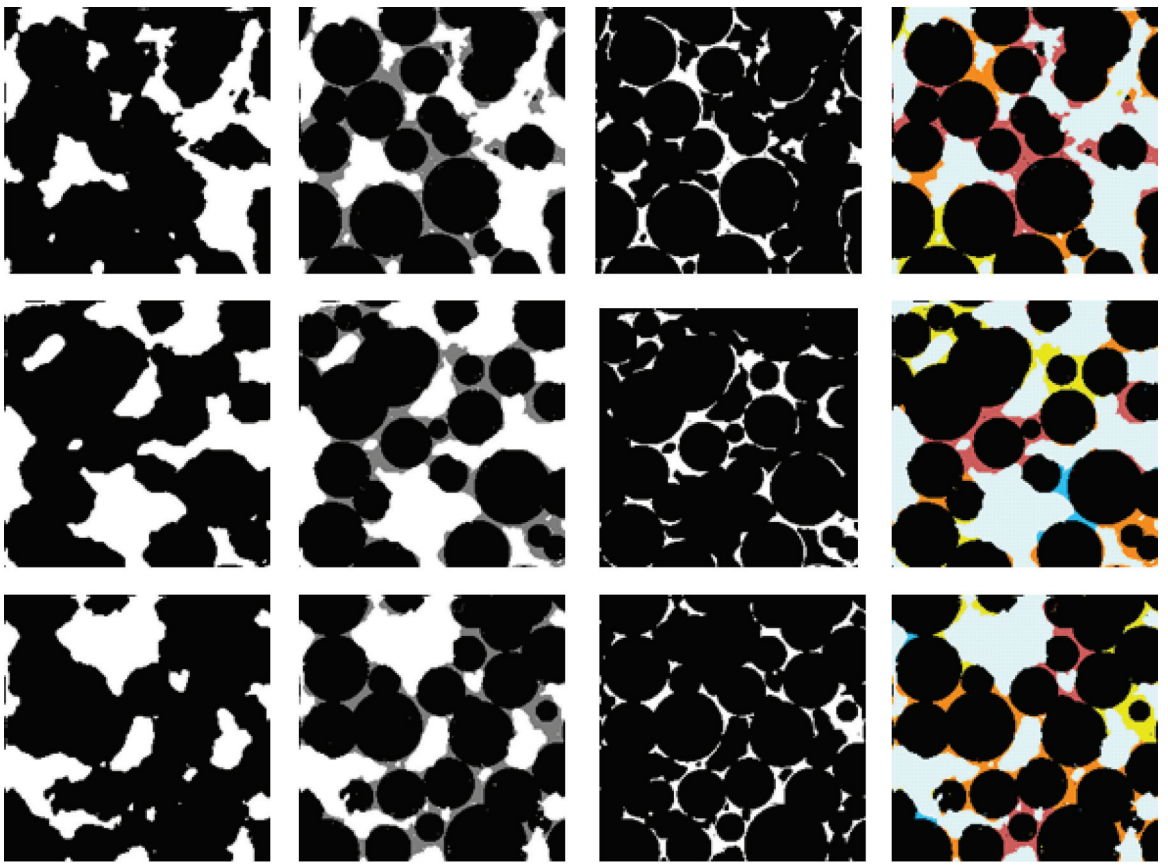


FIG. 10. The process of division of clay clusters by the clustering algorithm. (a) The 2D slices of the reconstructed model by the hybrid method. (b) The 2D slices after division of rock matrix and unclassified clay clusters. (c) The 2D slices of clay clusters in the reconstructed model. (d) The 2D slices after applying both HK and K-means algorithm.

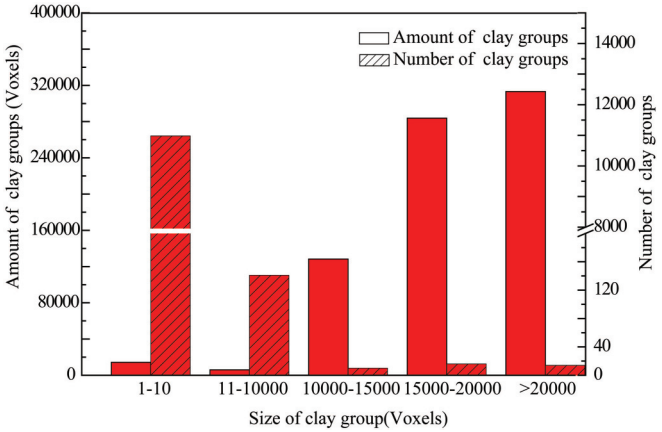


FIG. 11. Size distribution of clay clusters in the reconstructed model.

