Research Article

Simulating Nonhomogeneous Non-Gaussian Field by Using Iterative Rank-Dependent Reordering versus Translation Process-Based Procedure

H. P. Hong and M. Y. Xiao

Department of Civil and Environmental Engineering, University of Western Ontario, London, ON, Canada

Correspondence should be addressed to H. P. Hong; hongh@eng.uwo.ca

Received 16 February 2022; Accepted 9 April 2022; Published 30 May 2022

Academic Editor: Filipe J. Marques

Copyright © 2022 H. P. Hong and M. Y. Xiao. 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.

We compare two commonly used procedures, namely, the iterative rank-dependent reordering (IRDR) procedure and the translation process based procedure, for simulating homogeneous/nonhomogeneous non-Gaussian fields. We identify the limitations and the implicit assumptions of the procedures. We provide a new interpretation of the IRDR procedure, point out that there is no guarantee that the algorithm converges, and suggest modifications in terms of the initial samples, iteration involving decomposition, and convergence requirement to the IRDR procedure for it to become more efficient and robust. The numerical results show that, depending on the prescribed marginal probability distribution, the use of the IRDR procedure may not achieve a prescribed correlation function, a feature that is well-known if the translation process (i.e., Nataf translation system) based procedure is employed. It is shown that the performance of the modified IRDR procedure is comparable to that of the translation process based procedures in terms of limitations and matching the prescribed correlation function. The numerical results also show that the suggested modifications to IRDR in the present study make the algorithm more efficient and robust.

1. Introduction

Probabilistic assessment of the responses of engineering systems involves modeling and simulating nonhomogeneous non-Gaussian random vector and field [1–3]. The stochastic modeling could involve random variables with different prescribed non-Gaussian marginal probability distributions and correlation structures. One of the approaches [4–8] takes advantage of the Nataf translation system [9] that approximates a system of non-Gaussian random variables or field by a Gaussian system. The adoption of this approximation facilitates the simulation of the nonhomogeneous and non-Gaussian fields. A key aspect of this approach is to find an underlying or equivalent Gaussian Pearson product-moment correlation coefficient (which will be simply referred to as correlation or Pearson correlation) to a prescribed correlation coefficient for the non-Gaussian correlated variables. This evaluation is based on the solution of a two-dimensional integral equation defining the Nataf translation system for each pair of random variables. There are two well-known possible weaknesses of this approach. The first one is that the underlying Gaussian correlation may not exist for the prescribed combination of non-Gaussian random variables and correlation coefficient. The problem could be dealt with by using the closest approximation to the solution to the integral equation. The second problem is that the correlation matrix (or function) obtained by solving the integral equation may not be positive semidefinite or cannot be decomposed by Cholesky decomposition. This drawback could be overcome by using the nearest correlation coefficient matrix [10, 11]. The use of the nearest correlation coefficient matrix was considered by Kim and Shields [12] in assessing the underlying equivalent Gaussian correlation matrix for the non-Gaussian random field, where the integral equation was solved numerically. The availability of the underlying Gaussian correlation matrix or function simplifies the simulation of the non-Gaussian vector based on the
Gaussian copula [13] and the simulation of the non-Gaussian random field.

Another approach relies on the rank-dependent reordering (RDR) or shuffling of sampled non-Gaussian random variables to match the prescribed correlation structure [14]. This RDR procedure assumes that the Spearman correlation and Pearson correlation are a good approximation of each other. The use of rank-dependent reordering was considered, among others, by Florian [15]; Olsson et al. [16]; and Yang [17]. Florian [15] emphasized the use of reordering to update Latin hypercube sampling. As the Spearman rank correlation differs from the Pearson correlation, a modification was suggested in Yang [17], which applies a particle swarm optimization algorithm to minimize the difference in the Pearson correlation calculated from the samples to its target. More recently, Zheng et al. [18] proposed simulating the nonhomogeneous non-Gaussian field by iteratively reordering the samples which are initially generated using the marginal distributions; the reordering is carried out according to their ranking such that the correlation matrix of the reordered samples matches the prescribed correlation matrix. Although no link between their procedure and that of Iman and Conover [14] was pointed out, this procedure calls for iteratively applying RDR. Moreover, the limitation of using the iterative RDR (IRDR) procedure to simulate the nonhomogeneous and non-Gaussian field is not thoroughly explored. A comparison of the performance of using the iterative rank-dependent reordering procedure versus using the popular translation process (i.e., Nataf translation system) based procedure to simulate a non-Gaussian field is unavailable in the literature.

For completeness, we note that other popular methods to simulate the random field include the spectral representation method (SRM) and the method based on the Karhunen-Loève (KL) expansion. A random field represented by the KL expansion is controlled by the random coefficients [1, 12, 19–21]. Since the direct use of the KL expansion with random coefficients leads to the Gaussian field, Phoon et al. [19, 20] proposed an iterative procedure that modifies and shuffles the random expansion coefficients based on the ranking to simulate the non-Gaussian process. Kim and Shields [12] used the underlying Gaussian correlation for the non-Gaussian field and the KL expansion to simulate the random field. SRM [22–24] is based on the power spectral density (PSD) function of the random field, where the relation between the PSD function and the covariance function of the field is established based on the Wiener–Khintchine theorem. A discussion of using SRM for conditional simulation of nonstationary processes was given by Cui and Hong [25]. SRM was extended to simulate nonstationary/nonhomogeneous and non-Gaussian fields by many researchers, including Yamazaki and Shinozuka [26]; Grigoriu [27]; Deodatis and Micaletti [28]; Ferrante et al. [29]; Shields and Deodatis [30]; and Hong and Cui [31]. These studies use iterative procedures to find the underlying Gaussian PSD function; the samples of the non-Gaussian field are obtained by transforming the samples generated from the underlying Gaussian PSD function. Without the need to find the underlying Gaussian PSD function, iterative algorithms are presented by Schreiber and Schmitz [32] and Masters and Gurley [33] to generate the surrogate or to simulate the homogeneous non-Gaussian field and by Hong et al. [34–36] and Cui and Hong [37] to simulate the nonhomogeneous and non-Gaussian fields or vectors. These algorithms involve reordering samples such that the PSD function of the sampled field matches the prescribed PSD function. It is acknowledged that the above review of techniques to simulate the non-Gaussian field is not exhausting, considering different complex mathematical modeling and techniques employed on this topic in the literature; a comprehensive review is beyond the scope of the present study.

