Research Article

Characterization of Aquifer Multiscale Properties by Generating Random Fractal Field with Truncated Power Variogram Model Using Karhunen–Loève Expansion

Liang Xue,^{1,2} Diao Li,¹ Cheng Dai,³ and Tongchao Nan⁴

¹Department of Oil-Gas Field Development, College of Petroleum Engineering, China University of Petroleum, Beijing 102249, China

²State Key Laboratory of Petroleum Resources and Engineering, China University of Petroleum, Beijing 102249, China

³State Key Laboratory of Shale Oil and Gas Enrichment Mechanisms and Effective Development, Sinopec Group, Beijing 100083, China

⁴Department of Hydrosocieties, School of Earth Sciences and Engineering, Nanjing University, Nanjing 210093, China

Correspondence should be addressed to Tongchao Nan; tcnan@nju.edu.cn

Received 29 June 2017; Revised 31 October 2017; Accepted 28 November 2017; Published 19 December 2017

Academic Editor: Walter A. Illman

Copyright © 2017 Liang Xue et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The traditional geostatistics to describe the spatial variation of hydrogeological properties is based on the assumption of stationarity or statistical homogeneity. However, growing evidences show and it has been widely recognized that the spatial distribution of many hydrogeological properties can be characterized as random fractals with multiscale feature, and spatial variation can be described by power variogram model. It is difficult to generate a multiscale random fractal field by directly using nonstationary power variogram model due to the lack of explicit covariance function. Here we adopt the stationary truncated power variogram model to avoid this difficulty and generate the multiscale random fractal field using Karhunen–Loève (KL) expansion. The results show that either the unconditional or conditional (on measurements) multiscale random fractal field can be generated by using truncated power variogram model and KL expansion when the upper limit of the integral scale is sufficiently large, and the main structure of the spatial variation can be described by using only the first few dominant KL expansion terms associated with large eigenvalues. The latter provides a foundation to perform dimensionality reduction and saves computational effort when analyzing the stochastic flow and transport problems.

1. Introduction

Groundwater resources management and contamination prevention are the important issues to sustain the development of human society. One of the most challenging tasks to address these issues is to describe the heterogeneity of the hydrogeological parameters, such as hydraulic conductivity and porosity over scales. Many techniques have been devoted to obtain more information on the heterogeneity of the hydrogeological parameters in different scales, such as core samples analysis [1], direct push technology [2], and hydraulic tomography [3]. Due to the complex nature of the spatial distribution of these hydrogeological parameters, these parameters tend to be treated as random variables, and thus their spatial distribution can be quantitatively described using stochastic process [4, 5]. The governing equations

of groundwater flow and contaminant transport become stochastic partial differential equations. In this manner, the state variables, such as hydraulic head, velocity, and contaminant concentration, are predicted with uncertainty. The merit of the stochastic method is that it can produce the probabilistic distribution of the prediction. The best prediction (quantified by the first moment, i.e., mean) and the associated uncertainty (quantified by the second moment, variance/covariance) can be inferred from this distribution. Geostatistics was introduced to hydrogeology and became popular in numerical simulation of groundwater flow and solute transport [6–8]. The assumption underlying the traditional geostatistics [9] is that the spatial distribution of the hydrogeological property is statistically homogeneous; that is, the joint probability of hydrogeological parameters depends only on their relative locations, but not on the exact locations.

TABLE 1: Comparison of the existing methods to generate random fractal field.

Tool	Variogram type	Conditioning	Dimensionality reduction	Reference
GSLIB	Stationary	Yes	No	Deutsch and Journel, 1997
GSLIB-TPV	Stationary and nonstationary	Yes	No	Xue and Zhang, 2014; Xue et al., 2014
Spectral methods	Stationary and nonstationary	No	No	Heße et al., 2014
Fractal method	Stationary and nonstationary	No	No	Dieker, 2004
KL-TPV	Stationary and nonstationary	Yes	Yes	Current method

However, it is evident that the random hydrogeological field may manifest it as random fractals with statistically homogeneous increments [10–14]. Stochastic fractal models to characterize such random field include Gaussian-based fractional Brownian motion and fractional Gaussian noise [10, 12, 15], non-Gaussian-based fractional Levy motion or fractional Levy noise [13], and multifractals [16]. The Gaussian-based fractional Brownian motion model that can be characterized by power law variogram can be inferred from collected hydrogeological data, such as permeability or porosity, at several sites over diversified distance scales [12, 17]. The random field described using such model exhibits the self-affine scaling properties and thus can take the multiscale variation into account.

To analyze the scale-dependency of the random fractal field and investigate the effect of multiscale random fractals to the behavior of groundwater flow and contaminant transport, it requires generating the realizations of hydrogeological properties based on their stochastic characteristics. If measurements on the hydrogeological properties are directly available, the generated multiscale random fractal field should be further conditional on the measurement values. The most common method used to generate a Gaussian random field is sequential Gaussian simulation (SGSIM) due to the wide usage of geostatistical software library (GSLIB) developed by Deutsch and Journel [9]. Although power variogram model is included in GSLIB, it can not be used directly to generate a multiscale random fractals field. The traditional geostatistical method can only generate Gaussian random field with the assumption of stationarity or statistical homogeneity. Since the power variogram does not have a finite sill and integral scale and lacks an explicit form of covariance function, the SGSIM module in GSLIB does not have the support to use power variogram. The spectral method can be used to generate a multiscale random fractals field, which represents the random field in Fourier space and thus can be referred to as Fourier method [18]. The Fourier method has been improved to be a hierarchical method, such as the hybrid method [19] and Fourier-Wavelet method [20].