core were large communicating clay clusters obtained through the HK algorithm. The K-means algorithm finds a partition such that the squared error between the empirical mean of a cluster center and the voxels is minimized. Let c_i be the center of cluster R_i. The squared error between c_i and the voxels in cluster R_i is defined as

$$Z(C_i) = \sum_{x_i \subseteq R_i} \|x_i - c_i\|^2. \tag{8}$$

The goal of the K-means algorithm is to minimize the sum of the squared error over all clusters.

$$Z = \sum_{i=1}^{k} \sum_{x_i \subseteq R_i} \|x_i - c_i\|^2, \tag{9}$$

where $\|x_i - c_i\|^2$ is the distance between a voxel x_i and the cluster center c_i. The Euclidean distance between the kth cluster center and the ith voxel is as given by Eq. (10).

$$d(x_i, c_i) = \sqrt{(x_i - c_i)^2}. \tag{10}$$

The detailed illustration for the K-means algorithm applied in division of large clay clusters is as follows: (i) The k_{th} initial clustering centers were arbitrarily chosen from the set of voxels belonging to the large connecting clay cluster. (ii) The distance between the clustering centers and the voxels in the large communicating clay cluster were calculated and each voxel x_i was assigned to the nearest cluster center c_i. (iii) The updated clustering centers were obtained by calculating the mean distance between the clustering center and the voxel of the clay clusters. (iv) If the updated clustering centers were still changing, then the process went back to step (ii); otherwise it moved to step (v). (v) The position of cluster centers and divided clay clusters with size and number were the output. The schematic and process of the clustering algorithm is shown in Figs. 6 and 4S [please see the Supplemental Material [40] for the flow chart of the K-means algorithm (Fig. 4S)]. The initial clay clusters were attached in a disorderly manner on the grains [Fig. 6(a)]; after the application of the HK algorithm on the clay clusters, the number and size of connected unclassified clay clusters were obtained [Fig. 6(b)]; the large clay clusters were divided into smaller ones by means of the K-means algorithm [Fig. 6(c)]. The K-means algorithm divided the large communicating clay clusters into smaller ones with the structure of the digital core unchanged. At the same time, the porosity of the model remained unchanged. Finally, all the clay clusters in the digital core were labeled and clustered with number and size on the basis of both the HK and K-means algorithms.

C. Structural division of clay clusters

Montmorillonite, illite, chlorite, and kaolinite are the common clay minerals in the reservoir with various textures, for instance, intergrain filling, film-type filling, metasomatic texture, and others [66]. The different textures of clay minerals in the digital core included the surface filling of the grain (surface filling of single grain, lamellar filling on the surface of the grain, and others), intergrain filling (between two grains, among multiple grains), and metasomatic texture (filling inside the grain).

After the application of the clustering algorithm, the number and size distributions of the clay clusters were obtained. However, the clay minerals were not only determined by the size and the content, but also by the different textures. The clay clusters would be reflected as surface filling of the grain if the sites on the boundary of the clay cluster were related to the pore and the same grain [Fig. 7(a)]. Furthermore, in the reconstructed model the texture of the surface filling of the grain typically covered the surface filling texture, such as centroclinal, fan shaped, lepidoblastic, and others. The clay clusters would be considered as intergrain filling, if the neighboring sites of the clay cluster were occupied by the sites of pore and the different grains [Fig. 7(b)]. The intergrain filling typically covered the textures of clay clusters which were cemented and filled between different grains (two or more grains). If the neighboring sites of the clay cluster all belonged to the same grain of other kinds of minerals, then the clay clusters would be considered as metasomatic texture [Fig. 7(c)]. The metasomatic texture typically covered the textures of the clay clusters which were filled in other kinds of minerals. The clay clusters with metasomatic texture were scattered inside the grains. The smaller clusters (size <5 voxels) were taken as the object to judge whether the clay clusters were metasomatic texture or not. The textures of surface filling of the grain, intergrain filling, and metasomatic texture were tagged as A, B, and C, respectively [Fig. 7; please see the Supplemental Material [40] for the flow chart of structure judgment (Fig. 5S)].