The aim of the present study is twofold. First, we identify whether these two procedures share similar limitations in simulating homogeneous/nonhomogeneous non-Gaussian fields and the implicit assumptions of the procedures. Second, we provide a new interpretation of the IRDR procedure, point out that there is no guarantee that the algorithm converges, and suggest modifications in terms of the initial samples, iteration involving decomposition, and convergence requirement to the IRDR procedure for it to become more efficient and robust. In the following, we first summarize the RDR procedure and provide our interpretation of the IRDR. This serves to identify necessary modifications to the algorithm for it to become more efficient and robust. We then succinctly summarize the properties of the underlying Gaussian correlation based on the Nataf translation system [6, 7] to show the limitation of the Nataf translation system based procedure. This is followed by numerical examples comparing the performance of the IRDR procedure and the translation process based procedure and by conclusions.

2. Simulation Based on the IRDR Procedure and Suggested Modifications

Consider a one-dimensional random field \( Y(x) \) over the domain of \( x \) with prescribed marginal cumulative distribution function, \( F_x (x) \), and covariance or correlation function. For simplicity and without loss generality, it is considered, in the following, that \( Y(x) \) has zero mean and unit variance since the standard deviation and the mean values can be easily considered by scaling and shift.

In many practical applications, the data collection and data analysis for a random field are carried out at discrete points with a given sampling interval. Let \( \{ x \}_p = \{ x_1, \ldots, x_p \}^T \) represent a vector of \( x \) of size \( p \times 1 \) at the discrete points, where the superscript denotes the transpose and \( x_j \) represents the \( j \)-th element of the vector. It is considered that the correlation function of the field can be adequately represented by the correlation matrix (or covariance matrix) \( \Sigma_y \) of size \( p \times p \).

One could sample a matrix \( \{ U \} \) of size \( p \times m \) with its \( (i, j) \)-th element representing the value of the independent standard uniformly distributed random variable. A matrix \( \{ U \} \) of size \( p \times m \) with the \( (i, j) \)-th element \( z_{ij} = \Phi^{-1} (u_{ij}) \) can be formed. If \( Y(x) \) is a Gaussian random field, the simulation of \( m \) samples of the field can be obtained using
the well-known procedure based on the Cholesky decomposition,
\[
[Y] = [L_0][Z], \tag{1}
\]
where each column of \([Y]\) represents a sample of the random field and \(L_0\) denotes the lower triangular matrix obtained by decomposing \([C_{\rho_Y}]\) using the Cholesky decomposition. This decomposition can be carried out only if \([C_{\rho_Y}]\) is a positive semidefinite matrix [38]. The use of equation (1) ensures that the correlation matrix calculated from \([Y]\) equals the prescribed target \([C_{\rho_Y}]\).

To simulate non-Gaussian correlated random variables with a prescribed correlation matrix and to reduce potential spurious correlation, Iman and Conover [14] proposed a distribution-free RDR procedure in the context of Latin hypercube sampling. The procedure assumes that the Spearman (rank-dependent) correlation could be approximated by the Pearson (product-moment) correlation \([C_{\rho_Y}]\), which could be decomposed by using the Cholesky decomposition to obtain \([L_0]\). Multiple sets of the random vector or field \([Y]\) are sampled according to the marginal probability distribution (Gaussian or non-Gaussian) and Latin hypercube sampling technique. A row-wise matrix \([R]\) containing the row-wise ranking of \([Y]\) is formed. The samples of the correlated variables are then formed by reordering \([Y]\) according to the row-wise ranking of the matrix \([L_0] [R]\).

The RDRS procedure has been exploited in probabilistic analysis of engineering systems [15–17, 39]. Based on the RDR procedure, Olsson et al. [16] suggested simulating standard multivariate normal variate with a prescribed correlation matrix and then mapping the samples to the non-Gaussian space through the probability transformation. However, Olsson et al. [16] did not elaborate on which correlation matrix should be used. Following Iman and Conover [14], one could assume that the Pearson correlation \(\rho\) equals the target Spearman correlation \(\rho_s\). For bivariate normally distributed random variables, the relation between \(\rho_s\) and \(\rho\) is (e.g., [13])
\[
\rho_s = \frac{6}{\pi} \arcsin \left( \frac{\rho}{5} \right). \tag{2}
\]

The absolute difference between \(\rho_s\) and \(\rho\) is less than 0.0181, which occurs for \(\rho\) equal to \(\pm 0.594\). However, since \(\rho_s\) is usually unavailable and equation (2) is not applicable for the non-Gaussian field, it seems that \(\rho\) in the non-Gaussian space is employed to carry out the simulation as a practical measure. Based on the above considerations, the RDR procedure followed by Olsson et al. [16] for Latin hypercube sampling may be described as follows:

1. Calculate \([L_0]\) from \([C_{\rho_Y}]\) and sample \([Z]\) based on the standard normal distribution and the Latin hypercube or stratified sampling scheme but without considering correlation.
2. Calculate \([A_{\rho_Z}]=\{1/m\}[Z][Z]^T\), and apply the Cholesky decomposition to \([A_{\rho_Z}]\) to obtain the lower triangular matrix \([B_Z]\).

