This is the accepted manuscript made available via CHORUS. The article has been published as:

Reconstruction of nonstationary disordered materials and media: Watershed transform and cross-correlation function

Pejman Tahmasebi and Muhammad Sahimi

Phys. Rev. E **91**, 032401 — Published 3 March 2015

DOI: 10.1103/PhysRevE.91.032401

Reconstruction of Nonstationary Disordered Materials and Media: Watershed Transform and Cross-Correlation Function

Pejman Tahmasebi and Muhammad Sahimi†

Mork Family Department of Chemical Engineering & Materials Science, University of Southern California, Los Angeles, California 90089-1211

Nonstationary disordered materials and media - those for which the probability distribution function of any property varies spatially when shifted in space - are abundant and encountered in astrophysics, oceanography, air pollution patterns, large-scale porous media, biological tissues and organs, and composite materials. Their reconstruction and modeling is a notoriously difficult and largely unsolved problem. We propose a method for reconstructing a broad class of such media based on partitioning them into locally-stationary zones. Two methods are used for the patitioning. One is based on the Shannon entropy, while the second method utilizes a watershed transform. The locally-stationary zones are then reconstructed based on a cross-correlation function and one-dimensional raster path that we recently introduced [P. Tahmasebi and M. Sahimi, Phys. Rev. Lett. **110**, 078002 (2013)], with overlaps between the zones to ensure seamless transition from one zone to another. A large number of examples, including porous media, ecological systems, disordered materials and biological tissues and organs are reconstructed and analyzed to demonstrate the accuracy of the method.

PACS number(s): 47.46.+r, 81.05.Rm, 61.43.Bn, 05.20.-y, 02.70.-c, 89.90.+n

I. INTRODUCTION

Disordered media, ranging from man-made composite materials, to natural ones such as rock, biological tissues and organs, and tree patterns in forests are ubiquitous and of fundamental importance to many phenomena and processes of scientific and practical interests [1,2]. Thus, their characterization and modeling has been a long-standing problem. In principle, disordered materials and media may be divided into two broad classes. In one class are spatially-stationary media - those for which the probability distribution function (PDF) of any property does not change when shifted in space and, hence its various statistics do not vary in space. Modeling and analysis of stationary systems have made great progress [3-6], particularly over the last few years [7,8].

In the second class are spatially-nonstationary disordered (NSD) media, for which the PDF of any property and its various statistical properties do vary spatially, when shifted in space. Such media are also referred to has macroscopically-heterogeneous media, because there is no representative volume element such that if the media's properties are averaged over such a volume, they will not change if measured in larger volumes or length scales. A review of a broad class of disordered materials and media indicates that nonstationarity is more of a rule, than an exception. Examples of NSD systems are encountered in astrophys [9], oceanography [10], rock at large scales [11,12], spatial patterns of environmental pollution [13], and biological tissues and organs [14]. In addition, medical diagnostics based on computations with three-dimensional (3D) images [15] that are often nonstationary have become increasingly important.

Unlike stationary media though, the problem of modeling and analyzing the NSD systems has proven to be fraught with difficulties. Almost all of the past approaches to the problem suffer from shortcomings and inaccuracy, or are specific to a particular type of system without any generality. Even in the case of large-scale porous media, a highly important and much studied class of NSD systems, a methodology that can analyze them is still lacking, even though there are ample experimental data indicating that the spatial distributions of the porosity, permeability, the elastic moduli, and wave speeds of such porous media follow nonstationary stochastic distributions (see below).

The goal of this paper is to introduce a general method for accurate description of NSD media. We do this through an inverse method, usually referred to as a reconstruction technique

[1-8]: given a certain amount of data, one tries to construct a model of a NSD medium that closely matches the data, such that it also provides accurate predictions for those properties of the system for which there are no data or, if there such data, they are not used in the reconstruction. Exact reconstruction is not practical, but the method that we introduce is capable of generating accurate realizations for a wide variety of systems.

The rest of this paper is organized as follows. In the next section we describe how to partition a NSD system into locally-stationary zones. Sectin III describes the reconstruction of the locally-stationary zones based on a cross-correlation function and a one-dimensional raster path. To quantify the accuracy of the reconstruction we compute a multiple-point connectivity function, whjich is described in Sec. IV. The results are presented in Sec. V, while important generalizations of the method are discussed in Sec. VI. The paper is summarized in the last section.

II. SEGMENTATION OF A NONSTATIONARY SYSTEM INTO LOCALLY-STATIONARY REGIONS

Suppose that a d-dimensional data set is available for a NSD medium of interest, which may be numerical or in the form of an image, such as, for example, medical images or a cross-section of a disordered material. We consider a point $\mathbf{x}$ in the data set and a segment or window centered around it. The latter may be divided into two components: a stationary component and a nonstationary part that we refer to as *trends*. Due to the great progress over the last few years [3-8], the stationary component may be reconstructed accurately. Thus, the main task is to account for the trends that may be either deterministic or stochastic. If the trends are only mildly varying, or are of low-variance statistics type, then accounting for them is not too difficult. Our goal here is, however, devising a general method that is capable of modeling strongly-varying trends. Breaking a data set into a stationary part plus trends has already been developed for nonstationary time series and proven to be successful [16]. Here, we describe two methods for partitioning a nonstationary medium into a set of locally-stationary zones.

A. Partitioning based on the Shannon entropy

We describe the method for 2D media; its extension to 3D media will be clear. Let $\mathbf{G}$ represent the computational grid used in the reconstruction, partitioned into blocks or templates of size $L_x \times L_y$. Likewise, the target system (TS) to be reconstructed is partitioned into exactly the same number of blocks of the same size $L_x \times L_y$. The templates of $\mathbf{G}$ and the TS are denoted, respectively, by $\mathbf{T}_{\rm G}$ and $\mathbf{T}_{\rm TS}$. We denote by $\mathbf{D}_T(\mathbf{u})$ the *data event* at position $\mathbf{u}$ in $\mathbf{T}_{\rm G}$, i.e. a realization of the disorder generated to reconstruct the template $\mathbf{T}_{\rm G}$ centered at $\mathbf{u}$. We use "event" instead of "set," because the realization of the disorder may change during the reconstruction. Between every two neighboring blocks is an overlap region $\mathbf{O}$ of size $\ell_x \times \ell_y$, representing $\mathbf{D}_T$, where $\ell_x = L_x$ and $\ell_y \ll L_y$, if the overlap region is between two neighboring blocks in the vertical (y) direction, and $\ell_y = L_y$ and $\ell_x \ll L_x$, if the $\mathbf{O}$ is between two neighboring templates in the x direction; see Fig. 1. The purpose of the $\mathbf{O}$ regions is to preserve the continuity near the common boundary between two neighboring blocks within a stationary segment i, referred to as SS_i, as well as between neighboring blocks, but in two distinct segments, SS_i and SS_j.

Figure 1 shows the schematic representation of the approach. The segmentation is carried out along a 1D raster path, as shown in Fig. 1(a). Suppose, for example, that the 1D raster path is along the horizontal (x) direction. The algorithm begins at the path's origin in $\mathbf{G}$ - the leftmost bottom template, $\mathbf{T}_{\rm G}^{(1)}$ - and moves along the raster path. It first identifies those blocks of the TS that contain $\mathbf{T}_{\rm TS}^{(1)}$ and constitute a locally-stationary region. To identify them, one begins with $\mathbf{T}_{\rm TS}^{(1)}$ and computes its Shannon entropy [17], given by,

$$S_1 = -\sum_{i=1}^{n} p_i \ln p_i \ , \tag{1}$$

with $p_i = {\rm histogramofsamplei}/(\text{length of the sample})$ being the probability of having a state i in the TS, and n the number of pixels in the block. S_1 is calculated by the same method, if the TS is not represented by an image, but by a data set. Then, beginning with $\mathbf{T}_{\rm TS}^{(1)}$, we compute the corresponding entropies of the nearest neighbors, next nearest neighbors, etc., of $\mathbf{T}_{\rm TS}^{(1)}$ in both directions, and denoted them by $S_2,\ S_3,\cdots$. So long as entropies $S_1,\ S_2,\cdots$ constitute a second-order stationary set (or third order, if need be), i.e. one for which the average and variance (and the third moment in case of third-order stationarity) are spatially invariant, we consider the blocks as belonging to the same locally-stationary zone. But, if the stationarity of the computed entropies is lost, the search stops and one obtains a zone of the TS of size $m_1 \times n_1$ blocks that is locally stationary. Figure 1(b) shows a 3×3 zone. The same procedure is used for the rest of the TS blocks in order to identify other locally-stationary zones. Figures

1(c) - 1(e) present several such segments. Each zone is characterized by its own statistics, and the trends are the variations of such statistics between the zones.

B. Partitioning based on the watershed transform

The watershed transform (WST) is a method used for partitioning a given system into various distinct parts, and is based on an image. In doing so, the general shapes and topography of three distinct set of points are considered that are, (i) the local minima; (ii) the points between two different local minima, and (iii) the points with a probability (of occurence) equal to that of the local minima. The idea behind the WST is quite simple: Suppose that there is a "hole" in each local minimum through which water can flow with a constant flux. After a while, one sees the separation "bridges" or lines that prevent merging the water from different holes. Such lines represent the watershed boundaries that are good indicators for feature partitioning [18-21].

An important application of the WST is identifying similar and distinct segments in an image. Clearly, if a region contains small variations in some properties, it also has small gradients in the same properties. This is taken advantage of for identifying the distinct segments. First, the system is partitioned into cells or blocks. Then, each cell is expanded by adding the boundary points to it, i.e. the points in the neighboring cells. This is done by computing a distance function between two points $\mathbf{x}$ and $\mathbf{y}$, defined by

$$d(\mathbf{x},\mathbf{y}) = inf\left\{\int_{\gamma} ||\boldsymbol{\nabla} f[\gamma(s)]||ds\right\} \tag{2}$$

which defines a geodesic distance, where the infimum is over all the paths γ inside the domain that we consider. The topological distance between a point $\mathbf{x}$ and a set in the domain is the one that minimizes d, which also has the minimum slope among all the paths. In the present case the function f is simply the pixeled image TS. Clearly, if two pixels are similar, or if the average value of the pixels within two neighboring cells are equal or very close, then, $d \approx 0$. Thus, adding a new point is continued only if that point is at a smaller distance d to the region than to the other regions (hence, smaller or zero gradient in the property). The process of adding the points continues until there is no eligible point. Then, the remaining points represent the watershed lines or boundaries.

The WST method is most useful for use in identification of locally-stationary zones of disordered media in which there are large contrasts between the various zones. An example is a fractured porous medium in which not only the permeabilities of the porous matrix and the fractures differ by orders of magnitudes, the distribution of the orientations of the fractures may also vary from zone to zone. Another system for which the WST method is useful is a two-phase material in which the conductivities or elastic properties of the two phases differ very significantly. An example is shown in Fig. 2(a) in which the dark areas represent, for example, high conductivity or permeability regions, while the white areas indicate the opposite. It is straightforward to show that the medium is nonstationary. In this particular case one can visually detect three different zones. But, the aim is to use the WST to systematically identify such zones in a complex NSD medium.

The WST method is suitable for such systems due to the essence of the algorithm, namely, the aformentioned "flowing water" that creates large contrasts between the watershed lines and the catchment basins. Without defining any number of regions, or partitioning the system into several regions, we use the WST to identify and separate the locally-stationary regions. If, however, we use the algorithm suitable for stationary systems [8] for reconstructing the medium shown in Fig. 2(a), we obtain the realization shown in Fig. 2(b) that cannot clearly distinguish the distinct directional dark regions in various zones.

In practice, given the TS, an initial segmentation procedure consists of, (a) clustering neighboring pixels following an order based on increasing intensity gradients (and, hence, the distance d), and (b) splitting the TS into different locally-stationary zones. The equivalent to the watershed line is the line separating two different locally-stationary zones. Consider two "sinks," such as, for example, two opposite boundaries of a NSD medium, such as a fractured porous medium. If the TS is already pixeled, then one can directly use it. Otherwise, it is first partitioned into cells or blocks that are ranked according to the intensity of the gradients, leading to a ranked surface. They are sequentially occupied according to the rank, from the lowest to the highest. Neighboring occupied blocks are then connected and considered part of the same zone, except if their connection would promote the density of the two areas towards different sinks. This can happen when different orientations change the flow direction that creates a different sink. In this case, their connection is avoided as they should belong to different zones. The edge between them is part of the watershed. At the end of the process, the set of such edges forms one single watershed line that splits the TS into two separate zones.