In this work, we adopt a truncated power variogram model proposed by Di Federico and Neuman [21] to generate the multiscale random fractal field that can be characterized by power variogram model. The truncated power variogram can describe the multiscale characteristics of the random field and can take measurement scale, observation scale, and

window scale into account. This model also has a reasonable hydrogeological interpretation, which states that the power law variogram behavior of log hydraulic conductivity random field can be considered as a weighted integral of infinite hierarchies of mutually uncorrelated stationary fields. After truncating the higher and/or lower frequency, the truncated power variogram models become stationary and have explicit forms of covariance functions. Each hierarchy is referred to as a mode in the spectra of such random field. Mathematically, it has been proved that each mode can be characterized by either the exponential or the Gaussian variogram. It renders two types of truncated power variogram: truncated power variogram with exponential modes (TpVE) and truncated power variogram with Gaussian modes (TpVG). We have modified the GSLIB code to incorporate these truncated power variogram models to generate random fractal field, both unconditional and conditional, using sequential Gaussian simulation method [22]. Other methods can also generate random fractal field, such as the spectral methods [23] and fractal method [24]. However, these existing methods do not have the capability to generate the conditional random field. The conditioning data of hydrogeological properties are usually available and they are critical components to improve the characterization of the targeted field. Therefore, it is necessary to take these hard data into account by the conditioning technique. The comparison of the existing methods to generate random fractal fields is listed in Table 1. Here in this work, an innovative method to generate the multiscale random fractal field is proposed by incorporating the truncated power variogram models with Karhunen–Loève (KL) expansion method. The advantage of the proposed method is that KL expansion has the capability to conduct the model dimensionality reduction. If it is further incorporated with polynomial chaos expansion on the state variables, such as hydraulic head in the single-phase flow problem [25], saturation in the multiphase problem [26], and contaminant transport problem [27], the computational cost to perform uncertainty analysis in the hydrological research can be greatly reduced. The proposed KL-based multiscale random fractal field generator provides the foundation to establish a high-efficiency stochastic analysis framework.

The rest of the paper is arranged as follows: the theoretical bases of truncated power variogram model and KL expansion are introduced in Section 2; the results of unconditional and conditional multiscale random fractal field generations and

the capability of KL expansion to perform dimensionality reduction are presented in Section 3; and the conclusions are drawn in Section 4.

2. Methodology

2.1. Truncated Power Variogram Models. The heterogeneous log hydraulic conductivity, denoted as Y , random fractal field with statistical homogeneous increments, can be described using power variogram model:

$$\gamma(s) = A_0 s^{2H}, \quad (1)$$

where A_0 is a constant, H is Hurst coefficient ($0 < H < 1$), and s is the lag distance.

The log hydraulic conductivity random field is self-affine since the power variogram function has the scaling property:

$$\gamma(rs) = r^{2H} \gamma(s), \quad (2)$$

where r is any positive constant ($r > 0$).

The fractal dimension of the self-affine log hydraulic conductivity random field is

$$d = E + 1 - H, \quad (3)$$

where E is the Euclidean or topologic dimension.

Define the mode number $n = 1/I$ representing the spatial frequency of random fluctuation (where I is the integral scale), integrate continuously over all the possible modes within the range $[n_l, n_u]$, and weight each mode with a factor of $1/n$ [21],

$$\gamma(s, n_l, n_u) = \int_{n_l}^{n_u} \gamma(s, n) \frac{1}{n} dn, \quad (4)$$

where the lower and upper limits of mode number n_l and n_u correspond with the upper and lower limits of integral scale I_u and I_l due to the reciprocal relationship. The upper limit of integral scale I_u is proportional to the characterization length of window scale (e.g., the domain size of an aquifer), and the lower limit of integral scale I_l is proportional to the characterization length of data support or measurement scale (e.g., the size of the soil core). It can be noted that the truncated variogram $\gamma(s, n_l, n_u)$ is a linear combination of the variogram $\gamma(s, n)$ in each mode weighted by $1/n$ (where n is always positive, $n > 0$); therefore the property of truncated variogram $\gamma(s, n_l, n_u)$ holds that of $\gamma(s, n)$; that is, $\gamma(s, n_l, n_u)$ is conditionally negative definite function if variogram function $\gamma(s, n)$ (that can be in the form of exponential and Gaussian variogram as will be shown soon) is conditionally negative definite function. It also indicates that the corresponding truncated covariance function $C(s, n_l, n_u)$ is positive definite if covariance function $C(s, n)$ (here can be exponential and Gaussian covariance) is positive defined, since the truncated covariance function is stationary.

If each statistically homogeneous mode in an infinite hierarchy log hydraulic conductivity random field with uncorrelated spatial increments can be characterized by exponential variograms in terms of I integral scale,

$$\gamma(s, I) = \sigma^2(I) \left[1 - \exp\left(-\frac{s}{I}\right) \right], \quad (5)$$

where $\sigma^2(I) = AI^{2H}$ is the variance of log hydraulic conductivity and A is a constant.