3. Calculate
\[
[Z^*] = [L_0][B_Z^{-1}][Z]. \tag{3}
\]
4. Determine the matrix \([R]\) containing the ranks based on \([Z^*]\) according to its row-wise ranking and reorder the samples in \([Z]\) according to \([R]\) (so a newly ordered \([Z]\) is obtained).
5. Map \([Z]\) to \([Y]\) by using the probability transformation, where \(y_{ij} = F^{-1}(z_{ij})(\Phi(z_{ij}))\), and \(y_{ij}\) and \(z_{ij}\) denote the \((i, j)\)-th element of \([Y]\) and \([Z]\), respectively.

Several observations are worth mentioning. First, we interpret \([Z]\) and \([Z^*]\) in Steps 1 to 3 as projections of another underlying sampled Gaussian field \([Z_0]\); they are given by \([Z] = [B_Z][Z_0]\) and \([Z^*] = [L_0][Z_0]\). From the first projection, we find the underlying sampled Gaussian field \([Z_0]\), and its use in the second projection leads to equation (3). While the projected \([Z]\) leads to a covariance matrix that differs from the prescribed covariance matrix, \([Z^*]\) leads to the prescribed covariance matrix in the Gaussian space, that is,
\[
[Z^*][Z^*]^T = [L_0][B_Z^{-1}][Z][[L_0][B_Z^{-1}][Z]^T = [C_{\rho_Y}]. \tag{4}
\]

Because the row-wise rank of \([R]\), \([Z^*]\), and the resulting matrix \([Z]\) in Step 5 are identical, the Spearman correlation matrix from each of the three matrices is identical by definition. However, this Spearman correlation matrix is likely to be different from the Pearson correlation matrix \([C_{\rho_Y}]\). In other words, the Pearson correlation from \([Z]\) is not equal to \([C_{\rho_Y}]\), although the difference is small if \([Z]\) is Gaussian. A critical issue with this simulation procedure is that there is no basis to infer if the Pearson correlation matrix obtained from \([Y]\) equals \([C_{\rho_Y}]\) if \([Y]\) is strongly non-Gaussian. It can be shown that the steps shown in the above are applicable, whether \([Z]\) in Step (1) is sampled from independent or correlated normally distributed random variables. In fact, if \([Z]\) in Step (1) is sampled by considering the correlation matrix \([C_{\rho_Y}]\), the reordering is unnecessary.

Rather than carrying out the sampling in the Gaussian space and mapping it to a non-Gaussian space, Zheng et al. [18] proposed to carry out the initial sampling according to its marginal distribution directly and reordering the samples iteratively. There is no restriction that a stratified simulation is to be considered. More specifically, their procedure consists of generating samples \([Y]\) based on the marginal probability distribution of \(Y(x)\) and iteratively approximating the prescribed correlation matrix by reordering. Although it was not pointed out in Zheng et al. [18], as will be seen, the similarity between their procedure in each iteration and that given by Iman and Conover [14] is striking. The steps of this IRDR procedure are as follows:

1. Calculate \([L_0]\) based on the prescribed \([C_{\rho_Y}]\) and sample \([Y]\) of size \(p \times m\) according to the prescribed marginal distribution of \(Y(x)\).
2. Calculate \([A_{\rho_Y}]=\{1/m\}[Y][Y]^T\), and apply the Cholesky decomposition to \([A_{\rho_Y}]\) to obtain the lower
triangular matrix, \([A_{yp}]\), where \([A_{yp}] = [B_{yp}][B_{yp}]^T\). The calculation of \([A_{yp}]\) is based on the biased estimate, and the mean and standard deviation of the samples are considered to be equal to zero and unit. This is adequate since \(m\) is considered to be very large. Otherwise, the calculation of \([A_{yp}]\) should be modified accordingly.

(3) Calculate \([Y^*] = [L_0]^{-1}[B_{yp}]^{-1} [Y]\).

(4) Determine \([R]\) based on \([Y^*]\) according to the row-wise ranking; reorder \([Y]\) following the row-wise ranking in \([R]\) (and retain it as new \([Y]\)). Calculate \([A_{yp}] = (1/m) [Y][Y]^T\).

(5) If \(|[A_{yp}] - [C_{yp}]| > \epsilon \| [C_{yp}] \|\), iterate Steps (3) to (5), where \(\epsilon\) is a preselected tolerance value, and the Frobenius norm for the matrix is adopted in the present study.