Thus, the WST method may be summarized as follows:

(i) Partition the TS into cells or blocks of fixed sizes. If the TS is a pixeled image, it will not need partitioning.

(ii) Identify the edges between the background and the high contrast regions. If the disordered medium is more or less a binary one, i.e. one in which the two phases are more or less uniform, but with strong contrast between the two, we first carry out a pre-processing of the TS by filtering it. To do so, we use two matrices,

$$\mathbf{F}_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} , \quad \mathbf{F}_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \tag{3}$$

for filtering in the x and y directions in 2D systems (and similarly in 3D for which one must use nine matrices). Then, the quantities

$$\begin{aligned} C_x(n_1, n_2) &= \mathbf{F}_x * \mathcal{T} = \sum_{k_1=1}^{3} \sum_{k_2=1}^{3} F_x(k_1, k_2)\mathcal{T}(n_1 + k_1, n_2 + k_2) , \\ C_y(n_1, n_2) &= \mathbf{F}_y * \mathcal{T} = \sum_{k_1=1}^{3} \sum_{k_2=1}^{3} F_y(k_1, k_2)\mathcal{T}(n_1 + k_1, n_2 + k_2) , \end{aligned} \tag{4}$$

where $*$ denotes the convolution operator, $\mathcal{T}$ represents the target system TS, and the coordinates of the pixels are (n_1, n_2). Finally, the quantity

$$G(n_1, n_2) = \sqrt{C_x^2(n_1, n_2) + C_y^2(n_1, n_2)} , \tag{5}$$

is calculated. It is $G(n_1, n_2)$ that is used in the computations with the WST, rather than the TS itself, because the computations are typically much more accurate when a pre-processing is carried out and, thus, $G(n_1, n_2)$ is used.

(iii) Select a point at random, and calculate the distances d, defined by Eq. (2), between that point and the neighboring points.

(iv) Add the neighboring points to a list of the points that belong to the zone that contains the randomly-selected point, if the criterion for the addition is met. The criterion for addition is that the distance between two points must be smaller than a threshold value.

(v) If the list has been expanded (by adding new points to it), repeat (iv).

(vi) Label all the points in the immediate neighborhood of the list as the watershed line.

(vii) Repeat (iii) - (vi), until the identification of the locally-stationary zones is complete.

The same procedure is used if the TS is partitioned into blocks or cells, except that one first calculates the average property value (or pixels) within the cell, and then computes the distance function between a point representing the average and those in the neighboring blocks or cells. For example, the result for partitioning the nonstationary medium of Fig. 2(a) is shown in Fig. 3, indicating that the WST has correctly identified the three distinct nonstationary zones of the medium.

Note that one can use larger matrices $\mathbf{F}_x$ and $\mathbf{F}_y$, such as 5×5 or 7×7 matrices (in which case the sums in Eq. (4) will run from 1 to 5 or to 7, respectively), but doing so will also increase the computation time, and our own experience indicates that the 3×3 matrices suffice for good accuracy. One also can use other types of filters as a pre-processing step before using the WST. For example, by using an erosion or dilation filter one can remove or connect small features in a TS to other structures. Such filters allow one to increase the size of the system to be studied.

III. RECONSTRUCTION BASED ON A CROSS-CORRELATION FUNCTION

The next step is to reconstruct the identified locally-stationary zones. The reconstruction may be unconditional - one in which the reconstructed system does not have to honor *exactly* specific hard (quantitative) data in the TS - or conditional, one in which a certain amount of hard data must be honored. We first describe the unconditional case, after which the conditional case will be briefly described.

A. Unconditional reconstruction

The reconstruction of the locally-stationary zones is carried out by the method that we recently introduced [8] for stationary media, where we demonstrated that, using a *single* 2D slice of a 3D stationary medium as the input data, the method produces accurate realizations of the *entire* 3D medium. Thus, not only is the method accurate, it also addresses the long-standing practical problem of how to reconstruct a 3D medium based on 2D data. For completeness we describe the method here. Let $\mathrm{SS}_i(x, y)$ represent the datum at point (x, y) of a locally-stationary zone SS_i. Examining SS_i, one focuses on a portion $\mathbf{D}_T(\mathbf{u})$ of size $\ell_x \times \ell_y$ - the **O** region - and reconstructs it based on a matching between the **O** and the *entire* SS_i. We use a cross-correlation function (CCF) [8] to quantify the quality of the matching:

$$\mathcal{C}(j,k) = \sum_{x=0}^{\ell_x-1}\sum_{y=0}^{\ell_y-1} \mathcal{S}(x+j,y+k)\mathbf{D}_T(x,y)\ , \tag{6}$$

with j and k varying within the entire SS_i, where $\mathcal{S}$ is the stationary segment SS. Equation (6) indicates that the desired position of (j,k) - the best match with the SS_i - is one that maximizes $\mathcal{C}(j,k)$. Thus, the reconstruction of each locally-stationary segment proceeds as follows [8]. Note that Eq. (6) is completely similar to Eqs. (4); that is, the CCF is simply a convolution.

(i) As the location of the CCF's maximum is not known a priori, we use $1/\mathcal{C}$ and set a threshold $0 \leq \delta \leq 1$. If $\delta = 0$, the matching between the **O** regions and SS_1 is perfect, where as $\delta > 0$ generates an ensemble of realizations that do not match the SS_1 *exactly.* After some preliminary simulations we used, $\delta = 0.2$.

(ii) The blocks of **G** along the raster path, Fig. 1, are reconstructed one by one. For block 1 we generate several realizations of the disorder in SS_1, based on and constrained by its statistics (such as the volume fractions of the phases), or by sampling SS_1. For each realization the CCF between the **O** region of block 1 (to its right at the interface between blocks 1 and 2) and the *entire* SS_1 is computed. Any realization for which $1/\mathcal{C} < \delta$ is accepted; otherwise, it is discarded.

(iii) Once a large enough ensemble of acceptable realizations is generated, one of them is selected at random and inserted in block 1. Typically, one generates 30-50 acceptable realizations.

(iv) Next, block 2 along the raster path within SS_1 is filled up with an acceptable realization of the disorder by steps (ii) and (iii), the new **O** with the next block to be reconstructed is identified, and the procedure is repeated.

(v) Once the blocks of SS_1 along the raster path have been reconstructed, the algorithm moves into the next locally-stationary segment, SS_2, which may share a row of block with SS_1, or a row of the **O** regions, which make the transition from SS_1 to SS_2 seamless. The algorithm then reconstructs blocks 4 - 10 shown in Fig. 1(c), using only the properties of SS_2. Once the blocks along the 1D raster path at the bottom of the TS have been reconstructed, the path is continued on the next horizontal row of blocks shown in Fig. 1(b).

(vi) To reconstruct block 11 in Fig. 1, we proceed in the same way as for block 1 (using the image or statistics of SS_1). Note that block 12, for example, contains two $\mathbf{O}$ regions, one with block 10 and another with block 2. In this way all blocks of SS_1 and SS_2 are reconstructed. The algorithm then moves into the locally-stationary segment SS_3 and reconstructs its blocks along the raster path, until it reaches SS_4 and proceeds in a similar manner. The process continues until a realization of the completely reconstructed TS is obtained. Clearly, one can generate multiple realizations of the medium.

B. Conditional reconstruction

In this case the TS contains some hard data (HD) that must be honored exactly, due to which the data event $\mathbf{D}_T$ is the entire block, not just the overlap regions $\mathbf{O}$. Thus, conditional reconstruction is a two-step process. First, one identifies the realizations that honor the HD, and then determines the matching with $\mathbf{D}_T$. The algorithm computes the CCF and checks whether $1/\mathcal{C} < \delta$, where δ is the threshold. Those realizations that honor the HD are identified and one of them is selected randomly and inserted in the reconstruction grid $\mathbf{G}$.

A problem may arise if the TS is very large, or if the generated realizations do not honor the HD, leading to discontinuities and failure of the reconstruction. Such a case, which is extremely rare, may be addressed by several approaches. One is to increase the threshold δ to obtain new realizations that, although may have more significant differences with the $\mathbf{O}$ regions that are larger than those with a lower value of δ, they are still acceptable. But, increasing δ also allows the incorrect patterns to enter the ensemble, which may subsequently lead to the generation of a poor realization of disorder. An alternative approach is based on template splitting. First, the patterns of disorder that honor the HD with the initial threshold δ are identified. If, however, no such pattern exists, the sector template is split into smaller templates and the reconstruction proceeds with smaller parts. The splitting continues until the ensemble of the disorder patterns that honor the HD has at least one member.

If the medium under consideration is anisotropic, then, the proposed method will need modification for certain types of anisotropy. For example, if the anisotropy is due to layering, as is the case in large-scale porous media, then, each layer is by itself isotropic, and it is the stratification that gives rise to anisotropy. In that case, the method is still accurate. All one has to do is using smaller templates or grid blocks at the interfaces between the various layers. If, however, the anisotropy is due to other more complex factors, such as, for example, directiondependent pore-size distribution, then the method would need modification, a matter that is currently under study.

C. the parameters of the algorithm

The algorithm's parameters are the sizes of the templates and the overlap regions $\mathbf{O}$, and the threshold δ for accepting a realization of the disorder. The size of the $\mathbf{O}$ regions is a fraction of the templates' (see Fig. 1), and is fixed once the templates' is set. Our study of the stationary systems [8], as well as the nonstationary ones described here indicated that the templates' size is the most important factor. The size depends on the heterogeneity of the TS: for a relatively homogeneous TS a coarse grid suffices, whereas highly disordered media require grids with small blocks. Thus, the templates' size is decided by the desired precision.

IV. THE CONNECTIVITY FUNCTION

To quantify the accuracy of the reconstructed systems, we compute a general connectivity function for both the TS and its reconstruction, which is the multiple-point connectivity (MPC) function that has been used [8,22] in reconstructing large-scale geological formations. The MPC is the probability $p(\mathbf{r};s)$ of having a sequence of s points in a phase in a multiphase structure in a given direction $\mathbf{r}$. If an indicator function $I^{(i)}(\mathbf{u})$ is defined by

$$I^{(i)}(\mathbf{u}) = \begin{cases} 1\ , & \mathbf{u} \in \text{phase } i\ , \\ 0\ , & \text{otherwise}\ . \end{cases} \tag{7}$$

then, $p(\mathbf{r};s)$ is given by

$$p(\mathbf{r};s) = \text{Prob}\{I^{(i)}(\mathbf{u}) = 1,\ I^{(i)}(\mathbf{u}+\mathbf{r}) = 1,\ \cdots, I^{(i)}(\mathbf{u}+s\mathbf{r}) = 1\}\ . \tag{8}$$

For the results that are described in this paper we computed $p(\mathbf{r};s)$ for $s = 100$. We emphasize that matching $p(\mathbf{r};s)$ for a complex system with that of its reconstruction is a highly stringent test of the accuracy of the method, as one demands a large number of points to be in the same phase, as opposed to two-points requirement used in the past. For convenience we denote the MPC function by $p(\mathbf{r})$.

V. RESULTS

In what follows we present and discuss several examples, selected from a variety of disciplines. In almost all the cases, we generated several realizations of reconstructed TS, and computed the MPC function, all of which will be presented below.

A. Porous media

Various properties of large-scale porous media, such as oil reservoirs and groundwater aquifers, are correlated, with the correlations being long-ranged and non-decaying. The spatial distribution of the properties of such porous media, such as their permeability and elastic moduli are often described [12,23] by a fractional Brownian motion (FBM), a nonstationary stochastic process with a covariance function given by

$$C(r) - C(0) \sim r^{2H} \ , \tag{9}$$

where r is the distance between two points, and H is the Hurst exponent such that $H > 1/2$ ($H < 1/2$) indicates long-range positive or persistent (negative or antipersistent) correlations, whereas $H = 1/2$ represents the case in which the increments of the FBM are random. We generated a 3D system of size $120 \times 120 \times 64$ using the FBM with $H = 0.6$. Figure 4 presents the model porous medium along with three examples of its reconstructed model. Also shown is a comparison of the computed MPC functions $p(r)$ (where $r = |\mathbf{r}|$) for the realizations and the original TS. The agreement is illuminating in that, since the synthetic porous medium is precisely nonstationary, the method reproduces it very closely as well.