The consequent truncated power variogram model with exponential modes (TpVE) is

$$\gamma(s, I_l, I_u) = \gamma(s, I_u) - \gamma(s, I_l) \quad (6)$$

with the analytical form ($0 < H \leq 0.5$)

$$\begin{aligned} \gamma(s, I_m) &= \frac{AI_m^{2H}}{2H} \left[1 - \exp\left(-\frac{s}{I_m}\right) \right. \\ &\quad \left. + \left(\frac{s}{I_m}\right)^{2H} \Gamma\left(1 - 2H, \frac{s}{I_m}\right) \right], \end{aligned} \quad (7)$$

where $m = l, u$ and $\Gamma(\cdot, \cdot)$ is the incomplete gamma function.

The corresponding covariance function of TpVE model is

$$C(s, I_l, I_u) = C(s, I_u) - C(s, I_l) \quad (8)$$

with the analytical form

$$\begin{aligned} C(s, I_m) &= \frac{AI_m^{2H}}{2H} \left[\exp\left(-\frac{s}{I_m}\right) + \left(\frac{s}{I_m}\right)^{2H} \Gamma\left(1 - 2H, \frac{s}{I_m}\right) \right]. \end{aligned} \quad (9)$$

If each statistically homogeneous mode in an infinite hierarchy random field with uncorrelated spatial increments can be characterized by Gaussian variogram,

$$\gamma(s, I) = \sigma^2(I) \left[1 - \exp\left(-\frac{\pi s^2}{4I^2}\right) \right]. \quad (10)$$

The analytical form of truncated power variogram model with Gaussian modes (TpVG) is ($0 < H \leq 1$):

$$\begin{aligned} \gamma(s, I_m) &= \frac{AI_m^{2H}}{2H} \left[1 - \exp\left(-\frac{\pi s^2}{4I_m^2}\right) \right. \\ &\quad \left. + \left(\frac{\pi s^2}{4I_m^2}\right)^H \Gamma\left(1 - H, \frac{\pi s^2}{4I_m^2}\right) \right]. \end{aligned} \quad (11)$$

The corresponding analytical form of TpVG covariance is

$$\begin{aligned} C(s, I_m) &= \frac{AI_m^{2H}}{2H} \left[\exp\left(-\frac{\pi s^2}{4I_m^2}\right) \right. \\ &\quad \left. + \left(\frac{\pi s^2}{4I_m^2}\right)^H \Gamma\left(1 - H, \frac{\pi s^2}{4I_m^2}\right) \right]. \end{aligned} \quad (12)$$

If the lower limit λ_l approaches 0 (i.e., the point measurement scale) when the size of the log hydraulic conductivity measurement is so small that it can be treated as a point and the upper limit λ_u approaches infinity (i.e., infinitely large window scale or domain size), the truncated variogram models can also reduce the power variogram model (1) with the condition

$$A_0 = \frac{A\Gamma(1 - 2H)}{2H} \quad \text{for TpVE model}, \quad (13)$$

$$A_0 = \frac{A(\pi/4)^H \Gamma(1 - H)}{2H} \quad \text{for TpVG model}. \quad (14)$$

2.2. KL Expansion of Stochastic Process. The spatial distribution of hydrogeological variables, such as log hydraulic conductivity Y , can be treated as a stochastic process. The covariance function $C(\mathbf{x}, \mathbf{y})$, as in (9) for TpvE and in (12) for TpvG, of a zero-mean stochastic process, $Y(\mathbf{x})$, is bounded, symmetric, and positive definite, and it can be expanded as

$$C_Y(\mathbf{x}, \mathbf{y}) = \sum_{n=1}^{\infty} \lambda_n f_n(\mathbf{x}) f_n(\mathbf{y}), \quad (15)$$

where $\mathbf{x}, \mathbf{y}$ is the spatial coordinate in a topological space $\mathbf{D}$, the lag distance $s = |\mathbf{x} - \mathbf{y}|$, λ_n is the eigenvalue, and $f_n(\mathbf{x})$ is eigenfunction.

The eigenfunctions are orthogonal and form a complete set. They satisfy the condition

$$\int_D f_n(\mathbf{x}) f_m(\mathbf{x}) d\mathbf{x} = \delta_{nm}, \quad (16)$$

where δ_{nm} is the Kronecker delta function.

The KL expansion of the stochastic process can be expressed as

$$Y(\mathbf{x}) = \sum_{n=1}^{\infty} \xi_n \sqrt{\lambda_n} f_n(\mathbf{x}), \quad (17)$$

where ξ_n is the standard Gaussian random variable following the normal distribution $N(0, 1)$. The unconditional random field can be generated by using this equation by finding its eigenvalues and eigenfunctions with the stochastic characteristics known.