We note from the above that the IRDR procedure, except the iteration and no stratified sampling, is the same as the RDR procedure. Immediately after Step (4) in each iteration, the Spearman correlation matrices from \([R], [Y^*], \) and \([Y]\) are identical by definition. However, there is no guarantee that the algorithm converges for a selected tolerance. We make this assertion based on two factors. First, equation (2) is not applicable for non-normally distributed random variables, and the difference between the Spearman correlation and Pearson correlation, in general, is unknown. Second, the objective of the iterations is to reduce the difference between the Pearson correlation matrix calculated from the samples and the target (i.e., \(|[A_{yp}] - [C_{yp}]|\)), but the reordering is based on and aimed at matching Spearman correlation. Similar to equation (4), it is straightforward to show that the Pearson correlation matrix from \([Y^*]\) equals the prescribed \([C_{yp}]\). However, if \([Y^*]\) is used to represent the samples of the field, it may not match the prescribed marginal distribution of the non-Gaussian field; if \([Y]\) is used to represent the samples of the field, it may not match the prescribed Pearson correlation. One may also use the difference between \([Y^*]\) and \([Y]\) as a convergence measure. For the cases when the adopted convergence criterion in Step (5) is not achieved (examples will be presented shortly), we suggest assigning a maximum iteration number for the procedure and using \([Y]\) corresponding to the minimum \(|[A_{yp}] - [C_{yp}]|\) during the iteration as the samples of the random field.

Another problem with the algorithm for practical applications is that the assigned correlation matrix (which could be calculated using the prescribed correlation function) \([C_{yp}]\) may not be positive semidefinite. This will be elaborated by using numerical examples in the subsequent sections. A matrix must be at least positive semidefinite to be decomposed by the Cholesky decomposition [38]. To ameliorate this problem, we suggest substituting \([C_{yp}]\) with its corresponding nearest correlation matrix [10, 11]. The same could be done for \([A_{yp}]\) if it is not positive semidefinite. The use of the nearest correlation matrix makes the algorithm more robust, at least for when \([C_{yp}]\) is not positive semidefinite; the algorithm given in Qi and Sun [11] is employed for the numerical analysis to be carried out in the following sections.

Lastly, \([Y]\) in Step (1) could be simulated with or without an assigned correlation structure and the prescribed marginal distribution. The use of samples with an assigned correlation matrix could be advantageous in accelerating the convergence. For example, we could sample \([Y]\) in Step (1) based on Gaussian copula with the prescribed marginal distribution and the assigned correlation matrix equal to \([C_{yp}]\). This results in each element of \([Y]\) in Step (1) being equal to \(y_{i,j} = F_{x_i}^{-1}(\Phi(z_{\alpha,i,j}))\), where \(z_{\alpha,i,j}\) is the \((i, j)\)-th element of \([Z]\), and \([Z]\) equals \([L_0][Z]\).

The flowchart describing the modified IRSD procedure based on the above three considerations is shown in Figure 1.

3. Underlying Gaussian Correlation for Non-Gaussian Field Based on the Nataf Translation System

The Nataf translation system [9] approximates two correlated non-Gaussian variables by bivariate normally distributed random variables, where the underlying Gaussian correlation \(\rho_G\) is determined by solving the following integral equation [4, 6, 40]:

\[
c_{jk}(\rho_G) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F^{-1}_Y(x_j) F^{-1}_Y(x_k) \Phi(\psi_j; x_j, \rho_G) \Phi(\psi_k; x_k, \rho_G) \psi_j, \psi_k, \rho_G, d\psi_j d\psi_k,
\]

where \(c_{jk}(\rho_G)\) is the Pearson correlation coefficient between \(y(x_j)\) and \(y(x_k)\), \(\rho_G\) and \(\Phi(\psi_j; x_j, \rho_G)\) and \(\Phi(\psi_k; x_k, \rho_G)\) are the joint probability density function of two correlated standard normal variables, and \(\Phi\) is the standard normal distribution function. Several notable relations between \(c_{jk}(\rho_G)\) and \(\rho_G\) are as follows [6, 7]:

(a) \(|c_{jk}(\rho_G)| \leq |\rho_G|\). \(c_{jk}(\rho_G)\) is a strictly increasing function of \(\rho_G\) for continuous random variables with a finite variance value. \(c_{jk}(\rho_G) \geq 0\) if \(\rho_G \geq 0\) and \(c_{jk}(\rho_G) \leq 0\).

(b) The maximum and minimum values of \(c_{jk}(\rho_G)\), denoted as \(\tilde{\rho}_{NG,jk}\) and \(\tilde{\rho}_{NG,jk}\), equal \(c_{jk}(1)\) and \(c_{jk}(-1)\), respectively.

As a consequence of these properties \(c_{jk}(\rho_G) = 0\) if \(\rho_G = 0\). In order to find the underlying Gaussian symmetric correlation matrix with the diagonal elements equal to one for the non-Gaussian field, we could first determine \(\tilde{\rho}_{NG,jk}\) and \(\tilde{\rho}_{NG,jk}\). These values are used to obtain a modified \([C_{yp}]\), denoted as \([C_{yp}]\), with its \((j,k)\)-th element \(\rho^*_y(x_j, x_k)\) defined by

\[
\rho^*_y(x_j, x_k) = \begin{cases} 
\min(\rho_G(x_j, x_k), \tilde{\rho}_{NG,jk}), & \text{if } \rho_G(x_j, x_k) \geq 0 \\
\max(\rho_G(x_j, x_k), \tilde{\rho}_{NG,jk}), & \text{if } \rho_G(x_j, x_k) < 0 
\end{cases}
\]

where \(\rho_G(x_j, x_k)\) denotes the \((j,k)\)-th element of \([C_{yp}]\). For cases where \(\rho^*_y(x_j, x_k) \neq \tilde{\rho}_{NG,jk}\) or \(\rho^*_y(x_j, x_k) \neq \tilde{\rho}_{NG,jk}\) and \(j \neq k\), the underlying Gaussian correlation \(\rho_{G,jk}\) equals 1 and −1,
respectively. For the remaining cases and \( j \neq k \), \( \rho_{G, jk} \) could be obtained by solving equation (5) iteratively with \( c_{jk}(\rho_{G, jk}) = \rho_{c}^0(x_j, x_k) \), noting that the solution to \( \rho_{G, jk} \) is bracketed by \( c_{jk}(\rho_{G, jk}) \leq \rho_{c, jk} \leq 1 \) if \( 0 < c_{jk}(\rho_{G, jk}) \), and \( c_{jk}(\rho_{G, jk}) \leq 0 \) if \( c_{jk}(\rho_{G, jk}) < 0 \). If \( j = k \), \( \rho_{G, jk} = 1 \).