Another 3D example, shown in Fig. 5 [24], is also a synthetic complex porous medium of size $200 \times 150 \times 30$ in which the permeability varies spatially. Oil reservoirs typycally have such a structure [25]. The nonstationarity of the medium is due to different distributions of the orientations and thicknesses of the channels in various zones, and large contrasts between the channels and the background. Its reconstructed model along with a comparison between the computed MPC functions are also shown in Fig. 5. Once again, the agreement is excellent.

The next two examples are provided by 2D spatial distributions of ellipses. Figure 6 [26] presents one in which the TS is composed of a uniform background in which oriented ellipses, alligned in two distinct directions, at 45° and 135° relative to the horizontal line, have been distributed. The size of the system is 200×200. Shown also is the reconstructed system and a comparison of the computed MPC functions. The agreement is very good. A more complex example of the same type of system is shown in Fig. 7, where the TS contains a distribution of ellipses with some of them having two main variable properties, their orientation and thickness [27]. The orientations vary from the upper to lower parts, while the thickness of the particles is different in upper/lower and middle parts. The size of the system is 100×250. Figure 7 also presents reconstruction of the system and compares the computed MPC functions. Overall, the agreement is reasonable.

The last example is a cross section of a fracture surface of rock shown in Fig. 8, which was already described in the discussion of the watershed transform. The size of the system is 100×140, in which the fractures have a constant thickness, but are orientated [27] with the orientation distributions being different in various zones. Three examples of its reconstruction along with a comparison of the computed MPC functions are also shown in Fig. 8. The agreement is excellent. In fact, for some of the realizations one cannot even distinguish between the MPC functions for the TS and the realization.

B. Ecological systems

Next, we present the results for several ecological systems. The first example is one in which the TS is a shallow-water, tidal-dominated system [28] with a size 200×500. The comparison of the original TS and its reconstructed model, as well as the computed MPC functions for several rezalizations and the TS are shown in Fig. 9. The agreement between the two is good. Another example is provided by the delta of the Ganges River [29], a highly complex and nonstationary system. We reconstructed the map with a size of 400×300. The results are shown in Fig. 10. The agreement between the original map and the reconstructed model is very good, particularly given the complexity of the system.

To illustrate the result with conditional simulation, we consider the system shown in Fig. 11 [30], a highly complex nonstationary system of size 200×200. Also shown are the hard data for both the background and branching structure. The reconstructed system, along with a comparison of the computed MPC functions are also shown in Figure 11. Once again, the agreement is excellent. Clearly, honoring the HD imposes a more stringent constraint on the reconstruction process, resulting in even better agreement.

The next example is a cross section of a dicot wooden stem, magnified by a factor of 400 [31] and shown in Fig. 12. A stem is one of two major structural axis of a vascular plant. The image was digitized into a 300×400 grid, the same as that of the grid $\mathbf{G}$. Its reconstruction and a comparison of the computed MPC functions are also presented in Fig. 12. The agreement is excellent.

C. Disordered materials

One way of fabricating various structures and materials made of inorganic nanoparticles is through the self-assembly by exploiting the interfacial properties of the particles in a solution [32]. We consider an image of a self-assembly of nanosize spherical SiO_2 particles [33], shown in Fig. 13. The image was partitioned into a 300×400 grid and reconstructed using a grid $\mathbf{G}$ of the same size. The results are shown in Fig. 13, indicating good agreement.

Another example is provided by the cross section of a systetic fracture surface of a disordered material of size 100×200, shown in Fig. 14 [34]. Its reconstruction and comparison of the computed MPC functions are also shown in Fig. 14. The agreement is good, given the complexity of the system.

D. Biological tissues and organs

Imaging, computing and analyzing living organs have become a highly active research field [14,15]. Figure 15 presents a part of an image of a child's brain [35]. A grid $\mathbf{G}$ of size 354×481, the same as that of the TS, was used to reconstruct and analyze the image. The results are shown in Fig. 15. The reconstruction mimics closely the original image. The quality of the reconstructed model can be improved further and extended to full 3D images.

The final example is a cross section of human skin [36], a notoriously complex material. The agreement between the original image of size 200×400 and its reconstructed model, measured in terms of the closeness of the computed MPC function for both, is good.

E. Numerical comparison

To further test the accuracy of the reconstruction method, we also computed the effective permeabilities of the model porous media shown in Figs. 4 and 5. The effective permeabilities of the model in Fig. 4 are 0.124, 0.85, and 0.101 (arbitrary units) in the x, y, and z directions, respectively. The average effective permeabilities of the reconstructed models were 0.121, 0.88, and 0.100, respectively. Similarly, the effective permeabilities of the model porous medium in Fig. 5 were computed to be 1.125, 0.98, and 1.114 in the x, y, and z directions, while the corresponding average permeabilities of the reconstructed models turned out to be 1.135, 1.104, and 1.121. In both cases, the agreement is excellent.

VI. DISCUSSION

Two important aspects of the method that we describe in this paper deserve further discussions.

A. Reconstruction of three-dimensional systems using two-dimensional data

In two previous papers [8] we showed how the method that we proposed is capable of reconstructing 3D *stationary* materials and media based on *a single 2D slice of data.* Clearly, The same method cannot be used for nonstationary disordered media. While our work on extending a modification of the method proposed in [8] to nonstationary media is currently ongoing, we suspect that if a nonstationary medium contains n locally-stationary zones, *at most* n 2D slices of data, one from each zone, would be sufficient in order to reconstruct a 3D NSD medium. Thus, one uses the 2D slices as the hard data that must be honored and utilizes the method described in this paper to reconstruct the rest of the system. This will be described and discussed in a future paper.

B. Integrating several types of data

In this paper the TS was represented by an image. But, we point out that our reconstruction method is not the only one that uses a conceptual framework - the image of the TS - for reconstruction; others have done the same in the past. Moreover, with 3D printing approaching commercialization, reconstruction of 3D images based on 3D images obtained from printing will be a powerful tool for generating models of 3D disordered media.

But, what distinguishes the proposed method from the previous ones is that the CCF is a multiscale function and, as such, it contains information from all the length scales relevant to the medium under consideration. More importantly, in a separate paper [37] we have shown how to integrate various types of data, either in terms of images or numerical data (such as, for example, the porosity distribution or well log data), with the basic method based on the CCF that we introduced previously [8] and used in this paper to reconstruct the locally-stationary zones in a NSD medium. Briefly, suppose that one has m types of hard (quantitative or numerical) and/or "soft" (qualitative) data. One computes, according to the number of variables at hand, the corresponding CCFs and, hence, selects the final pattern of heterogeneity based on such functions. Thus, Eq. (6) is generalized to,

$$\mathcal{C}_{\text{overall}}(\mathbf{x}) = \mathcal{C}_{\mathcal{T}}(\mathbf{D}_T, \mathcal{T}) + \sum_{k=1}^{m} \omega_k \mathcal{C}_{k\mathcal{T}}(k\mathbf{D}_T, k\mathcal{T}) \ , \tag{10}$$

where $0 \leq \omega_k \leq 1$ is the weight for the kth TS or data set, $\mathcal{C}_{k\mathcal{T}}$ its corresponding CCF, and $k\mathbf{D}_T$ is the data event for the kth type of data. One may include any type data, both numerical and in the form of images. Extensive discussions and examples are described elsewhere [37].

VII. SUMMARY

This paper describes a method for reconstruction of nonstationary materials and media. The method is based on partitioning a nonstationary medium into locally-stationary zones, and reconstructing each zone based a cross-correlation function and a one-dimensional raster path that we introduced previously [8]. The segmentation of the nonstationary system into locally-stationary zones is done by two distinct methods. One is based on Shannon entropy, while the second method is based on a new technique introduced in the present paper, which is based on watershed transform. It was demonstrated that the method generates highly accurate reconstruction and realizations a wide variety of materials and media.

In our view, the advantages of the approach are, (i) low computational cost; (ii) high acuracy; (iii) applicability to any type of nonstationary system; (iv) the possibility of incorporating hard data to be honored exactly, and (v) the possibility of carrying the computations in parallelizd mode.

Past experience [1,2] has indicated that a reconstruction method that regenerates only low-order statistics of a disordered medium is not sufficient for reproducing the high-order ones and, thus, a high quality model. Our method addresses this deficiency by directly generating realizations of *multiple-scale structures* in a stochastic manner that reproduce the medium's *multiple-point* statistics through the CCF and the overlap regions.

Clearly, the method is not expected to be accurate for all types of materials and media. Our work has targetted highly heterogeneous materials and media, and the resulted presented here and in Ref. [8] demonstrate that the method is accurate for reconstructing them. Whether it can be equally accurate for other types of media remains for future investigations.

ACKNOWLEDGMENTS

The initial stages of the work was supported in part by the Department of Energy. This work has also received partial support from the RPSEA Consortium.

[†] E-mail address: moe@usc.edu

[1] S. Torquato, *Random Heterogeneous Materials* (Springer, New York, 2002).

[2] M. Sahimi, *Heterogeneous Materials I & II* (Springer, New York, 2003).

[3] N. F. Berk, Phys. Rev. Lett. **58**, 2718 (1987); P. M. Adler, C. G. Jaquin, and J. A. Quiblier, Int. J. Multiphase Flow **16**, 691 (1990); P. M. Adler, C. G. Jaquin, and J.-F. Thovert, Water Resour. Res. **28**, 1571 (1992); P. Spanne, J.-F. Thovert, C. J. Jacquin, W. B. Lindquist, K. W. Jones, and P. M. Adler, *ibid.*, **73**, 2001 (1994); A. P. Roberts and M. Teubner, Phys. Rev. E **51**, 4141 (1995).

[4] C. Manwart and R. Hilfer, Phys. Rev. E **59**, 5596 (1999); C. Manwart, S. Torquato, and R. Hilfer, *ibid.* **62**, 893 (2000); B. Biswal, P.-E. Øren, R. J. Held, S. Bakke, and R. Hilfer, *ibid.* **75**, 061303 (2007); H. Kumar, C. L. Briant, and W. A. Curtin, Mech. Mater. **38**, 818 (2006);

[5] H. Hamzehpour and M. Sahimi, Phys. Rev. E **74**, 026308 (2006); H. Hamzehpour, M. R. Rasaei, and M. Sahimi, *ibid.* **75**, 056311 (2007).

[6] C. L. Y. Yeong and S. Torquato, Phys. Rev. E **57**, 495 (1998); *ibid.* **58**, 224 (1998); Y. Jiao, F. H. Stillinger, and S. Torquato, *ibid.* **76**, 031110 (2007); *ibid.* **77**, 031135 (2008).

[7] Y. Jiao, F. H. Stillinger, and S. Torquato, Proc. Natl. Acad. Sci. USA **106**, 17634 (2009); C. E. Zachary and S. Torquato, Phys. Rev. E **84**, 056102 (2011).

[8] P. Tahmasebi and M. Sahimi, Phys. Rev. E **85**, 066709 (2012); Phys. Rev. Lett. **110**, 078002 (2013).

[9] R. A. Frazin, M. D. Butala, A. Kemball, and F. Kamalabadi, Astrophys. J. **635**, L197 (2005).

[10] A. Seppäen, M. Vauhkonen1, P. J. Vauhkonen1, E. Somersalo, and J. P. Kaipio, Inverse Problems **17**, 467 (2001); R. T. Lemos and B. Sansó, J. Am. Statist. Asso. **104**, 5 (2009).

[11] L. Bertin, G. Evensen, and H. Wackernagel, Inverse Problems **18**, 1 (2002); M. Honarkhah and J. Caers, Math. Geosci. **42**, 487 (2010).

[12] M. Sahimi, *Flow and Transport in Porous Media and Fractured Rock*, 2nd ed. (Wiley-VCH, Weinheim, 2011).

[13] J. L. Mennis and L. Jordan, Ann. Asso. Am. Geog. **95**, 249 (2005).

[14] T. McInerney and D. Terzopoulos, Comput. Med. Graph. **19**, 69 (1995).