The eigenvalues and eigenfunctions of the covariance function can be solved from the Fredholm equation

$$\int_D C_Y(\mathbf{x}, \mathbf{y}) f(\mathbf{x}) d\mathbf{x} = \lambda f(\mathbf{y}). \quad (18)$$

If there are n_Y measurements $Y_1(\mathbf{x}_1), Y_2(\mathbf{x}_2), \dots, Y_{n_Y}(\mathbf{x}_{n_Y})$, it may require generating random field conditioning on these measurements. The generation of the conditional random field can be achieved by the conditional Kriging. The conditional Kriging mean and variance are, respectively,

$$E(Y(\mathbf{x}))^{(c)} = E(Y(\mathbf{x})) + \sum_{i=1}^{n_Y} \mu_i(\mathbf{x}) [Y(\mathbf{x}_i) - E(Y(\mathbf{x}))], \quad (19)$$

$$C_Y^{(c)}(\mathbf{x}, \mathbf{y}) = C_Y(\mathbf{x}, \mathbf{y}) - \sum_{i=1}^{n_Y} \mu_i(\mathbf{x}) C_Y(\mathbf{x}_i, \mathbf{y}),$$

where $\mu_i(\mathbf{x})$ are weighting functions and the subscript c represents "conditional."

The weighting function can be computed by using the following Kriging equations:

$$\sum_{i=1}^{n_Y} \mu_i(\mathbf{x}) C_Y(\mathbf{x}_i, \mathbf{y}) = C_Y(\mathbf{x}, \mathbf{y}), \quad j = 1, 2, \dots, n_Y. \quad (20)$$

Since the set of eigenfunctions $\{f_n(\mathbf{x})\}$ is complete, we can expand $\mu_i(\mathbf{x})$ on the basis of $\{f_n(\mathbf{x})\}$,

$$\mu_i(\mathbf{x}) = \sum_{k=1}^{\infty} \mu_{ik} f_k(\mathbf{x}), \quad (21)$$

where μ_{ik} is the coefficients to be determined.

Substitute this expansion to Kriging equation (20), multiply $f_m(\mathbf{x})$, and integrate both sides with respect to $\mathbf{x}$ over the domain D :

$$\sum_{i=1}^{n_Y} \mu_{im} C_Y(\mathbf{x}_i, \mathbf{y}) = \lambda_m f_m(\mathbf{y}). \quad (22)$$

Similar to unconditional case, the conditional eigenvalues $\lambda_n^{(c)}$ and eigenfunctions $f_n^{(c)}(\mathbf{x})$ of the conditional covariance functions $C_Y^{(c)}(\mathbf{x}, \mathbf{y})$ can be solved from the following Fredholm equation:

$$\int_D C_Y^{(c)}(\mathbf{x}, \mathbf{y}) f^{(c)}(\mathbf{x}) d\mathbf{x} = \lambda^{(c)} f^{(c)}(\mathbf{y}). \quad (23)$$

To determine $f_n^{(c)}(\mathbf{x})$, we expand it in terms of unconditional eigenfunctions $f_n(\mathbf{x})$ in a similar way to Kriging coefficients. Write the conditional eigenfunction as

$$f^{(c)}(\mathbf{x}) = \sum_{m=1}^M d_m f_m(\mathbf{x}). \quad (24)$$

Substituting this expression into (23), multiply $f_m(\mathbf{y})$ and integrate it with respect to $\mathbf{y}$ over D ,

$$\begin{aligned} \sum_{k=1}^M (\lambda_m \delta_{km}) d_k - \sum_{k=1}^M \left(\sum_{i=1}^{n_Y} \mu_{ik} \lambda_m f_m(\mathbf{x}_i) \right) d_k \\ = \sum_{k=1}^M (\lambda^{(c)} \delta_{km}) d_k. \end{aligned} \quad (25)$$

Write it in the matrix form

$$(\mathbf{A} - \lambda^{(c)} \mathbf{E}) \mathbf{d} = 0, \quad (26)$$

where the components of $\mathbf{A} = (a_{km})_{M \times M}$, $a_{km} = \lambda_m \delta_{km} - \sum_{i=1}^{n_Y} \mu_{ik} \lambda_m f_m(\mathbf{x}_i)$, and $\mathbf{E}$ is $M \times M$ identity matrix. This equation indicates that the problem of finding the eigenvalues of conditional covariance function can be transformed into the problem of finding the eigenvalues of matrix $\mathbf{A}$.

After finding the eigenvalue of matrix $\mathbf{A}$, $\lambda^{(c)}$, the corresponding eigenvector $\mathbf{d}$ can be used to construct the conditional eigenfunction of the conditional covariance

$$f_n^{(c)}(\mathbf{x}) = \sum_{i=1}^M d_{mi} f_i(\mathbf{x}). \quad (27)$$

Once the eigenvalue $\lambda^{(c)}$ and eigenfunction $f_n^{(c)}(\mathbf{x})$ of conditional covariance function are obtained, the conditional

log hydraulic conductivity random field can be generated using

$$Y(\mathbf{x}) = \sum_{n=1}^N z_n \sqrt{\lambda_n^{(c)}} f_n^{(c)}(\mathbf{x}), \quad (28)$$

where N is the number of terms required to generate the conditional random field and $z_n \sim N(0, 1)$ is the standard Gaussian random variable.