The underlying Gaussian correlation matrix \( [C_G] \) is formed by \( \rho_{G, jk} \). If \( [C_G] \) is not a positive semidefinite matrix, \( [C_G] \) is replaced by its corresponding nearest correlation matrix \( [\bar{C}_G] \) [10, 11].

The use of the Nataf translation system with the KL expansion to simulate the random field was considered by Kim and Shields [12]. Rather than using the KL expansion, for the field that is represented at \( p \) points (i.e., at \( [x]_p \)), we could simply sample the Gaussian field based on the Cholesky decomposition of \( [C_G] \) and then map the sampled Gaussian field to the non-Gaussian field to obtain \( [Y] \) with its \((i, j)\)-th element \( y_{ij} = F_{y(x),x}^{-1}(\Phi(z_{G,ij})) \), where \( z_{G,ij} \) is the \((i, j)\)-th element of \( [Z_G] = [L_G][Z] \), and \( [L_G] \) denotes the lower triangular matrix obtained by applying the Cholesky decomposition to \( [C_G] \) or \( [\bar{C}_G] \). For simplicity, this procedure is referred to as the Nataf translation system based on the NTS procedure in the following.

It must be emphasized that the exposition so far is focused on the one-dimensional random field. Since the simulation of the random field is carried out at discrete points by using the NTS procedure and the IRDR procedure, these procedures can be directly extended to the n-dimensional field by viewing and arranging all the discrete points in multidimension as a long vector.

4. Comparison of Matched Correlation, Underlying Gaussian Correlation, and Spearman Correlation

4.1. Prescribed Conditions for the Random Fields. A set of eight correlation functions were considered by Kim and Shields [12] to test their procedure with the use of the KL expansion to simulate the random field. For each prescribed correlation function, they considered two marginal distributions: one is weakly non-Gaussian, and the other is strongly non-Gaussian. These cases are listed in Table 1. Cases 1 to 4 describe homogeneous non-Gaussian fields, while Cases 5 to 8 describe nonhomogeneous non-Gaussian fields. The consideration of the beta distribution with the parameters given in Table 1 results in weakly non-Gaussian fields. The consideration of the shifted lognormal distribution with the distribution parameters assigned in Table 1 leads to strongly non-Gaussian fields. A simple spectral analysis indicates that Cases 3, 4, 7, and 8 represent narrow-band fields. The use of KL expansion, which was considered in Kim and Shields [12] in connection with the NTS procedure and in Zheng et al. [18] and in Xiao and Hong [41] in connection with the IRDR procedure, is beyond the scope of the present study.

4.2. Correlation from Sampled Fields Based on the IRDR Procedure. In this section, the numerical analysis results obtained by iteratively using the IRDR procedure to simulate the non-Gaussian field are presented. For the analysis, unless otherwise indicated, the tolerance value \( \varepsilon \) is set equal to \( 10^{-4} \), a maximum iteration is set equal to 20, and the field is represented at points with the separation \( \Delta x \) between two adjacent points equal to 0.1 for Cases 1 to 4 and equal to 0.05 for Cases 5 to 8. In all cases, the number of points representing the field equals 101. The value of \( \varepsilon \) for the first point is shifted from 0 to 0.002 for Cases 5 to 8, and the value of \( \varepsilon \) for the last point is shifted from one to 0.998 for Cases 6 and 8 to avoid zero correlation in the diagonal of the correlation matrix. Note that the selection of \( \Delta x \) does not affect the application of the considered two procedures; the selected values of \( \Delta x \) used for the considered examples are sufficiently small to represent the autocorrelation function. The adopted tolerance of \( \varepsilon = 10^{-4} \), which is defined as the relative difference in the Frobenius norm for the correlation matrix, for the convergence criterion, is based on the consideration that such a relative difference is deemed sufficiently small. The
Table 1: Definition of the cases (see [12]).

<table>
<thead>
<tr>
<th>Case</th>
<th>Prescribed autocorrelation function for homogeneous cases (1 to 4) or autocovariance function for nonhomogeneous cases (5 to 8)</th>
<th>Distribution type and notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$C_{\rho_1}(x_j, x_k) = \exp(-</td>
<td>x_j - x_k</td>
</tr>
<tr>
<td>2</td>
<td>$C_{\rho_2}(x_j, x_k) = \exp(-</td>
<td>x_j - x_k</td>
</tr>
<tr>
<td>3</td>
<td>$C_{\rho_3}(x_j, x_k) = \exp(-</td>
<td>x_j - x_k</td>
</tr>
<tr>
<td>4</td>
<td>$C_{\rho_4}(x_j, x_k) = \exp(-</td>
<td>x_j - x_k</td>
</tr>
<tr>
<td>5</td>
<td>$C_{\rho_5}(x_j, x_k) = \exp(-</td>
<td>x_j - x_k</td>
</tr>
<tr>
<td>6</td>
<td>$C_{\rho_6}(x_j, x_k) = \exp(-</td>
<td>x_j - x_k</td>
</tr>
<tr>
<td>7</td>
<td>$C_{\rho_7}(x_j, x_k) = \exp(-</td>
<td>x_j - x_k</td>
</tr>
<tr>
<td>8</td>
<td>$C_{\rho_8}(x_j, x_k) = \exp(-</td>
<td>x_j - x_k</td>
</tr>
</tbody>
</table>
considered maximum iteration is set equal to 20 since, as will be seen, the lowest relative difference in the Frobenius norm for the correlation matrix during the application of the procedure is achieved in less than 20 iterations.