[15] H. Müller, N. Michoux, D. Bandon, and A. Geisbuhler, Int. J. Med. Information **73**, 1 (2004); H. Handels and J. Ehrhardt, Methods Inf. Med. **48**, 11 (2009).

[16] R. Friedlich, J. Peinke, M. Sahimi, and M. R. Rahimi Tabar, Phys. Rep. **506**, 87 (2011).

[17] C. E. Shannon, Bell Syst. Technol. J. **27**, 379 (1948).

[18] S. Beucher, in *Mathematical Morphology in Image Processing*, edited by E.R. Dougherty (Marcel Dekker, New York, 1993), p. 433.

[19] L. Najman and M. Schmitt, Signal Processing **38**, 99 (2994).

[20] A. Bleau and L. J. Leon, Computer Vision Image Understanding **77**, 317 (2000).

[21] item J. Cousty, G. Bertrand, L. Najman, and M. Couprie, IEEE Trans. Pattern Anal. Machine Intelligence **31**, 1362 (2009).

[22] S. Krishnan and A. G. Journel, Math. Geol. **35**, 915 (2003).

[23] M. Sahimi and S. E. Tajer, Phys. Rev. E **71**, 046301 (2005).

[24] S. Castro, J. Caers, and T. Mukerji, *the Stanford VI Reservoir: 18th Annual Report*, Stanford Center for Reservoir Forecasting, Stanford University (2005).

[25] M. R. Rasaei and M. Sahimi, Transp. Porous Media **72**, 311 (2008).

[26] M. Honarkhah, *Stochastic Simulation of Patterns Using Distance-Based Pattern Modeling*, Ph.D. dissertation, Stanford University (2011).

[27] L. de Vries, J. Carrera, O. Falivene, O. Gratacs̃, and L. Slooten, Math. Geosci. **41**, 29 (2009).

[28] A. Harding, S. Strebelle, M. Levy, J. Thorne, D. Xie, S. Leigh, R. Preece, and R. Scamman, in *Geostatistics Banff 2004*, edited by O. Leuangthong and C. V. Deutsch, vol. 14 (Springer, Berlin, 2004), p. 559.

[29] Ganges River delta, Landsat. 7 image, USGS/EROS and NASA Landsat; see https://www.google.com/search?q=Satellite+picture+of+Ganges+River+delta&biw=1600&bih=77

[30] B. Arpat, *Sequential Simulation with Patterns*, Ph.D. Thesis, Stanford University (2005).

[31] See, http://www.onfocuslaboratories.com/support/gallery2/main.php?g2_itemId=152

[32] M. Li, H. Schnablegger, and S. Mann, Nature **402**, 393 (1999).

[33] See, http://www.fei.com/resources/image-gallery/Selfassembly-SiO2-Nanosphere.aspx

[34] T. Chugunova and L. Hu, Math. Geosci. **40**, 133 (2008).

[35] See, http://thoughtbroadcast.com/tag/mri/

[36] See, http://www.visualphotos.com/image/1x9083004/black_skin

[37] P. Tahmasebi and M. Sahimi, Transp. Porous Media (to be published).

[38] A. Boucher, Comput. Geosci. **35**, 1151 (2009).

Captions

Fig. 1 (Color online). (a) A two-dimensional grid **G** with the overlap regions **O**. Arrows indicate the direction of the raster path. (b) - (e) Four locally-stationary zones. Numbers indicate the order by which the templates are reconstructed.

Fig. 2. (a) A nonstationary fracture surface [38], and (b) its reconstruction by the method for stationary media [8].

Fig. 3 (Color online). Segmentation of fracture surface of Fig. 2 by using the watershed transform.

Fig. 4 (Color online). The TS is a 3D model of a large-scale porous medium, generated by the FBM. Shown are its reconstruction and a comparison between the computed MPC functions $p(r)$.

Fig. 5 (Color online). The TS is a synthetic complex three-dimensional porous medium in which the permeability varies spatially. The nonstationarity is due to different distributions of the orientations and thicknesses of the channels in various zones.

Fig. 6. The TS is a distribution of oriented ellipses, alligned in two distinct directions, at 45° and 135°, relative to the horizontal line.

Fig. 7. The TS is a distribution of ellipses in which some ellipses have two main spatially-varying properties, orientation and size. The orientations are from the upper to lower part, and the thicknesses are different in the upper/lower and middle parts.

Fig. 8. The TS is a fracture pattern in a cross section of rock. Shown are its reconstruction and the comparison of the MPC functions.

Fig. 9 (Color online). The TS is a shallow-water tidal-dominated system.

Fig. 10 (Color online). The TS is the delta of Ganges River.

Fig. 11. Conditional reconstruction of the delta of a river with many channels or, more generally, a disordered branched structure. We also indicate the spatial distribution of the hard data. Note that the nonstationarity of the system is due to the average and variance of the channels' thickness and orientation being different in all the direction.

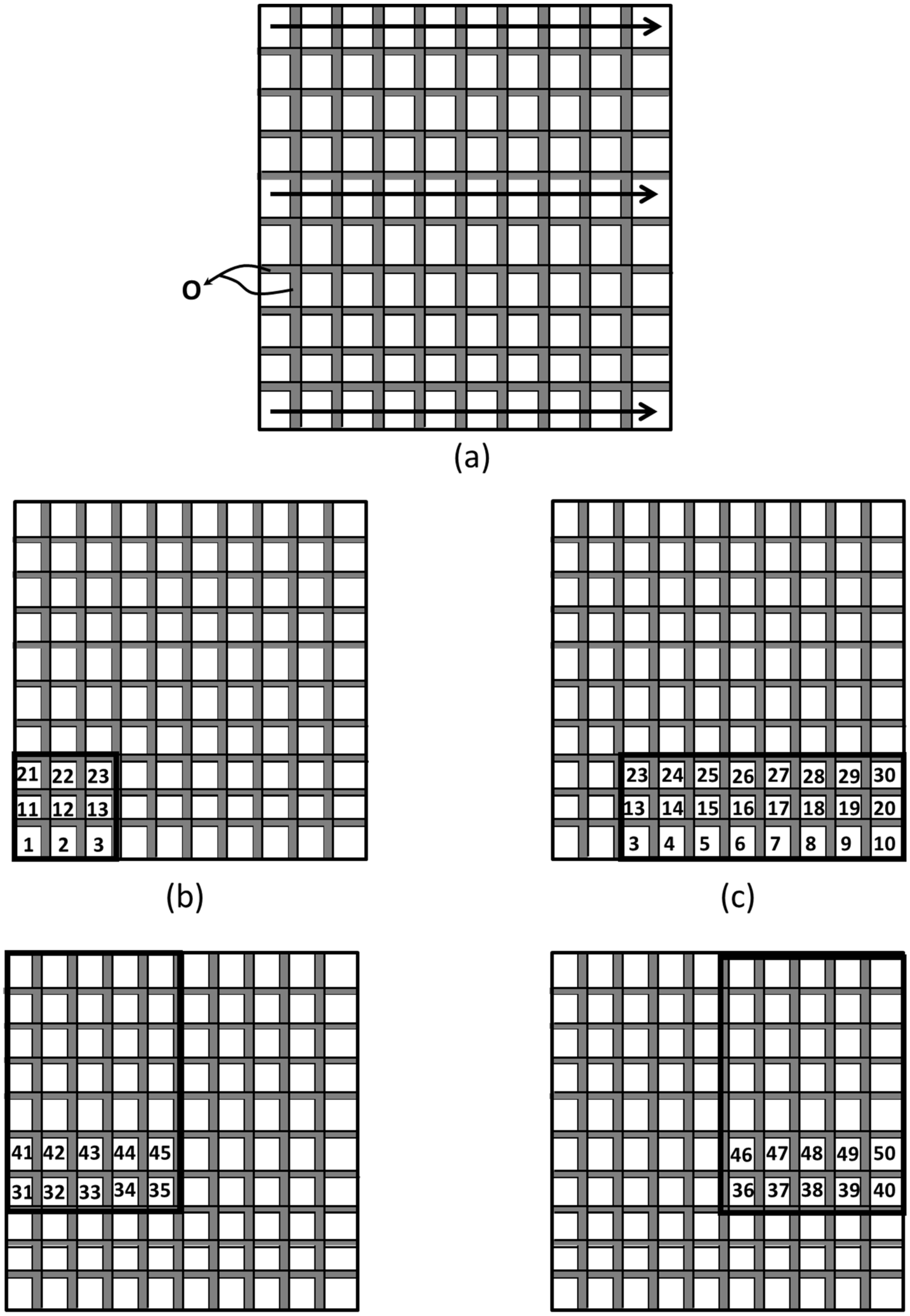
Fig. 12 (Color online). A cross section of the TS, dicot wooden stem, and its reconstruction. Shown are the computed MPC functions.

Fig. 13 (Color online). The TS is a self-assembly array of SiO_2 nanosphere. Shown are its reconstruction, along with the computed MPC functions $p(r)$.

Fig. 14. The TS is a synthetic fracture surface. The fractures have a constant thickness, but are orientated, with the orientation distribution being different in various regions.

Fig. 15. The TS is a 2D image of a child's brain. Shown are a realization of its reconstructed image and the comparison of the computed MPC functions.

Fig. 16. The TS is a cross section of human skin



(a)

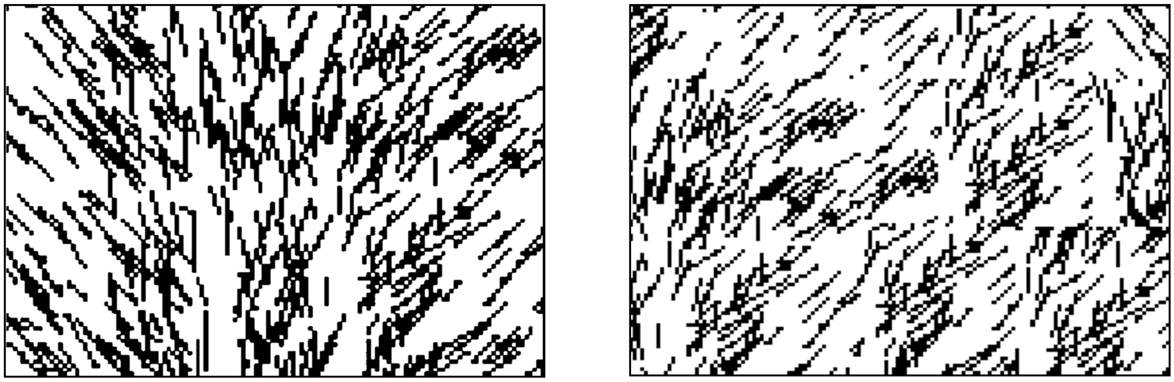
(b)

(c)

(d)

(e)

Figure 1



(a) (b)