V. RECONSTRUCTION OF DIGITAL CORE CONTAINING CLAY MINERALS

In the hybrid method, the size of the grain was determined by both the initial grain size distribution and the clay content in the reservoir. Hence the initial configuration was determined by the geometry and statistics of the reference model. After the division of rock matrix and unclassified clay minerals, the clay clusters were counted and divided on the basis of the clustering algorithm and the structural judgment. In comparison with the information about variant clay minerals of the reservoir, such as distributions and textures, the clay clusters were

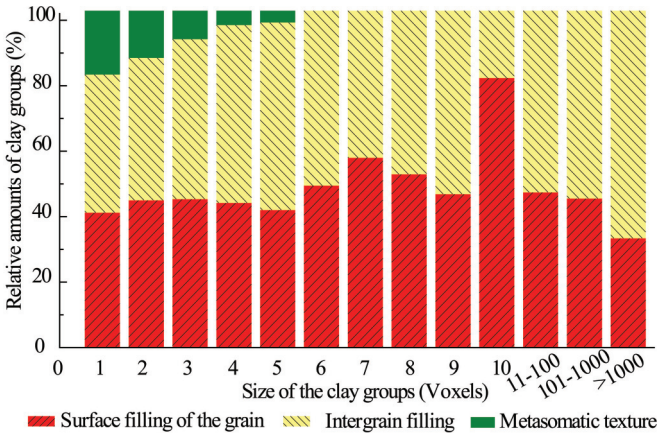


FIG. 12. Distribution of different textures of clay clusters in the digital core.

assigned different properties such as type, size, number, and texture. Eventually, the digital core containing clay minerals was reconstructed by the clay clusters were assigned different properties [Fig. 8; please see the Supplemental Material [40] for the process of reconstruction of a digital core containing clay minerals (Fig. 7S)], and all of the clay clusters were labeled with type, size, number, and texture.

VI. RESULT AND DISCUSSION

A. Geometry and statistics of the digital core

The digital core had not only the dispersion of grains, but also the multiphase systems of arbitrary geometry. All the work of this study was based on the model reconstructed by the hybrid method. Although the clay minerals were divided and reconstructed in the final digital core, but the morphological structure of the digital core remained unchanged. The accuracy of the hybrid method was systematically proved in the work of Liu *et al.* [28]; the simulation results of the hybrid method were compared with different methods (experiment, x-ray CT image, and typical simulated annealing algorithm). Consequently, the identical porosity, the transport properties of the formation factor, permeability, and the local porosity are similar. The simulated results of formation factor and permeability suggest that the 3D digital core reconstructed by the hybrid method has similar transport properties as the x-ray microimage. These results were in agreement with the experimental measurements. Therefore, in this work, we just discussed geometry parameters of the model and the correlation functions used in the hybrid method.

There are a variety of correlation functions that can be used in the procedure of reconstruction, including the AC function, LP function, two-point cluster function, chord-length distribution function, and many others [9,28,50,67,68]. In the present study, the AC function, LP function, and fractal dimension were applied as the control and statistical functions. These three statistical functions contain substantial structural information about the digital core and yet are simple enough to be implemented. Meanwhile, the three control functions reflect the correlation between the digital core and the reference model.

In a statistically inhomogeneous system, the AC function is defined as

$$A_2^{(j)}(\mathbf{r}_1,\mathbf{r}_2) = \langle I^{(j)}(\mathbf{r}_1) I^{(j)}(\mathbf{r}_2) \rangle, \tag{11}$$

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are two arbitrary points in the system, the angular brackets denote an ensemble average, and the characteristic function $I^{(j)}(\mathbf{r})$ is defined as

$$I^{(j)}(\mathbf{r}) = \begin{cases} 1, & \text{when } \mathbf{r} \text{ is in phase } j \\ 0, & \text{otherwise} \end{cases}. \tag{12}$$