The correlation matrix \([C_{xy}]\) is calculated by using the correlation function shown in Table 1 for each case. If \([C_{xy}]\) is not positive semidefinite, it is replaced by its nearest correlation matrix \([\tilde{C}_{xy}]\) [11]. It was found that \([C_{xy}]\) is not positive semidefinite only for Cases 2 and 4. The value of the element of \([\tilde{C}_{xy}] - [C_{xy}]\) is shown in Figure 2 for Cases 2 and 4. The results presented in the figure indicate that the absolute value of each element in \([C_{xy}] - [\tilde{C}_{xy}]\) is less than \(10^{-5}\).

Further analysis results by varying \(p\) from 5 to 100 indicate that the first value of \(p\) that the Cholesky decomposition to \([C_{xy}]\) breaks down equals 17 (i.e., \(\Delta x = 0.125\)) for Case 2 and 25 (i.e., \(\Delta x = 0.0833\)) for Case 4. This shows that the proposed modification to the IRDR procedure by considering the nearest correlation matrix is necessary.

By considering a sample size equal to 50,000 (i.e., \(m = 50,000\)), samples of the random field for each combination of cases and the marginal distributions listed in Table 1 are generated. For the sampling, in this section, the initial samples of \([Y]\) are obtained based on the marginal distribution alone (i.e., Option 1 shown in Figure 1). Typical samples of the random fields by using the IRDR procedure are shown in Figure 3. The Pearson correlation in the non-Gaussian space is calculated by using 50,000 samples for each considered combination. This calculated correlation is

![Figure 2: Value of the element of \([\tilde{C}_{xy}] - [C_{xy}]\) at \((x_k, x_j)\) for Cases 2 and 4.](image1)

![Figure 3: Typical samples generated by using the IRDR procedure.](image2)
The autocorrelation (AC) function calculated based on different assumptions and approaches is depicted in Table 2.

### Table 2: List of symbols used for the correlation matrix.

<table>
<thead>
<tr>
<th>Correlation matrix</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{Y}$</td>
<td>Pearson correlation obtained based on the prescribed correlation function</td>
</tr>
<tr>
<td>$C_{Y}$</td>
<td>Nearest correlation matrix to $C_{Y}$, if $C_{Y}$ is not positive semidefinite</td>
</tr>
<tr>
<td>$C_{Y}$, $R_{DS}$</td>
<td>Pearson correlation from $Y$ that are sampled using the IRDR procedure</td>
</tr>
<tr>
<td>$C_{Y}$, $R_{DS}$</td>
<td>Pearson correlation in Gaussian space using samples mapped from $Y$ that is simulated using the IRDR procedure</td>
</tr>
<tr>
<td>$C_{Y}$, $R_{DS}$</td>
<td>Spearman correlation from $Y$ that is sampled using the IRDR procedure</td>
</tr>
<tr>
<td>$C_{Y}$, $R_{DS}$</td>
<td>Underlying Gaussian correlation obtained from the prescribed correlation function and equation (5)</td>
</tr>
<tr>
<td>$C_{Y}$, $R_{N}$</td>
<td>Nearest correlation matrix to $C_{Y}$, if $C_{Y}$ is not positive semidefinite</td>
</tr>
<tr>
<td>$C_{Y}$, $N_{TS}$</td>
<td>Pearson correlation from $Y$ that are sampled using the NTS procedure</td>
</tr>
<tr>
<td>$C_{Y}$, $N_{TS}$</td>
<td>Spearman correlation from $Y$ that is sampled using the NTS procedure</td>
</tr>
</tbody>
</table>

The results presented in Figures 4 and 6 indicate that, for the weakly non-Gaussian cases, $C_{Y,R_{DS}}$ is practically equal to its target and that the difference between $C_{Y,R_{DS}}$ and its target can be large for strongly non-Gaussian cases. The differences are most noticeable for Cases 3, 4, 7, and 8 if the marginal distribution is the shifted lognormal distribution.

To investigate whether the Pearson and Spearman correlations of the sampled fields follow the same trend, we calculated the Spearman correlation from the sampled fields $C_{Y,R_{DS}}$. For the cases when the lognormal distribution is used as the marginal distribution, the differences in the autocorrelation function calculated based on $C_{Y,R_{DS}}$, $C_{Y,R_{DS}}$, and $C_{Y,R_{DS}}$ are presented in Figure 7 for Cases 1 to 4 and in Figure 8 for Cases 5 to 8. The results for the cases when the beta distribution is used as the marginal distribution are not plotted since these differences are small and negligible.

Results presented in Figures 7 and 8 indicate that the Spearman and Pearson correlations from the sampled fields follow the same trend for the considered cases, except for Case 4 with the lognormal distribution as the marginal distribution. This implies that, in general, the accuracy of matching the prescribed Pearson correlation is influenced by the Spearman correlation. It must be emphasized that whether $C_{Y,R_{DS}}$ provides an accurate representation of the actual Spearman correlation cannot be inferred since the actual Spearman correlation is not given.