Figure 2

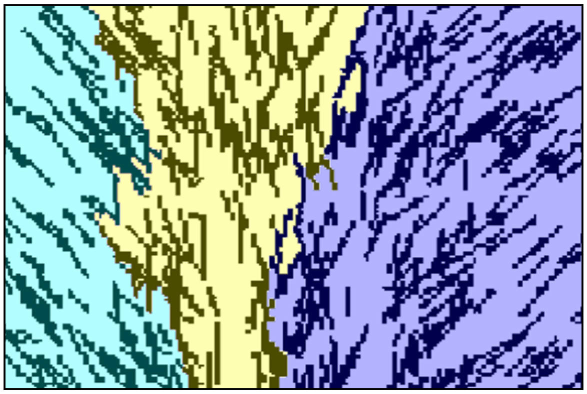


Figure 3

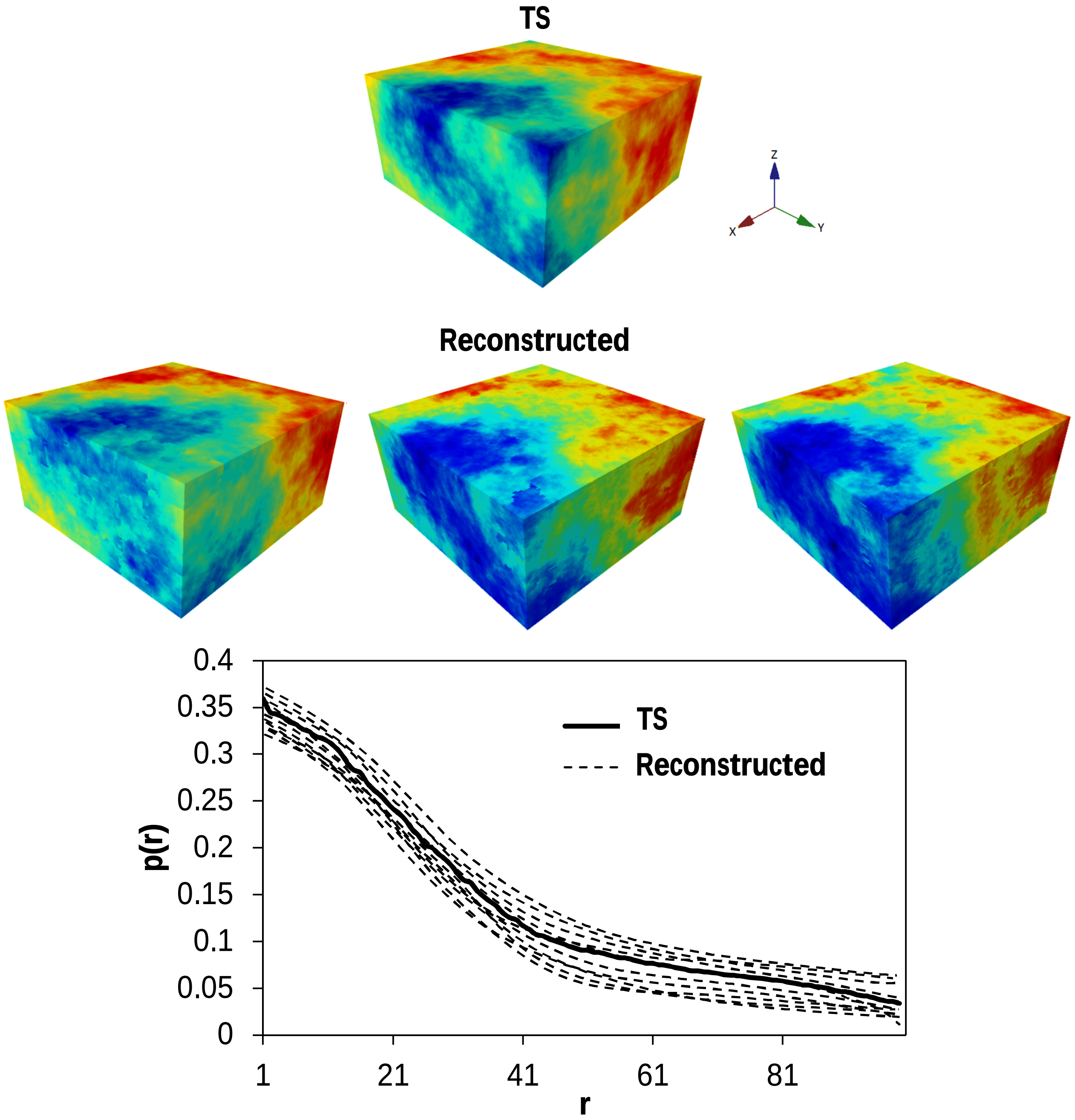

TS
Z
X
Y
Reconstructed
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
p(r)
1
21
41
61
81
r
TS
Reconstructed

Figure 4

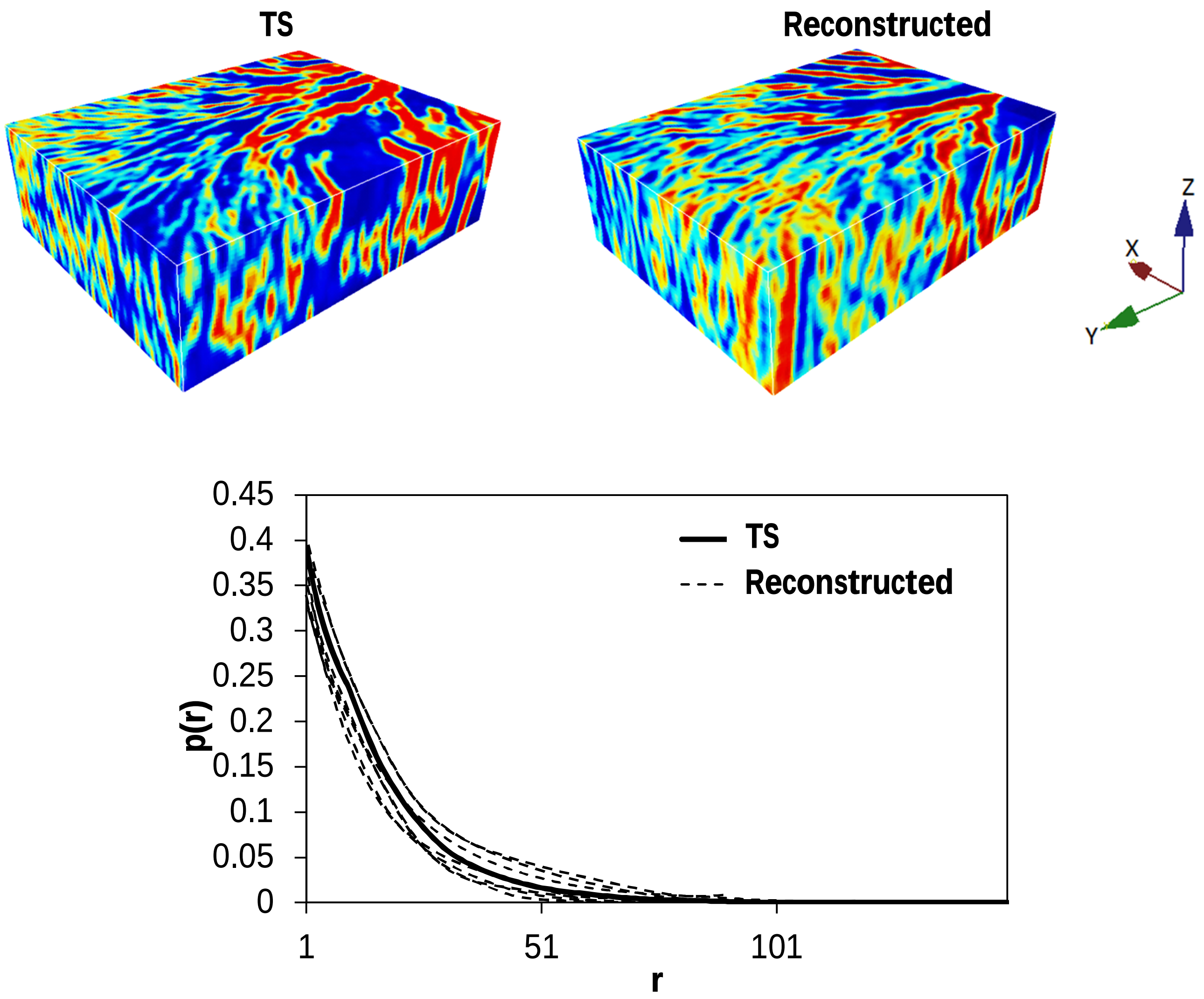

TS
Reconstructed
Z
X
Y
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
p(r)
TS
Reconstructed
1
51
101
r

Figure 5

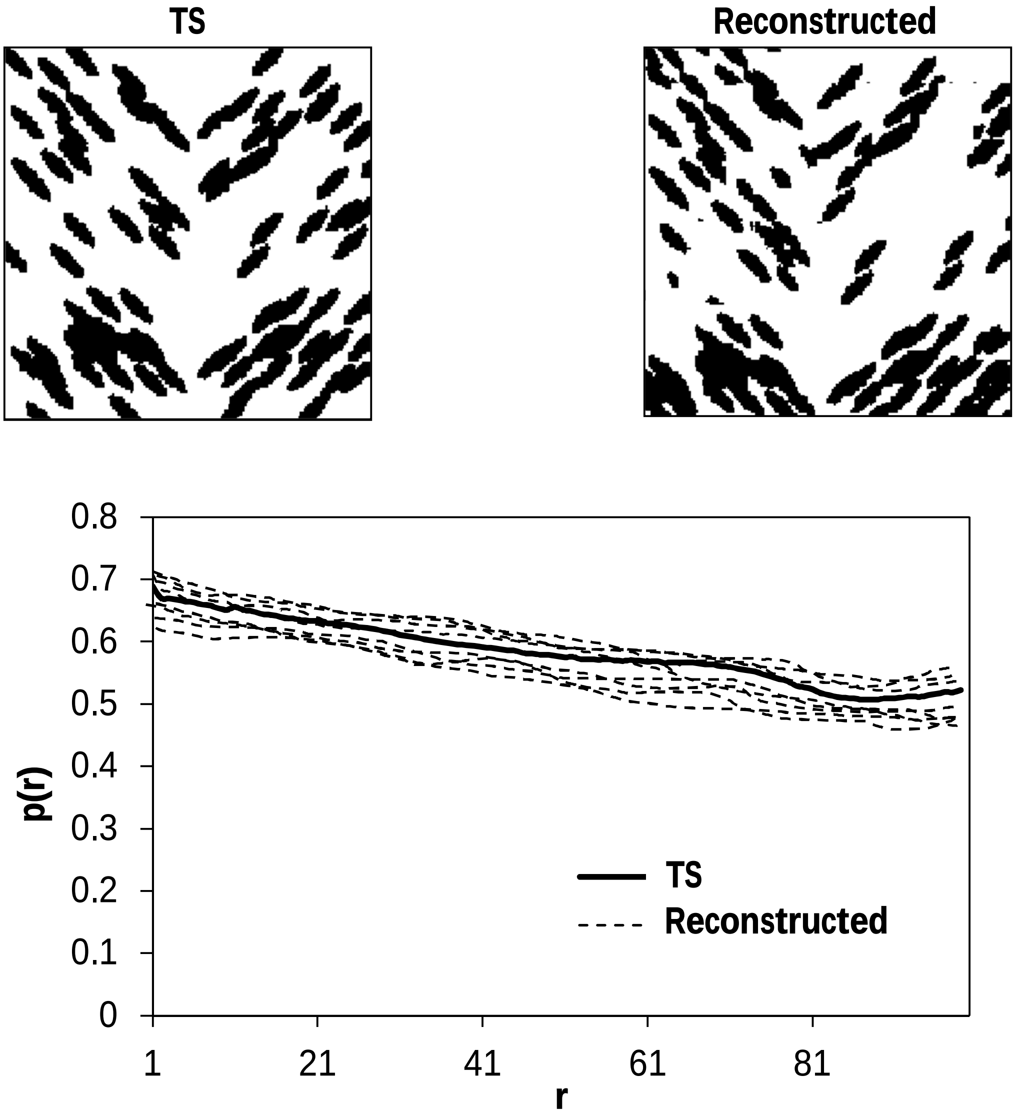

TS
Reconstructed
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
p(r)
1
21
41
61
81
r
TS
Reconstructed