The AC function can be interpreted as the probability of finding two points at positions $\mathbf{r}_1$ and $\mathbf{r}_2$ both in phase j. Another important statistical function of the basic structure is the LP function $L^{(j)}(\mathbf{r}_1,\mathbf{r}_2)$. The LP function is defined as the probability of finding a line segment spanning $\mathbf{r}_1$ to $\mathbf{r}_2$ that lies entirely in phase j. The LP function contains some connection information and it reflects certain long-range information about the system. In the isotropic medium, $A_2^{(j)}(\mathbf{r}_1,\mathbf{r}_2)$ and $L^{(j)}(\mathbf{r}_1,\mathbf{r}_2)$ depend only on the distance $r = |\mathbf{r}_1 - \mathbf{r}_2|$ between two points and can be expressed simply as $A_2^{(j)}(r)$ and $L^{(j)}(r)$. For all media with a volume fraction of ϕ_j, $A_2^{(j)}(0) = \phi_j$ and $\lim_{r\to\infty} A_2^{(j)}(r) = \phi_j^2$, $L^{(j)}(0) = A_2^{(j)}(0) = \phi_j$.

The fractal dimension can reflect the structure irregularity and discontinuity [28]. The fractal dimension was determined by box-counting measurements and defined as

$$D_B = \lim_{k\to\infty} \frac{\ln N_{\delta_k}}{-\ln \delta_k}, \tag{13}$$

where δ_k is the side length of a square grid, N_{δ_k} is the number of square grids with the side length of δ_k.

The statistical functions control the morphological structure of the digital core. The comparisons for the reference model and reconstructed digital core are presented in Fig. 9. It was clear that the correlation functions of pores and matrix agreed well with that of the reference model and the isolated points in the initial system were gathered to form the pore space under the control of the correlation functions. The reconstructed digital core obtained via the hybrid method retained the connectivity and spatial distribution. The fractal dimension of the reservoir had a high fitting degree with the reference model; the fractal dimension was 2.31.

Meanwhile, the geometry parameters, such as distribution of pore and throat, are the important parameters of the formation. In this work, we used the method proposed by Hu and Blunt [6] and Okabe and Blunt [69], who used the maximal ball algorithm to extract the geometry parameters of the pore and throat. Table I compares some parameters of the pore and throat in the reference model and in the reconstructed model. The average radius of pore and throat were 218 and 84 μm, respectively. The deviations were $<1.17\%$. The geometry parameters of the reconstructed model were in good agreement with the reference model.

B. Statistics of clay clusters in the digital core

The initial configuration was established by the hybrid method [Fig. 10(a)], after the division of rock matrix and unclassified clay minerals in the reconstructed model [Fig. 10(b)]. The HK algorithm was applied to label all of the clay clusters

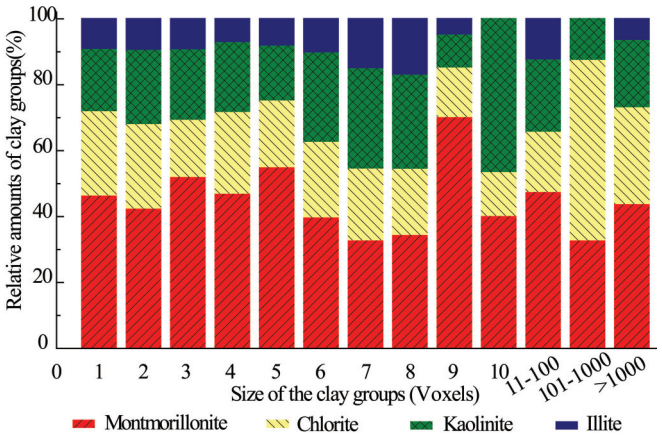


FIG. 13. Distribution of the different clay clusters in the digital core.

[Fig. 10(c)]. Meanwhile, the large communicating clay clusters were divided into smaller ones by applying the K-means clustering algorithm [Fig. 10(d)]. Ultimately, the number and size of the clay clusters were obtained by the clustering algorithm.

The biggest clay cluster in the digital core was 27 953 voxels, while the smallest clay cluster was 1 voxel equivalent to the number of 9432 (Fig. 11). Clay clusters with the size below 11 voxels accounted for about 1.91% of all clay clusters. The major clusters were almost 97.29% of all the clay clusters which were distributed with the size between 10 000 and 25 000 voxels. The distributive properties of the clay clusters were similar to those of the clay minerals in the reservoir with the characteristic of major big clusters and scattered smaller clusters.