An observation on the use of the IRDR procedure for the considered cases deserves comment. It is observed that, for the cases where the convergence is achieved, it is achieved in less than five iterations, as shown in Figure 9. The figure shows that the reduction of the error is most significant for the first reordering, implying that the RDR procedure [14] for the Latin hypercube sampling is equally effective for nonstratified sampling. Additional iterations could further reduce the error. However, the figure also shows that convergence is not always achieved, and the algorithm diverges for Cases 2 and 4 where the nearest correlation matrix is used. In such cases, $Y$ corresponding to the minimum $\varepsilon = ||A_{Y}|| - ||C_{Y}||$ is replaced by $C_{Y}$ for Cases 2 and 4 is used as the sampled fields; this adopted criterion is explained in the previous sections. After a few iterations, if $\varepsilon$ remains unchanged, it implies that there is no additional reordering. In all cases, the procedure is very efficient.

#### 4.3. Effect of the Initial Samples on Simulating Random Field by Using the IRDR Procedure.

In the previous section, the initial samples of the random fields are based on the marginal distribution alone (i.e., Option 1, see Figure 1). The simulation of the random fields is repeated but uses Option 2 to generate initial samples (i.e., using correlated samples defined by the Gaussian copula). To see the impact of Option 2 on the simulated fields, the value of $\varepsilon$ obtained in each iteration for all the considered cases is shown in Figure 10. The figure indicates that the error without considering reordering is already small, suggesting that the use of Gaussian copula with the correlation matrix in the Gaussian space equal to that in the non-Gaussian space already provides good approximations for the considered cases. The trends presented in Figure 10 are similar to those shown in Figure 9, except that the iteration number in Figure 10 is downshifted by one as compared to that depicted in Figure 9. The error for some cases decreases but increases for other cases due to reordering. If the error increases with an increased iteration, it indicates that the use of the reordering cannot further improve over the initial samples of the field that are generated based on the Gaussian copula with the correlation matrix equal to $C_{Y}$ (or $C_{Y}$).

As the obtained $C_{Y,R_{DS}}$ by using Option 2 to generate initial samples for the weakly non-Gaussian cases are practically identical to those shown previously based on Option 1, only the results for the strongly non-Gaussian cases are presented in the following. The AC function based on $C_{Y,R_{DS}}$ and its target are presented in Figure 11 for Cases 1 to 4 and in Figure 12 for Cases 5 to 8. A comparison of the results presented in Figures 4 and 11 and in Figures 6 and 12 indicates that $C_{Y,R_{DS}}$ is consistent whether Option 1 or Option 2 is used to generate the initial samples. Therefore, it is recommended to use Option 2 to generate initial samples in the IRDR procedure for efficiency.
4.4. Comparison of the Results from the IRDR and NTS Procedures. The numerical analysis by using the NTS procedure to simulate the non-Gaussian field for each combination of cases shown in Table 1 is carried out and presented in this section. Again, the field is discretized at 101 points for the simulation. As the results for the weakly non-Gaussian fields are very close to the Gaussian fields and are consistent with those obtained based on the IRDR procedure, only the results for cases that are strongly non-Gaussian are given in the following.

First, the underlying Gaussian correlation (or its corresponding nearest correlation) matrix $[C_G]$ is obtained by solving equation (5) for Cases 1 to 8 with the lognormal distribution as the marginal distributions listed in Table 1. Also, since $[C_G]$ for Cases 1 to 8 is not positive semidefinite, its corresponding nearest correlation matrix, $[\tilde{C}_G]$, is calculated and used in the following.
Figure 5: Comparison of the correlation matrix for Cases 5 to 8 by considering the beta distribution as the marginal distribution.

Figure 6: Comparison of the correlation matrix for Cases 5 to 8 with the lognormal distribution as the marginal distribution function.
Figure 7: Differences in the autocorrelation function calculated based on $[C_{\rho Y}]_{RDS} - [C_{\rho Y}]_{RDS}$ and $[C_{\rho sY}]_{RDS} - [C_{\rho sY}]$ for Cases 1 to 4 ($[C_{\rho Y}]$ for Cases 2 and 4 is replaced by $[C_{\rho sY}]$).

Figure 8: Continued.
Figure 8: Differences in the elements of (a) $[C_{\rho Y}]_{RDS} - [C_{\rho Y}]_{RDS}$ and of (b) $[C_{\rho Y}]_{RDS} - [C_{\rho Y}]$ for Cases 5 to 8.

Figure 9: Error $\epsilon$ versus iteration by applying the IRDR procedure. Option 1 is used to generate the initial samples.
Figure 10: Error $\epsilon$ versus iteration by applying the IRDR procedure $[C_{\rho Y}]$ is replaced by $[\tilde{C}_{\rho Y}]$ for Cases 2 and 4; Option 2 is used to generate initial samples.