Figure 6

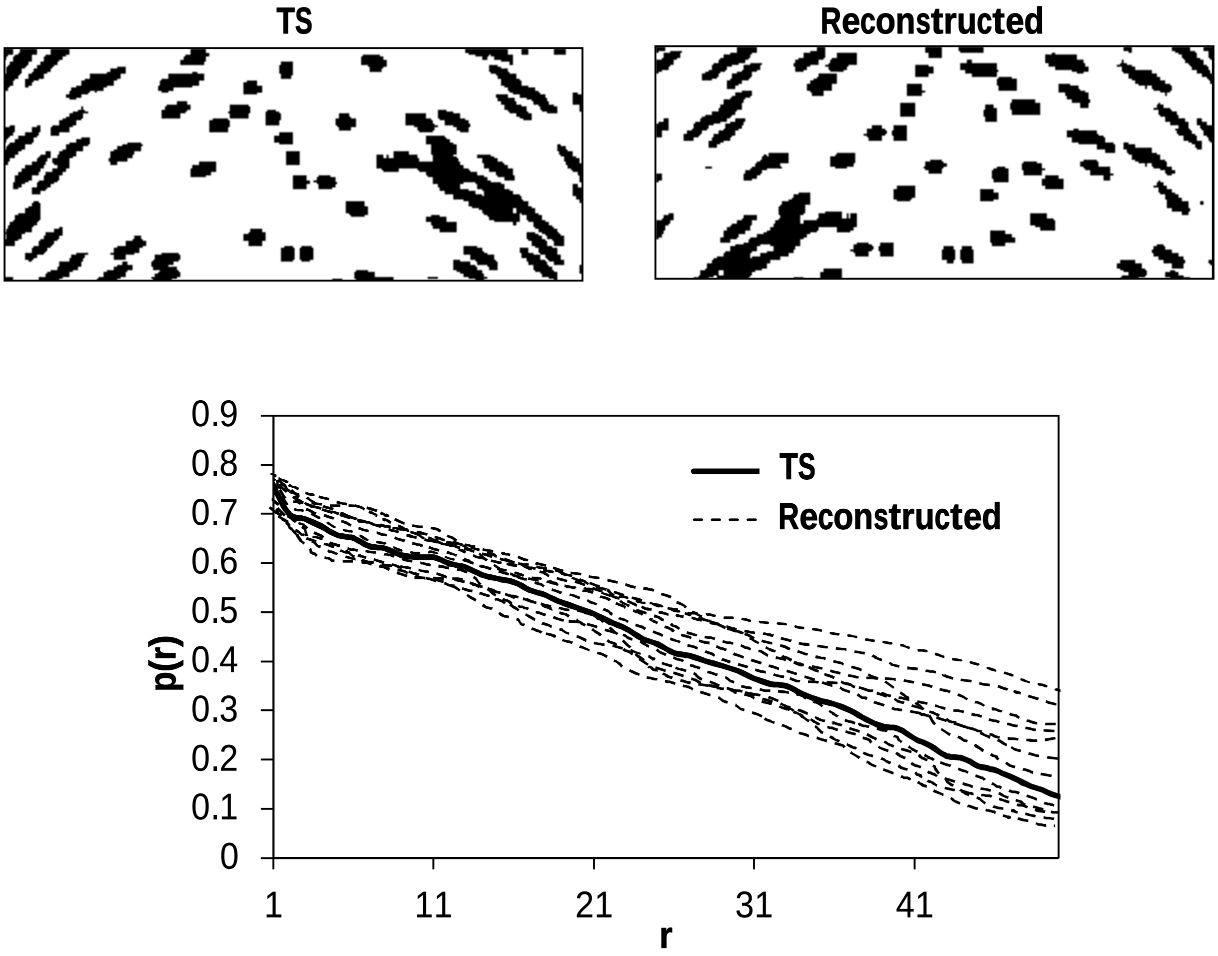

TS
Reconstructed
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
p(r)
1
11
21
31
41
r
TS
Reconstructed