The structures of the clay clusters were tagged as three major textures with surface filling of the grain, intergrain filling, and metasomatic texture. The 67.13% of the entire digital core was intergrain filling which accounted for 4685 clay clusters. The second major texture was surface filling of the grain with a value of 32.30% and it accounted for 4530 clusters in the reconstructed model. The metasomatic texture dispersed inside the grains was 0.28%, which corresponded to 1560 clay clusters in the digital core. The main textures of the large clusters were surface filling of the grain and intergrain filling. The smaller clusters presented (the size smaller than 5 voxels) were scattered in the model with three different textures. The relative amounts of clay clusters are shown in Fig. 12.

C. Distributions of clay clusters in the digital core

After the statistic and the judgment of the clay clusters in the reconstructed digital core, each clay cluster was assigned the properties of type, size, number, and texture. Compared with the type, content, distribution, and texture of clay minerals from the reservoir, in the digital core, the different kinds of clay minerals in the digital core were evenly distributed in terms of size (voxels). See Fig. 13.

Montmorillonite was the most common clay mineral in the reconstructed digital core. More specifically, the content of montmorillonite was 40.84%. The major textures of

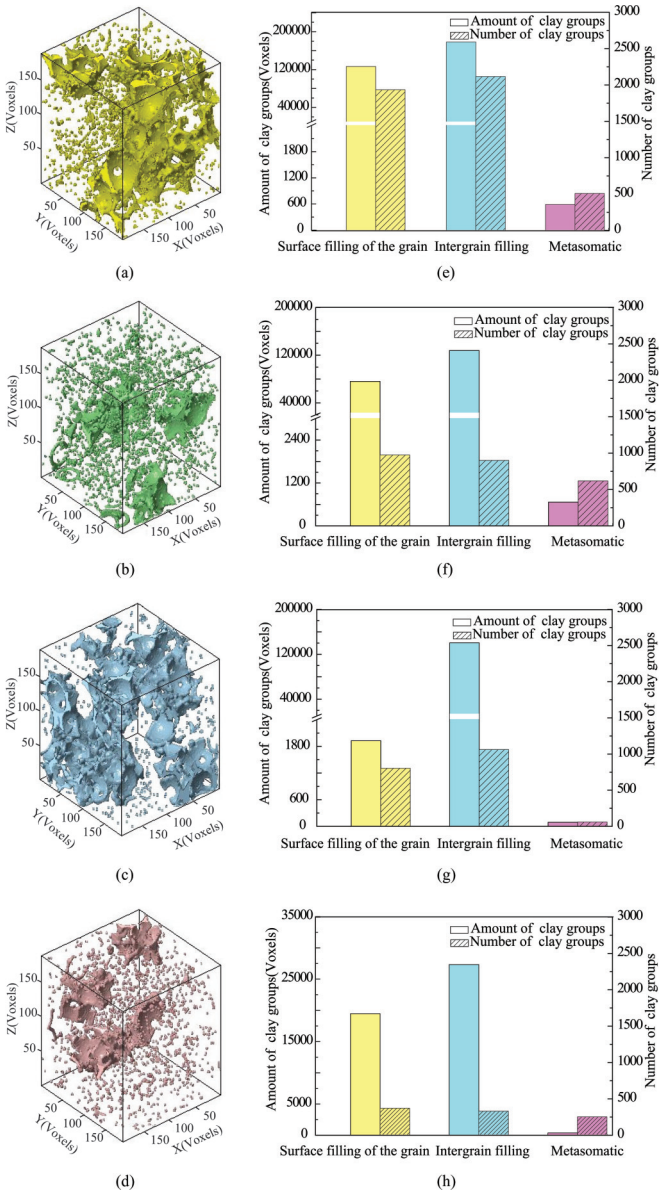


FIG. 14. Configurations and distributions of different clay minerals in the digital core. (a–d) The 3D configurations of montmorillonite, chlorite, kaolinite, and illite. (e–h) The amount and number distributions of clay clusters in the digital core.

montmorillonite were intergrain filling and surface filling of the grain which corresponded to numbers of 2117 and 1935, respectively. The intergrain filling of montmorillonite was 41.41%, and the surface filling of the grain was 58.39% in the digital core. The biggest clay cluster with the texture of intergrain filling was 22 716 voxels and the range of surface filling of the grain was from 1 to 21 273 voxels. The configuration of montmorillonite was mostly connected clay clusters on the matrix of the digital core [Figs. 14(a) and 14(b)]. Particularly, the content of chlorite in the digital core was 27.43%. Furthermore, surface filling of the grain was counted to be 37.14%, which corresponded to a number of 975. The biggest clay cluster with surface filling texture of grain was 21 193 voxels, whereas the smallest one was 1 voxel.