3. Results and Discussion

3.1. Equivalency Condition of Truncated Power Variogram and Power Variogram Models. To investigate the equivalency of statistically homogeneous or stationary truncated power variogram models and nonstationary power variogram model with homogeneous increments, a set of different I_u values in the truncated power variogram models are used in the increasing order to compare with the corresponding power variogram model. The results obtained by TpvE and TpvG models are very similar to each other, and it may be redundant to show both results in this work. Due to the larger range of H parameter in the TpvG model and thus its flexibility (especially in the inverse modeling process when observation scale needs to be accounted for), only the results obtained by TpvG model is shown in this paper. The lower limit I_l approaching to 0 corresponds with a point scale measurement. In reality, most of the measurements, such as the permeability measurements of soil cores, satisfy this condition since its scale is much smaller than the scale of domain size. Therefore, we set I_l as 0 and vary I_u in an increasing order to show the equivalency of TpvG with power variogram within a prescribed finite range of lag distance. As shown in (1) and (11), the parameter set is $\{A, H, I_u\}$ for TpvG model with a point measurement scale and $\{A_0, H\}$ for power variogram model. The relationship of A and A_0 is described by (14). Here, we set the parameter set of TpvG as $\{A, H\} = \{0.02, 0.25\}$ and the corresponding parameter set for power variogram model is $\{A_0, H\} = \{0.046, 0.25\}$ and then vary I_u from 10^2 to 10^6 . The obtained results are shown in Figure 1. It indicates that when I_u is small (e.g., $I_u = 10^2$), the TpvG model is associated with a finite sill and integral scale, which are 0.4 and 33.33, respectively. With the increasing of I_u , sill and integral scale increases. When I_u increases to 10^5 in this case, the TpvG model can approximate the power variogram model accurately in the range of the investigated lag distance.

3.2. Unconditional Multiscale Random Fractal Field. After determining the proper parameters for TpvG model to approximate the spatial distribution of multiscale random fractal field characterized by power variogram model (here we further increase I_u to 10^6 and the parameter set $\{A, H, I_u\} = \{0.02, 0.25, 10^6\}$ in TpvG is used to ensure the sufficient accuracy to approximate the power variogram with parameter set $\{A_0, H\} = \{0.046, 0.25\}$), a two-dimensional multiscale random fractals field can be generated by using KL expansion method with the TpvG model given by (12) as the covariance function. The procedures to generate the

FIGURE 1: The equivalency of TpvG and power variogram models.

unconditional multiscale random fractal field are shown in Figure 2(a). In this case, the generated 50×50 random multiscale fractal field is shown in Figure 3. As can be found in Figure 3, the upper region is associated with relative large values and the lower right region is associated with relative small values. To confirm the quality of the generated unconditional multiscale random fractal field, the sample variogram of the generated 2500 data values is used to compare with its theoretical counterpart. The sample variogram can be estimated using

$$\hat{\gamma}(s) = \frac{1}{2N_s} \sum_{(i,j) \in N_s} |\mathbf{x}_i - \mathbf{x}_j|^2, \quad (29)$$

where N_s is the number of pairs of samples i, j such that $|\mathbf{x}_i - \mathbf{x}_j| = h$. This task can be achieved by using the GAMV module in GSLIB code. As can be observed in Figure 4, the sample variogram model can reproduce the variogram structure of theoretical power variogram with parameter set $\{A_0, H\} = \{0.046, 0.25\}$ very well, which justifies the quality of the generated multiscale random fractal field. Although not shown here, a series of similar tests with different power variogram model parameter values have also been performed. They include the cases when the Hurst coefficient $0 < H < 0.5$ (indicating negatively correlated spatial increments), $0.5 < H < 1$ (indicating a positively correlated spatial increments), and $H = 0.5$ (indicating independent spatial increments). All the test results confirm that sample variograms can coincide well with their corresponding theoretical variograms.

3.3. Capability of Dimensionality Reduction by Using KL Expansion. One of the significant advantages to use KL expansion to generate the multiscale random fractal field is that it has the capability of reducing the dimensionality of the field in probabilistic space. This can be achieved by

FIGURE 2: Flowchart to generate the random fractal fields: (a) unconditional; (b) conditional.

FIGURE 3: The generated unconditional multiscale random fractal field using KL expansion with the power variogram model parameters $\{A_0, H\} = \{0.046, 0.25\}$.

truncating out the expanded KL terms with small eigenvalues. The eigenvalues of covariance function are plotted in Figure 5. It can be found that the eigenvalues decrease dramatically with the increasing of the number of the remained terms after truncation. Approximately, all the eigenvalues are less than 1 after 200 terms. As stated in Chang and Zhang [28], each eigenvalue represents the energy and information content for each term to express the spatial variability of the random field. To determine the number of terms that needs to be retained

FIGURE 4: Comparison of the theoretical power variogram with parameters $\{A_0, H\} = \{0.046, 0.25\}$ and its corresponding sample variogram.

to accurately represent the random field, Chang and Zhang [28] propose to define an energy criterion E_c :

$$E_c = \frac{\sum_{i=1}^{N_r} \lambda_i}{\sum_{i=1}^{\infty} \lambda_i} = \frac{\sum_{i=1}^{N_r} \lambda_i}{D\sigma^2}, \quad (30)$$

FIGURE 5: Eigenvalues of KL expansion.

TABLE 2: The relationship of energy criterion and number of retained terms.