Figure 7

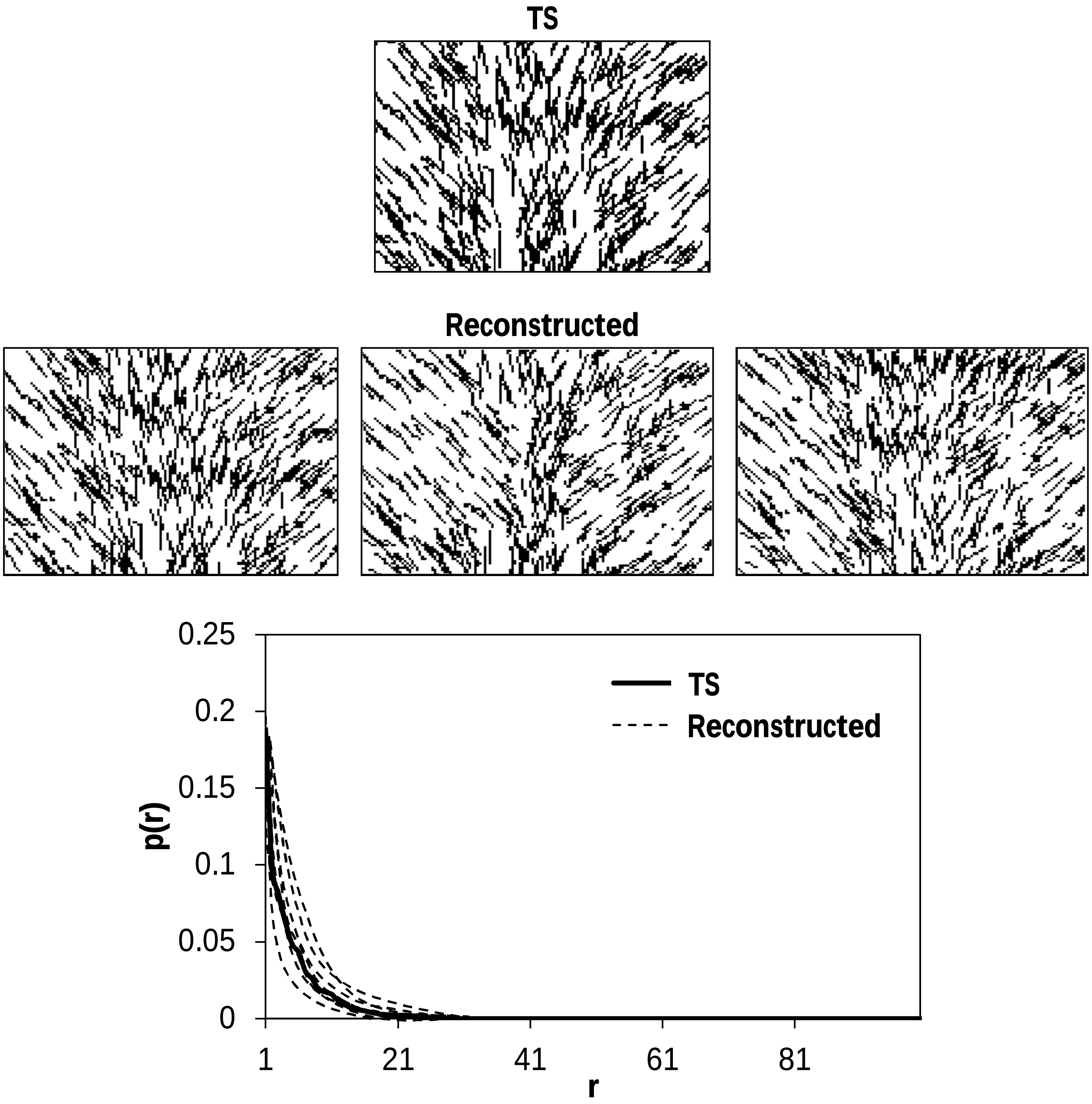

TS
Reconstructed
0.25
0.2
0.15
0.1
0.05
0
p(r)
1
21
41
61
81
r
TS
Reconstructed

Figure 8

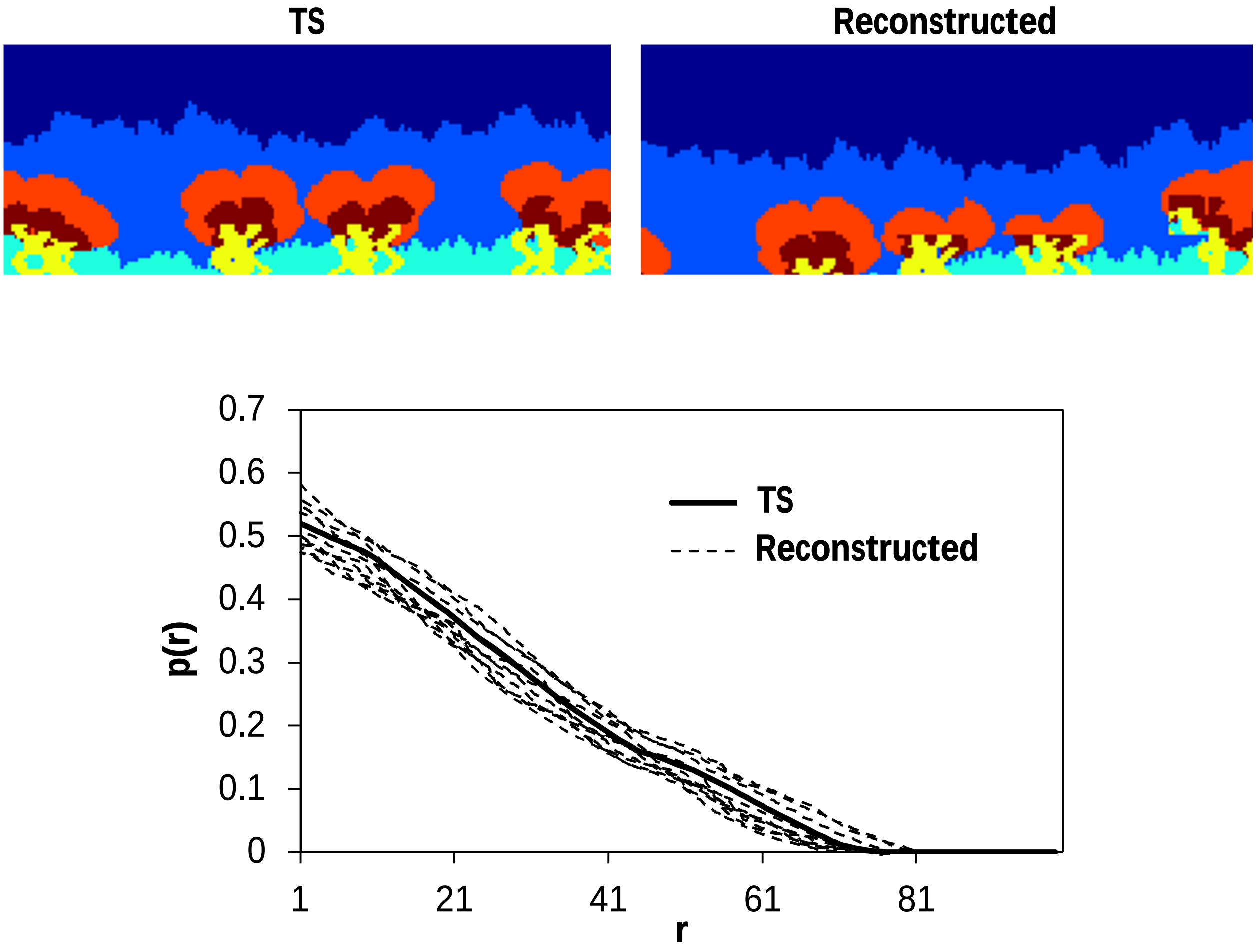
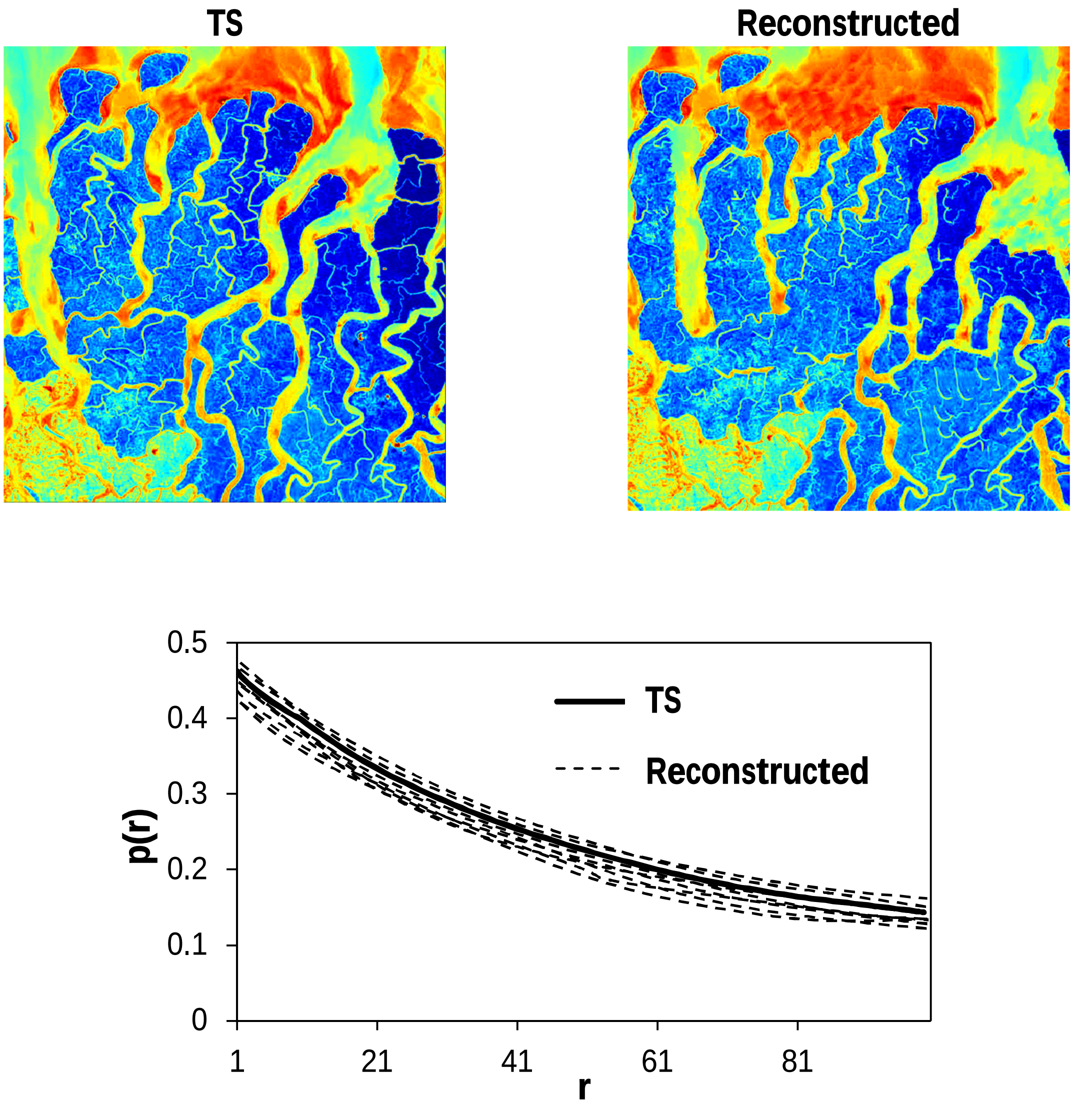
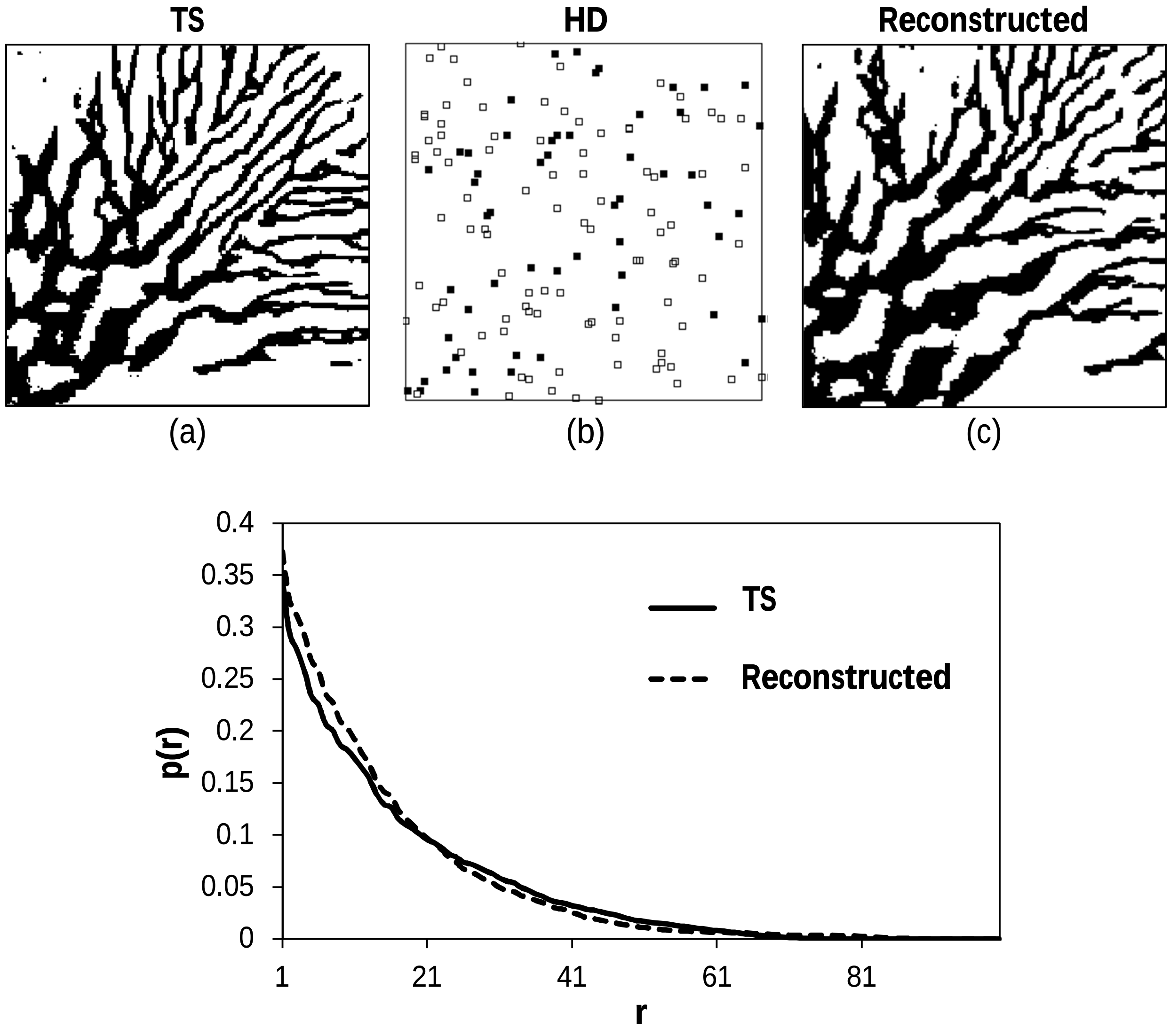

TS
Reconstructed
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
p(r)
1
21
41
61
81
r
TS
Reconstructed

Figure 9



MPC

Figure 10



(a) (b) (c)

(d)

Figure 11

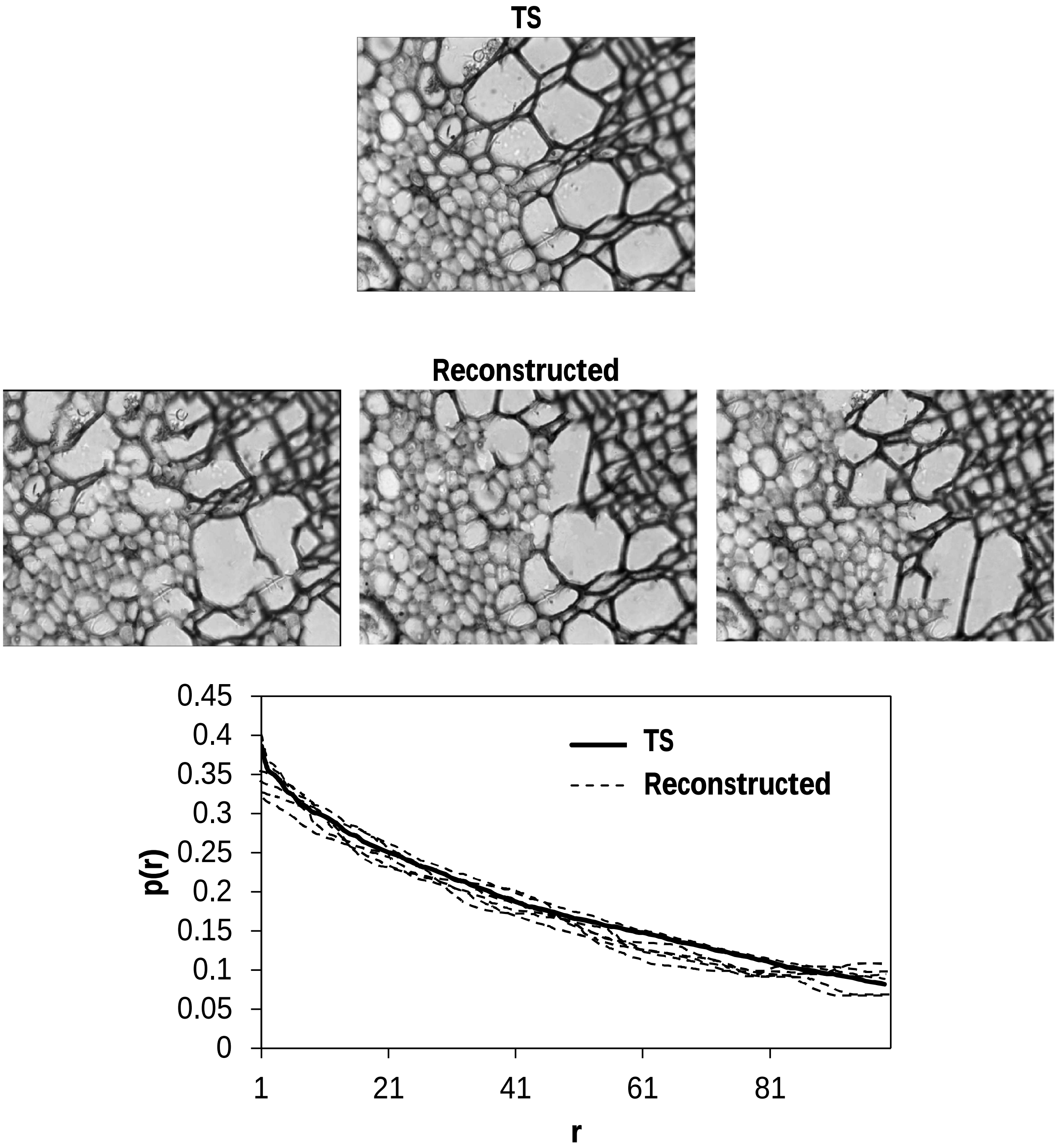

TS
Reconstructed
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
p(r)
1
21
41
61
81
r
TS
Reconstructed