In addition, intergrain filling of chlorite accounted for 62.65% which was related to the number of 900. In contrast, the biggest clay cluster with the form of intergrain filling was 22 767 voxels and the smallest one was 1 voxel, so chlorite was often annular or connected clusters in the digital core [Figs. 14(c) and 14(d)]. It is a known fact that the principal texture of kaolinite is intergrain filling. In contrast, the major texture in the reconstructed model was intergrain filling. Therefore, the configuration of kaolinite was distributed continuously with connected clusters. The large clay clusters of kaolinite all belonged to the texture of intergrain filling and the content was 98.58%. The biggest cluster with intergrain filling was 27953 voxels [Figs. 14(e) and 14(f)]. The content of illite in the reconstructed model was only 6.28%. The textures of illite included intergrain filling, surface filling of the grain, and metasomatic texture. The intergrain filling and surface filling of the grain accounted for 58.12% and 41.32%, respectively [Figs. 14(g) and 14(h)]. Last but not least, metasomatic texture was also common in the montmorillonite and chlorite. The clay clusters with metasomatic texture were distributed sporadically (Fig. 14). The numbers of metasomatic texture in the montmorillonite, chlorite, and illite were 504, 619, and 244, respectively. The digital cores reconstructed by means of the clustering algorithm constitute clay minerals showing similar structure, distribution, and textures with the respective core from the reservoir.

VII. CONCLUSION

The clustering algorithm efficiently improved the simulation of clay minerals. In addition, the combination of HK and K-means algorithms provided an alternative method for microscopic simulation of a reservoir, especially for the differences of the clay minerals' properties in the sandstone reservoir. In this paper, the reconstructed digital core was constituted of various minerals. The initial model was acquired via the hybrid method having the same statistics and geometry as those of the reservoir. The HK algorithm was applied to gather the number and size of different connected clay clusters. The K-means clustering algorithm was applied to divide the large connecting clusters into smaller clusters on the basis of differences in the clusters' nature. Exclusive consideration of statistics and division of clay clusters was carried out which was solely based on types, content, and textures of clay minerals in the reservoir. The distribution of the clay minerals in the digital core was in accordance with that in the reservoir. There were three typical textures of clay clusters in the digital core such as metasomatic texture, surface filling of the grain, and intergrain filling. The distributions and textures of the clay minerals of the digital core were quite reasonable.

We considered the improvement of the numerical modeling technique by reconstruction of the various clay minerals with different types and textures. In this context, we believe that the algorithm which is proposed in this paper has wide applications to study and simulate different models containing clay minerals, such as the formation damage caused by the clay minerals during oil production, the differences of the lithology-electric relation caused by different clay minerals' characteristics, and the difference in adsorption of variant substances (gas, asphaltene, ions, and others) during the well production due to difference in clay minerals, and many others, some of which will be studied and reported on in the near future.

ACKNOWLEDGMENT

The authors express their appreciation to the National Natural Science Foundation of China (Grants No. 51074029, No. 51104173, and No. 51474034) for the financial support of this work.