N_r	E_c
10	53.42%
50	69.59%
100	74.83%
200	79.43%
500	85.16%
1000	90.07%
1500	93.74%
2000	97.02%
2500	100.00%

where N_r is the number of the retained terms, D is the domain size, and σ^2 is the variance of the random field. The relationship of E_c and N_r is shown in Table 2. As can be observed in Table 2, the first 10 terms associated with large eigenvalues has already taken up to more than 50% of the total energy. To investigate how the number of the retained terms affects the generation of the multiscale random fractal field, a set of realizations of multiscale random fractal field are generated with different number of retained terms (e.g., 50, 100, 200, 500, 1000, and 1500) in KL expansion, as shown in Figure 6. It can be found that even only 50 leading terms, taking up to approximately 70% of the total energy, are retained in KL expansion during the generation of the multiscale random fractal field; the general pattern of the spatial variation can be captured. With the increasing of the number of the retained terms in KL expansion, the details of the heterogeneous characteristics become more and more revealed. When 1500 leading terms, corresponding to approximately 94% of the total energies, are retained in the KL expansion, the generated multiscale random fractal field is almost indistinguishable from that generated with full

KL expansion by comparing Figure 6(f) and Figure 3. This finding indicates that the heterogeneous characteristics of the random field are dominated by the leading KL expansion terms with large eigenvalues, which provides the foundation to conduct the dimensionality reduction. Since the small details of the heterogeneous characteristics may not affect the flow and transport characteristics dramatically, the KL expansion terms with small eigenvalues can be safely truncated to improve computational efficiency in the stochastic analysis of flow and transport. For example, it has been demonstrated that the mean and variance of hydraulic head obtained with only the first few leading KL expansion terms (e.g., $N_r = 6$ for a 1D case) is very close to those obtained from thousands of Monte Carlo simulation with full models [25]. Similar observations are also reported in Li et al. [26] and Liao and Zhang [27]. In a quantitative sense, if the polynomial chaos expansion (PCE) is used to construct a surrogate model for the state variables (e.g., hydraulic head) and KL expansion on the model parameter (e.g., log hydraulic conductivity), the number of model evaluation required to obtain the PCE coefficients is $(N_r + p)!/N_r!/p!$, where p is the degree of PCE. For a given p , the number of model evaluation is greatly reduced when N_r decreases; hence the construction of surrogate model can be more efficient.

3.4. Generation of Conditional Multiscale Random Fractal Field. When direct measurement data are available, it may require generating the multiscale random fractal field that can be conditional on these measurements. To take the measurements into account, the procedures to generate the conditional random fractal field are shown in Figure 2(b), and the measurement values are obtained from a realization of the generated unconditional field to maintain the underlying random fractal characterizations. If there are 100 measurement data (the location of the measurements is shown in Figure 7(a)) collected from the unconditional multiscale random fractal field, the two generated conditional random fractal field can be depicted as shown in Figures 7(b) and 7(c). It can be observed that the generated values on the measurement locations honor the measurement values as shown in Figure 7(d), and the spatial correlation of the values on the unmeasured locations is governed by the given power variogram. Therefore, both the measurement values and the multiscale random fractal features can be represented in the generated unconditional field. With the aid of the generated multiscale random fractal field conditioning on the available measurements on the hydrogeological properties, the conditional stochastic analysis on flow and transport can be further performed.

3.5. Advantages and Limitations of TPV Models in Simulating Real Fields. This study is focused on random field generation, but it will be helpful to discuss when and how the proposed method can be used to characterize an aquifer based on the real observations. Due to the diversity and complexity of the geologic heterogeneity, there exists no versatile geostatistical model which works consistently well for all cases. Different models and tools have been developed to describe specific characteristics, for example, (non)stationarity, log-normal

FIGURE 6: Generated multiscale random fractal field with different numbers of retained terms in KL expansion: (a) 50; (b) 100; (c) 200; (d) 500; (e) 1000; (f) 1500.

FIGURE 7: Conditional random fractal fields: (a) locations of the measurement data; (b) and (c) realizations of conditional random fractal fields; (d) comparison of measurements and generated values on the 100 measurement locations.

distribution, channeling, and self-similarity or fractal scaling [9, 21, 29]; [Carle and Fogg, 1996]. Compared to traditional geostatistical models employed in GSLIB, TPV models have some additional and unique merits. First, they can capture fractal scaling, a widely reported feature in real sediments, rocks, aquifers, and reservoirs ([30–37], among many others). Fractal scaling in many cases can be significant but cannot be described by traditional models. Second, unlike traditional models which are limited to stationarity assumption, TPV models are able to treat true fields as nonstationary in general and allow them to be stationary through upper cutoffs representing the impact of finite observation ranges or geologic homogeneity. That is, TPV models are compatible to both stationary fields and nonstationary fields, as demonstrated in Figures 1 and 4.

In the meanwhile, TPV models inherit the advantages of the traditional models; for example, they can take the

anisotropy, nested structures, and conditioning effect into account. In order to capture the property of geological layering in real fields, one may choose long correlation lengths in the horizontal directions and a much shorter one in the vertical direction. More specifically, the parameters H and/or I_u above can be anisotropic to reflect the anisotropy of geologic layering.

As discussed above, it may be unwise to expect a versatile tool to handle all the fields. TPV model has its limitations in certain circumstances. The main limitation is that TPV, as a member of two-point statistical tools, is weak in featuring apparently observed connectivity, branching, or channeling. At the present stage, to solve this problem one may also need postprocessing using tools like sequential indicator simulation or directly resort to multipoint statistics (MPS [29, 38]). Combining fractal scaling and connectivity will be a promising work in the future.