Figure 11: Comparison of autocorrelation function calculated based on $[C_{\rho Y}]_{RDS}$ and its target ($[C_{\rho Y}]$ for Cases 2 and 4 is replaced by $[\tilde{C}_{\rho Y}]$). Option 2 is used to generate initial samples.
Simulation of the non-Gaussian random field, according to the NTS procedure, is carried out for a sample size of 50,000 for each of the considered cases. The Pearson correlation $[C_{xy}]_{NTS}$ and the Spearman correlation $[C_{xy}]_{NTS}$ are calculated by using $[Y]$ that are sampled based on the NTS procedure. A comparison of $[C_{xy}]_{NTS} - [C_{xy}]_{NTS}$ and $[C_{xy}]_{NTS} - [C_{xy}]$ for Cases 1 to 4 is compared in Figure 13. The difference between $[C_{xy}]_{NTS}$ and $[C_{xy}]_{NTS}$ and between $[C_{xy}]_{NTS}$ and $[C_{xy}]$ for Cases 5 to 8 is shown in Figure 14. A comparison of the results presented in Figures 7 and 13 indicates that while the results presented in Figure 7 for Cases 1 and 2 are mostly less than zero, those shown in Figure 13 are greater than zero. For Cases 3 and 4, although there are differences, the maximum magnitude of the differences is consistent. Also, a comparison of the results presented in Figures 8 and 14 indicates that the trends observed in Figure 14 are consistent with those that can be observed in Figure 8. These observations and the comparison shown in Figures 4 to 6 imply that $[C_{xy}]_{NTS}$ is consistent with $[C_{xy}]_{IRDR}$, suggesting that the use of the IRDR procedure and the NTS procedure, at least quantitatively, is equally adequate and also shares similar limitations in matching the prescribed correlation.

Figure 12: Comparison of the elements of the correlation matrix for Cases 5 to 8 with the lognormal distribution as the marginal distribution function. Option 2 is used to generate initial samples.
Case 1, Lognormal distribution

\[ |C_{\rho Y}^{NTS} - |C_{\rho Y}^{NTS} \]

Case 2, Lognormal distribution

\[ |C_{\rho Y}^{NTS} - |C_{\rho Y}^{NTS} \]

Case 3, Lognormal distribution

\[ |C_{\rho Y}^{NTS} - |C_{\rho Y}^{NTS} \]

Case 4, Lognormal distribution

\[ |C_{\rho Y}^{NTS} - |C_{\rho Y}^{NTS} \]

**Figure 13:** Differences in the autocorrelation function calculated based on \(|C_{\rho Y}^{NTS} - |C_{\rho Y}^{NTS}\) and \(|C_{\rho Y}^{NTS} - |C_{\rho Y}^{NTS}\) for Cases 1 to 4 with the lognormal distribution as the marginal distribution function (\(|C_{\rho Y}\) for Cases 2 and 4 is replaced by \(|C_{\rho Y}|).\)

**Figure 14:** Continued.
To quantify the differences in matching the target correlation matrix by using the IRDR procedure and NTS procedure, the smallest and largest values of the elements in $[C_{xy}]_{IRDR} - [C_{xy}]$ and in $[C_{xy}]_{NTS} - [C_{xy}]$ are calculated and shown in Table 3 for all cases by considering the beta distribution or the lognormal distribution as the marginal distribution. The results presented in Table 3 show that the IRDR procedure outperforms the NTS procedure for the weakly non-Gaussian fields; this is reversed for the strongly non-Gaussian fields. Since there is no need to evaluate the underlying Gaussian correlation in the IRDR procedure, this procedure is much more efficient than the NTS procedure.

### Table 3: The smallest and largest values of the elements in $[C_{xy}]_{IRDR} - [C_{xy}]$ and in $[C_{xy}]_{NTS} - [C_{xy}]$.

<table>
<thead>
<tr>
<th>Target correlation</th>
<th>Case</th>
<th>Beta distribution as marginal</th>
<th>Lognormal distribution as marginal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$[C_{xy}]<em>{IRDR} - [C</em>{xy}]$</td>
<td>$[C_{xy}]<em>{NTS} - [C</em>{xy}]$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Smallest</td>
<td>Largest</td>
</tr>
<tr>
<td>Positive</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-0.001</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-0.049</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-0.008</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>-0.001</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>-0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>Negative</td>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.000</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-0.017</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.000</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.000</td>
<td>0.002</td>
</tr>
</tbody>
</table>

### 5. Conclusions

In the present study, we emphasize the underlying assumption for the iterative rank-dependent reordering (IRDR) procedure, indicating that while the objective of simulating the random field is to match the prescribed Pearson correlation, the iterative algorithm reorders the samples based on the Spearman correlation. This mismatching results in that the Pearson correlation of the samples of the random field differs from its target, especially if the random field is highly non-Gaussian. The differences are shown by numerical examples.
It is pointed out that the Pearson correlation matrix for a given correlation function may not be positive semidefinite for practical applications. In such cases, we suggest replacing the target Pearson correlation matrix with its nearest correlation matrix. Since the IRDR procedure may not converge for a specified tolerance, we suggest incorporating correlated initial samples and taking a practical approach in assigning the sample fields (see Figure 1). In short, we propose modifications to the IRDR procedure in terms of the initial samples, iteration involving decomposition, and convergence requirements to make the IRDR more efficient and robust.

Results show that the limitation in matching the prescribed Pearson correlation by using the IRDR procedure to sample the nonhomogeneous and non-Gaussian fields is quantitatively similar to that of using the Nataf translation system. The numerical results for the considered examples also show that the IRDR procedure outperforms the translation process-based procedure for the weakly non-Gaussian; this trend is reversed for the considered strongly non-Gaussian fields.

Data Availability

All data used in the manuscript are available on request to H.P. Hong (the corresponding author of the manuscript).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors gratefully acknowledge the financial support received from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2016-04814, for HPH), the China Scholarship Council (no. 201807980004, for MYX), and the University of Western Ontario.

References