[1] P.-E. Øren and S. Bakke, Transp. Porous Media **46**, 311 (2002).
[2] D. Coelho, J. F. Thovert, and P. M. Adler, Phys. Rev. E **55**, 1959 (1997).
[3] W. Zhu, W. Yu, and Y. Chen, J. Appl. Geophys. **85**, 37 (2012).
[4] Y. Yang, J. Yao, C. Wang, Y. Gao, Q. Zhang, S. An, and W. Song, J. Nat. Gas Sci. Eng. **27**, 496 (2015).
[5] J. Joos, T. Carraro, A. Weber, and E. Ivers-Tiffée, J. Power Sources **196**, 7302 (2011).
[6] H. Dong and M. J. Blunt, Phys. Rev. E **80**, 036307 (2009).
[7] P. Spanne, J. F. Thovert, C. J. Jacquin, W. B. Lindquist, K. W. Jones, and P. M. Adler, Phys. Rev. Lett. **73**, 2001 (1994).
[8] Y. Wang, T. Zhang, J. Liu, and J. Zhang, in *Proceedings of the International Conference on Computational Intelligence and Software Engineering, 2009, CiSE 2009* (IEEE, New York, 2009), p. 1.
[9] P. Tahmasebi and M. Sahimi, Phys. Rev. Lett. **110**, 078002 (2013).
[10] M. G. Rozman and M. Utz, Phys. Rev. E **63**, 066701 (2001).
[11] Y. Jiao, F. H. Stillinger, and S. Torquato, Phys. Rev. E **77**, 031135 (2008).
[12] J. F. Thovert, F. Yousefian, P. Spanne, C. G. Jacquin, and P. M. Adler, Phys. Rev. E **63**, 061307 (2001).
[13] C. L. Y. Yeong and S. Torquato, Phys. Rev. E **58**, 224 (1998).
[14] L. M. Pant, S. K. Mitra, and M. Secanell, Phys. Rev. E **90**, 023306 (2014).
[15] D. Chen, Q. Teng, X. He, Z. Xu, and Z. Li, Phys. Rev. E **89**, 013305 (2014).
[16] T. Zhang, D. Li, D. Lu, and J. Yang, Sci. China: Phys., Mech. Astron. **53**, 122 (2010).
[17] S. Chen, H. Li, and Y. Jiao, Phys. Rev. E **92**, 023301 (2015).
[18] L. Xu, X. Liu, and L. Liang, J. Nat. Gas Sci. Eng. **21**, 907 (2014).
[19] P. Tahmasebi, M. Sahimi, and J. Caers, Comput. Geosci. **67**, 75 (2014).
[20] H. Rezaee, D. Marcotte, P. Tahmasebi, and A. Saucier, Risk Assess. **29**, 893 (2015).
[21] P. Tahmasebi, A. Hezarkhani, and M. Sahimi, Comput. Geosci. **16**, 779 (2012).