4. Conclusion

Traditional geostatistical method used to characterize the heterogeneity of the hydrogeological properties is assumed that the random field is statistically homogeneous or stationary. However, growing evidences show that the heterogeneity of the hydrogeological random field is a nonstationary stochastic process but with a statistically homogeneous increments. Such field can be characterized by random fractals with the spatial variation described by power variogram model. It can be used to characterize the multiscale effect of the random field. To investigate the scaling effect of the random fractals on the flow and transport, it is necessary to generate such random fractals field numerically. Since power variogram does not have a corresponding covariance function, we adopt the stationary truncated variogram models to approximate the power variogram model and combine these models with Karhunen–Loève (KL) expansion to generate the random fractal field.

The main findings in this work lead to the following conclusions:

- (1) The truncated power variogram model can approximate the power variogram model in the range of the investigation region when the upper limit of the integral scale is very large under the point measurement scale condition. The truncated power variogram model is stationary, which has an explicit form of the covariance function and can be used in the traditional geostatistical analysis based on the stationary assumption.
- (2) KL expansion is an effective way to generate random field by decomposing the covariance function into its eigenvalues and eigenfunctions. The multiscale random fractal field can be generated by combining the stationary truncated power variogram model with KL expansion.
- (3) The main structure of the spatial heterogeneity is characterized by the dominating terms in the KL expansion with relative large eigenvalues. The eigenvalues of the TpvG covariance function decreases rapidly, which provides a foundation to conduct dimensionality reduction by truncating out the KL expansion terms associated with small eigenvalues.
- (4) The conditional multiscale random fractal field can be generated by using the conditional Kriging techniques by solving the Kriging equations when the measurement data are available. The generated values of conditional multiscale random fractal field honor the measurement values on the measurement locations.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is funded by the National Natural Science Foundation of China (Grants nos. 41402199 and 41602250), National Science and Technology Major Project (Grants nos. 2016ZX05037003 and 2016ZX05060002), the Science Foundation of China University of Petroleum, Beijing (Grant no. 2462014YJRC038), China Postdoctoral Science Foundation (Grants nos. 2016M591353 and 2017M611782), the Science Foundation of Sinopec Group (Grant no. P16058), the independent research funding of State Key Laboratory of Petroleum Resources and Prospecting (Grant no. PRP/indep-4-1409), and China Geological Survey (Grant no. DD20160293).

References

- [1] E. A. Sudicky, "A natural gradient experiment on solute transport in a sand aquifer: Spatial variability of hydraulic conductivity and its role in the dispersion process," *Water Resources Research*, vol. 22, no. 13, pp. 2069–2082, 1986.
- [2] P. Dietrich and C. Leven, "Direct push-technologies," in *Groundwater Geophysics*, R. Kirsch, Ed., pp. 347–366, Springer, Berlin, Germany, 2006.
- [3] S. J. Berg and W. A. Illman, "Three-dimensional transient hydraulic tomography in a highly heterogeneous glaciofluvial aquifer-aquitard system," *Water Resources Research*, vol. 47, no. 10, Article ID W10507, 2011.
- [4] G. Dagan, *Flow and Transport in Porous Formations*, Springer, New York, NY, USA, 1989.
- [5] L. W. Gelhar, *Stochastic Subsurface Hydrology*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1993.
- [6] J. P. Delhomme, "Spatial variability and uncertainty in groundwater flow parameters: A geostatistical approach," *Water Resources Research*, vol. 15, no. 2, pp. 269–280, 1979.
- [7] R. J. Hoeksema and P. K. Kitanidis, "Prediction of transmissivities, heads, and seepage velocities using mathematical modeling and geostatistics," *Advances in Water Resources*, vol. 12, no. 2, pp. 90–102, 1989.
- [8] R. A. Carlson and J. L. Osiensky, "Geostatistical analysis and simulation of nonpoint source groundwater nitrate contamination: A case study," *Environmental Geosciences*, vol. 5, no. 4, pp. 177–186, 1998.
- [9] C. V. Deutsch and A. G. Journel, *GSLIB: Geostatistical Software Library and User's Guide*, Applied Geostatistics Series, Oxford University Press, Oxford, UK, 2nd edition, 1997.
- [10] S. P. Neuman, "Universal scaling of hydraulic conductivities and dispersivities in geologic media," *Water Resources Research*, vol. 26, no. 8, pp. 1749–1758, 1990.
- [11] L. W. Gelhar, C. Welty, and K. R. Rehfeldt, "A critical review of data on field-scale dispersion in aquifers," *Water Resources Research*, vol. 28, no. 7, pp. 1955–1974, 1992.
- [12] F. J. Molz and G. K. Boman, "Further evidence of fractal structure in hydraulic conductivity distributions," *Geophysical Research Letters*, vol. 22, no. 18, pp. 2545–2548, 1995.
- [13] S. Painter, "Evidence for non-Gaussian scaling behavior in heterogeneous sedimentary formations," *Water Resources Research*, vol. 32, no. 5, pp. 1183–1195, 1996.
- [14] L. Tennekoon, M. C. Boufadel, D. Lavalley, and J. Weaver, "Multifractal anisotropic scaling of the hydraulic conductivity," *Water Resources Research*, vol. 39, no. 7, pp. 113–117, 2003.