Figure 12

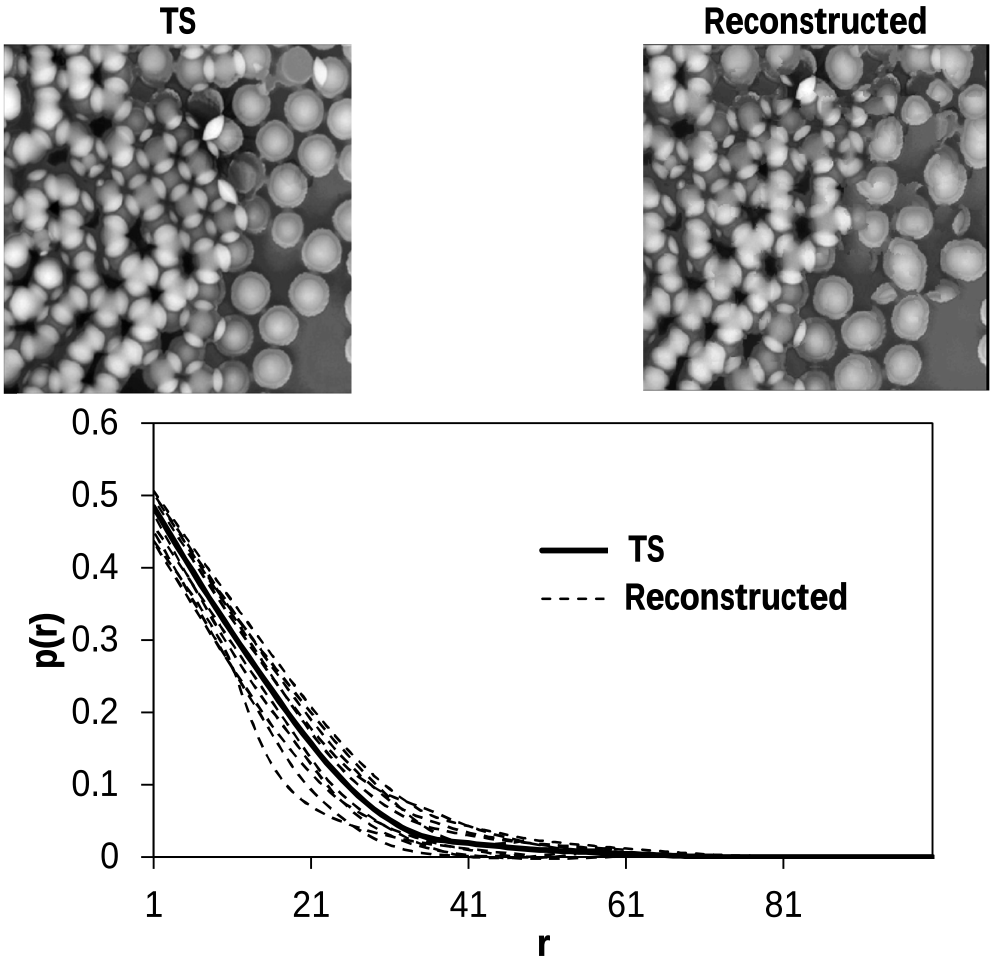

TS
Reconstructed
0.6
0.5
0.4
0.3
0.2
0.1
0
p(r)
1
21
41
61
81
r
TS
Reconstructed

Figure 13

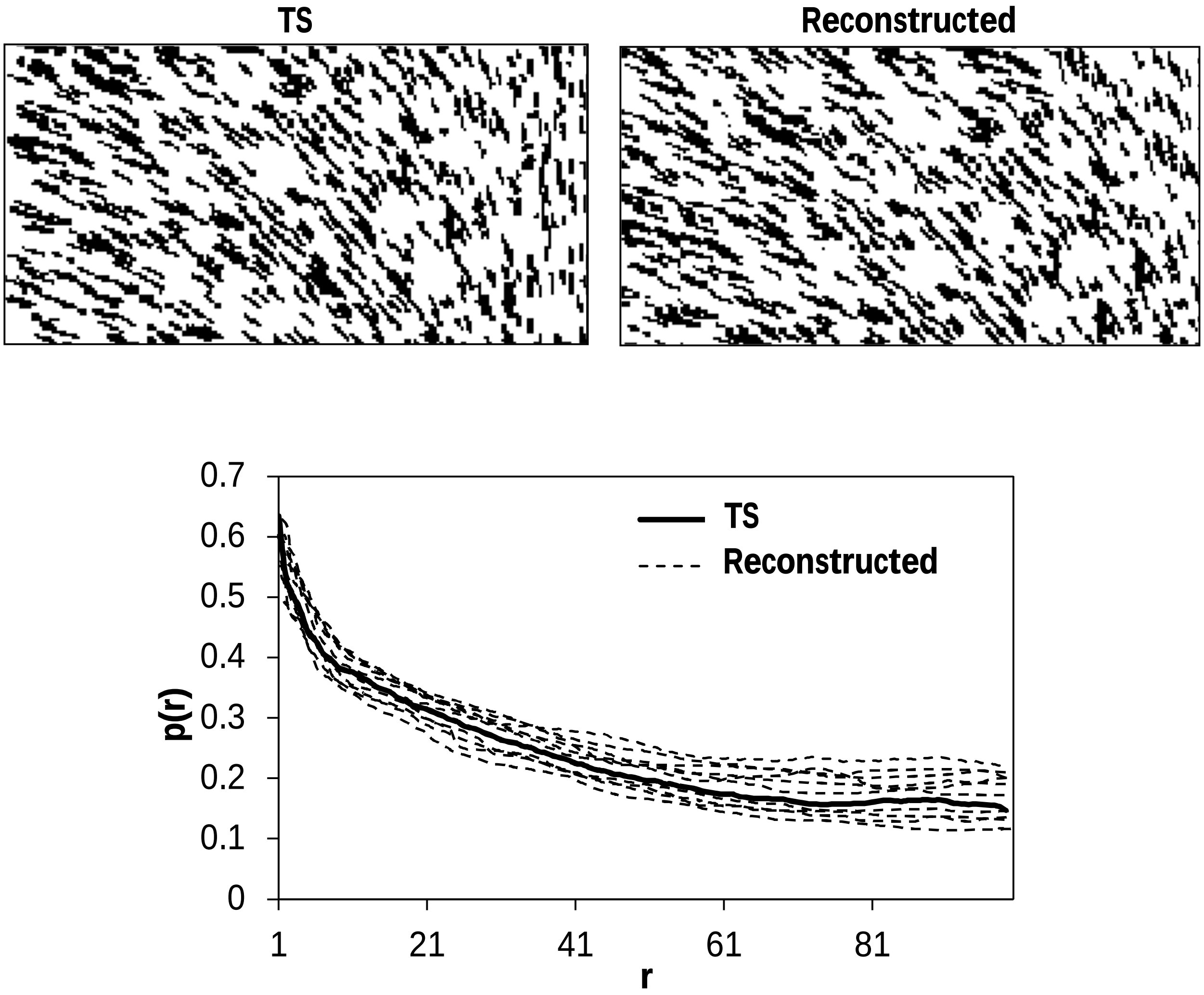

TS
Reconstructed
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
p(r)
1
21
41
61
81
r
TS
Reconstructed

Figure 14

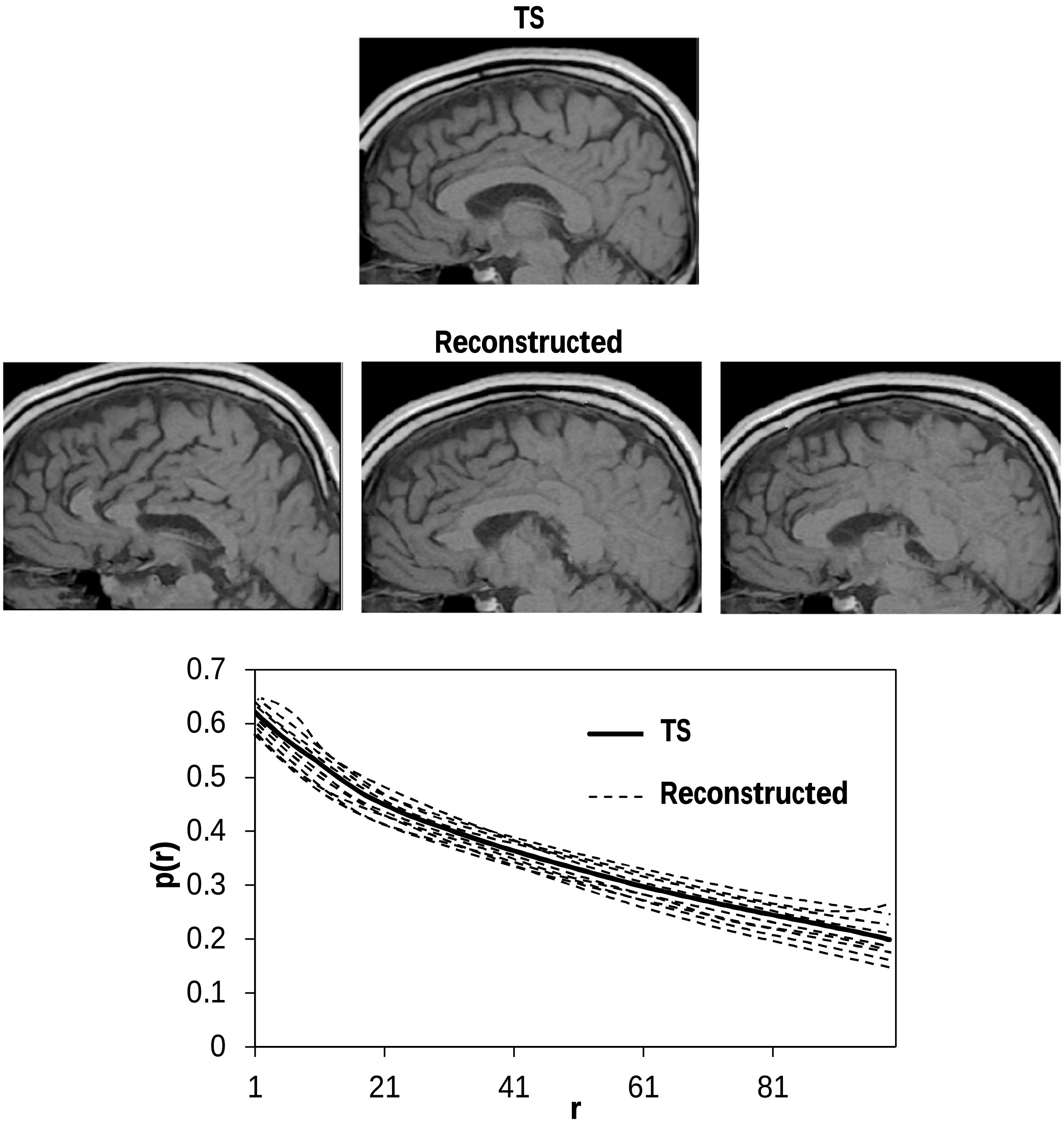

TS
Reconstructed
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
p(r)
1
21
41
61
81
r
TS
Reconstructed

Figure 15

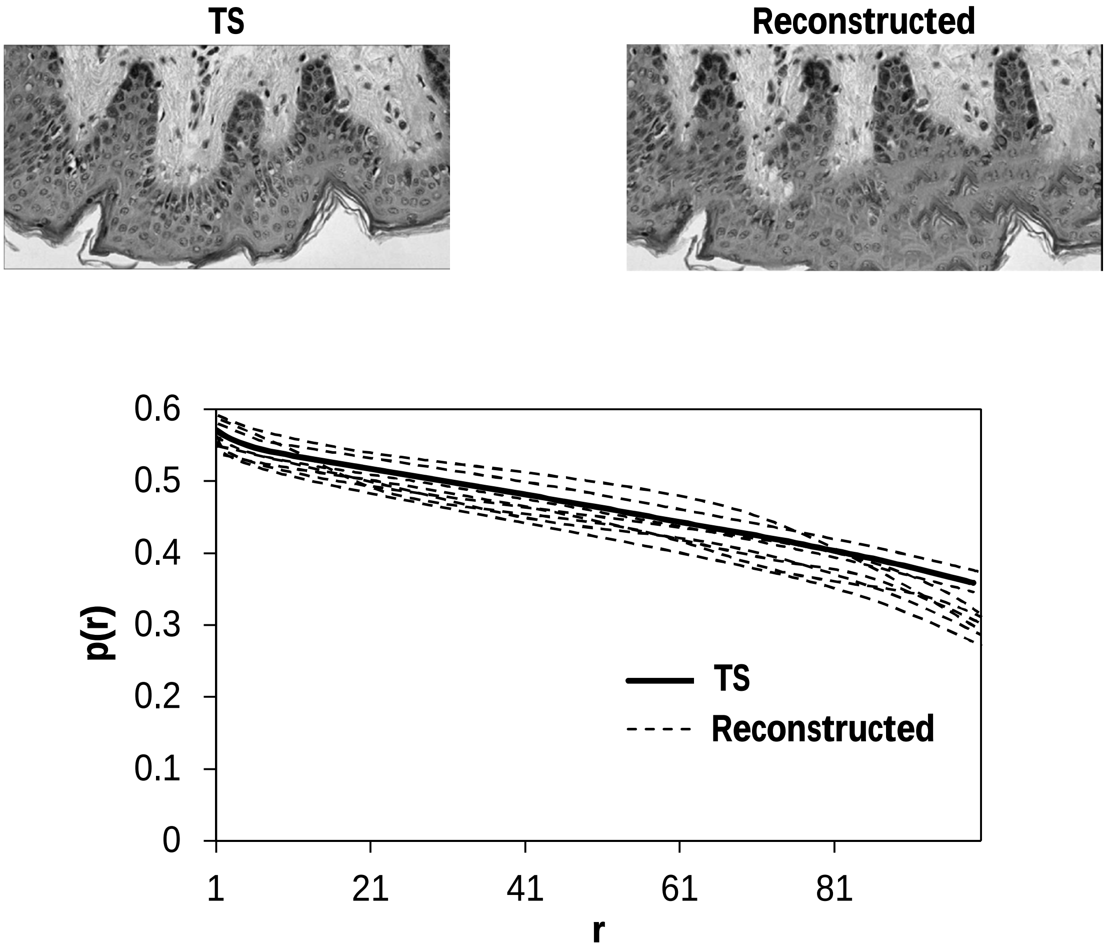

TS
Reconstructed
0.6
0.5
0.4
0.3
0.2
0.1
0
p(r)
1
21
41
61
81
r
TS
Reconstructed

Figure 16