[22] P. Tahmasebi and M. Sahimi, Water Resour. Res. **52**, 2074 (2016).
[23] H. Okabe and M. J. Blunt, Phys. Rev. E **70**, 066135 (2004).
[24] M. D. Rintoul and S. Torquato, J. Colloid Interface Sci. **186**, 467 (1997).
[25] C. J. Gommes, Y. Jiao, and S. Torquato, Phys. Rev. E **85**, 051140 (2012).
[26] C. Manwart and R. Hilfer, Phys. Rev. E **59**, 5596 (1999).
[27] M. Politis, E. Kikkinides, M. Kainourgiakis, and A. Stubos, Microporous Mesoporous Mater. **110**, 92 (2008).
[28] X. Liu, J. Sun, and H. Wang, Appl. Geophys. **6**, 105 (2009).
[29] G. Jin, T. W. Patzek, and D. B. Silin, in *Proceedings of SPE Annual Technical Conference and Exhibition, Houston, 2004* (Society of Petroleum Engineers, Richardson, TX, 2004).
[30] A. P. Roberts, Phys. Rev. E **56**, 3203 (1997).
[31] M. Gao, X. He, Q. Teng, C. Zuo, and D. Chen, Phys. Rev. E **91**, 013308 (2015).
[32] P. Tahmasebi, F. Javadpour, and M. Sahimi, Sci. Rep. **5**, 16373 (2015).
[33] P. Tahmasebi, F. Javadpour, M. Sahimi, and M. Piri, Adv. Water Resour. **89**, 91 (2016).
[34] V. Osipov, N. N. Bik, and N. Rumjantseva, Appl. Clay Sci. **2**, 363 (1987).
[35] S. Ashoori, M. Abdideh, and A. Alavi, Geocarto Int. **31**, 241 (2015).
[36] T. Saarenketo, J. Appl. Geophys. **40**, 73 (1998).
[37] G. P. Matthews, C. J. Ridgway, and J. S. Small, Mar. Pet. Geol. **13**, 581 (1996).
[38] A. Jullien, C. Proust, L. Le Forestier, and P. Baillif, Appl. Clay Sci. **21**, 143 (2002).
[39] G. Jin, T. Patzek, and D. Silin, Presented at Spe Western Regional/AAPG Pacific, 2003 (unpublished).
[40] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.96.043304 for the flow charts and schematic diagrams of the mentioned algorithms.
[41] J. Hoshen and R. Kopelman, Phys. Rev. B **14**, 3438 (1976).
[42] J. Hoshen, Pattern Recognit. Lett. **19**, 575 (1998).
[43] J. Hoshen, M. W. Berry, and K. S. Minser, Phys. Rev. E **56**, 1455 (1997).
[44] H. C. Kim and G. Cho, Ann. Nucl. Energy **23**, 1445 (1996).
[45] L. Zhang and N. A. Seaton, Chem. Eng. Sci. **51**, 3257 (1996).
[46] W. H. Jo and M. B. Ko, Macromol. **27**, 7815 (2002).
[47] S. Haimov, M. Michalev, A. Savchenko, and O. I. Yordanov, IEEE Trans. Geosci. **27**, 606 (1989).
[48] S. Frijters, T. Krueger, and J. Harting, Comput. Phys. Commun. **189**, 92 (2015).
[49] A. Margolina, H. Nakanishi, D. Stauffer, and H. E. Stanley, J. Phys. A: Gen. Phys. **17**, 1683 (1984).
[50] D. C. Rapaport, J. Phys. A: Gen. Phys. **19**, 291 (1998).
[51] D. C. Rapaport, J. Stat. Phys. **66**, 679 (1992).
[52] A. K. Jain, Pattern Recognit. Lett. **31**, 651 (2010).
[53] J. M. Constantin, M. W. Berry, and B. T. Vander Zanden, Int. J. High Perform. Comput. Appl. **11**, 34 (1997).
[54] Z. Huang, Data Min. Knowl. Discovery **2**, 283 (1998).
[55] A. K. Jain, ACM Comput. Surv. **31**, 264 (1999).
[56] A. Likas, N. Vlassis, and J. J. Verbeek, Pattern Recognit. **36**, 451 (2003).
[57] C. G. Whitney, Earth-Sci. Rev. **24**, 148 (1987).
[58] Y. Ju, J. Zheng, M. Epstein, L. Sudak, J. Wang, and X. Zhao, Comput. Methods Appl. Mech. Eng. **279**, 212 (2014).
[59] P. M. Adler, C. G. Jacquin, and J. F. Thovert, Water Resour. Res. **28**, 1571 (1992).
[60] J. Wang, S. Wei, S. Hao, D. Zhang, and Y. Ju, in *Proceedings of the 3rd International Congress on Image and Signal Processing (CISP), 2010* (unpublished), p. 2824.
[61] H. Das, P. T. Cummings, and L. Van, Comput. Chem. Eng. **14**, 1351 (1990).
[62] L. Ingber, Math. Comput. Modell. **12**, 967 (2002).
[63] X. Wang, X. Z. Gao, and S. J. Ovaska, in *Proceedings of the 2nd International and Interdisciplinary Conference on Adaptive Knowledge Representation and Reasoning (AKRR), 2008* (unpublished), p. 41.
[64] P. Sundari, U. Fauzi, Z. Irayani, and S. Viridi, in *Proceedings of the International Conference on Physics and Its Applications, 2011, Bandung, Indonesia*, edited by K. Basar and S. Viridi, AIP Conf. Proc. No. 1454 (AIP, New York, 2012), p. 121.
[65] A. Al-Futaisi and T. W. Patzek, Phys. A **321**, 665 (2003).
[66] R. H. Bennett and M. H. Hulbert, *Clay Microstructure* (Springer, Netherlands, 1986).
[67] G. Lei, P. C. Dong, Z. S. Wu, S. H. Gai, S. Y. Mo, and Z. Li, J. Nat. Gas Sci. Eng. **21**, 986 (2014).
[68] P. Tahmasebi and M. Sahimi, Phys. Rev. E **91**, 032401 (2015).
[69] H. Okabe and M. J. Blunt, J. Pet. Sci. Eng. **46**, 121 (2005).