- [15] J. Eggleston and S. Rojstaczer, "Inferring spatial correlation of hydraulic conductivity from sediment cores and outcrops," *Geophysical Research Letters*, vol. 25, no. 13, pp. 2321–2324, 1998.
- [16] M. C. Boufadel, S. Lu, F. J. Molz, and D. Lavalley, "Multifractal scaling of the intrinsic permeability," *Water Resources Research*, vol. 36, no. 11, pp. 3211–3222, 2000.
- [17] H. H. Liu and F. J. Molz, "Discrimination of fractional Brownian movement and fractional Gaussian noise structures in permeability and related property distributions with range analyses," *Water Resources Research*, vol. 32, no. 8, pp. 2601–2605, 1996.
- [18] F. Ruan and D. McLaughlin, "An efficient multivariate random field generator using the fast Fourier transform," *Advances in Water Resources*, vol. 21, no. 5, pp. 385–399, 1998.
- [19] P. R. Kramer, O. Kurbanmuradov, and K. Sabelfeld, "Comparative analysis of multiscale Gaussian random field simulation algorithms," *Journal of Computational Physics*, vol. 226, no. 1, pp. 897–924, 2007.
- [20] C. Cameron, "Relative efficiency of Gaussian stochastic process sampling procedures," *Journal of Computational Physics*, vol. 192, no. 2, pp. 546–569, 2003.
- [21] V. Di Federico and S. P. Neuman, "Scaling of random fields by means of truncated power variograms and associated spectra," *Water Resources Research*, vol. 33, no. 5, pp. 1075–1085, 1997.
- [22] L. Xue and D. Zhang, "A multimodel data assimilation framework via the ensemble Kalman filter," *Water Resources Research*, vol. 50, no. 5, pp. 4197–4219, 2014.
- [23] F. Heße, V. Prykhodko, S. Schlüter, and S. Attinger, "Generating random fields with a truncated power-law variogram: A comparison of several numerical methods," *Environmental Modeling and Software*, vol. 55, pp. 32–48, 2014.
- [24] T. Dieker, *Simulation of fractional Brownian motion [M.S., thesis]*, Department of Mathematical Sciences, University of Twente, Enschede, The Netherlands, 2004.
- [25] H. Li and D. Zhang, "Probabilistic collocation method for flow in porous media: Comparisons with other stochastic methods," *Water Resources Research*, vol. 43, pp. 6627–6632, 2007.
- [26] H. Li, H. Chang, and D. Zhang, "Stochastic collocation methods for efficient and accurate quantification of uncertainty in multiphase reservoir simulations," in *Proceedings of the SPE Reservoir Simulation Symposium 2009*, pp. 509–520, February 2009.
- [27] Q. Liao and D. Zhang, "Probabilistic collocation method for strongly nonlinear problems: 1. Transform by location," *Water Resources Research*, vol. 49, no. 12, pp. 7911–7928, 2013.
- [28] H. Chang and D. Zhang, "A comparative study of stochastic collocation methods for flow in spatially correlated random fields," *Communications in Computational Physics*, vol. 6, no. 3, pp. 509–535, 2009.
- [29] S. Strebelle, "Conditional simulation of complex geological structures using multiple-point statistics," *Mathematical Geology*, vol. 34, no. 1, pp. 1–21, 2002.
- [30] T. Hewett, "Fractal Distributions of Reservoir Heterogeneity and Their Influence on Fluid Transport," in *Proceedings of the SPE Annual Technical Conference and Exhibition*, New Orleans, LA, USA, 1986.
- [31] S. Painter and L. Paterson, "Fractional Lévy motion as a model for spatial variability in sedimentary rock," *Geophysical Research Letters*, vol. 21, no. 25, pp. 2857–2860, 1994.
- [32] A. G. Guzman and S. Neuman, "Field air injection experiments," in *Apache Leap Tuff INTRAVALEXperiments: Results and Lessons Learned*, T. Rasmussen, S. Rhodes, A. Guzman, and S. Neuman, Eds., Nuclear Regulatory Commission, 1996.
- [33] F. J. Molz, H. Rajaram, and S. Lu, "Stochastic fractal-based models of heterogeneity in subsurface hydrology: Origins, applications, limitations, and future research questions," *Reviews of Geophysics*, vol. 42, no. 1, 2004.
- [34] D. Veneziano and A. K. Essiam, "Nonlinear spectral analysis of flow through multifractal porous media," *Chaos, Solitons & Fractals*, vol. 19, no. 2, pp. 293–307, 2004.
- [35] V. Ganti, A. Singh, P. Passalacqua, and E. Foufoula-Georgiou, "Subordinated Brownian motion model for sediment transport," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 80, no. 1, 2009.
- [36] M. Siena, A. Guadagnini, M. Riva, and S. P. Neuman, "Extended power-law scaling of air permeabilities measured on a block of tuff," *Hydrology and Earth System Sciences*, vol. 16, no. 1, pp. 29–42, 2012.
- [37] A. Guadagnini, S. P. Neuman, M. G. Schaap, and M. Riva, "Anisotropic statistical scaling of soil and sediment texture in a stratified deep vadose zone near Maricopa, Arizona," *Geoderma*, vol. 214–215, pp. 217–227, 2014.
- [38] B. Jafarpour and M. Khodabakhshi, "A probability conditioning method (PCM) for nonlinear flow data integration into multipoint statistical facies simulation," *Mathematical Geosciences*, vol. 43, no. 2, pp. 133–164, 2011.