Simulation of non-Gaussian processes using fractile correlation

Kok-Kwang Phoon*, Ser-Tong Quek, Hongwei Huang

Department of Civil Engineering, National University of Singapore, Block E1A #07-03, 1 Engineering Drive 2, Singapore, Singapore 117576

Received 23 July 2002; revised 17 September 2003; accepted 17 September 2003

Abstract

The difficulties of simulating non-Gaussian stochastic processes to follow arbitrary product–moment covariance models and arbitrary non-Gaussian marginal distributions are well known. This paper proposes to circumvent these difficulties by prescribing a fractile correlation function, rather than the usual product–moment covariance function. This fractile correlation can be related to the product–moment correlation of a Gaussian process analytically. A Gaussian process with the requisite product–moment correlation can be simulated using the Karhunen–Loève (K–L) expansion and transformed to satisfy any arbitrary marginal distribution using the usual CDF mapping. The fractile correlation of the non-Gaussian process will be identical to that of the underlying Gaussian process because it is invariant to monotone transforms. This permits the K–L expansion to be extended in a very general way to any second-order non-Gaussian processes. The simplicity of the proposed approach is illustrated numerically using a stationary squared exponential and a non-stationary Brown–Bridge fractile correlation function in conjunction with a shifted lognormal and a shifted exponential marginal distribution.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Non-Gaussian stochastic process; Non-stationary process; Fractile correlation; Product–moment correlation; Karhunen–Loève expansion; CDF mapping

1. Introduction

The simulation of non-Gaussian processes based on memoryless non-linear transform of an underlying Gaussian process is still an area of active research. Such processes are known as translation processes [1]. The transform has to be monotonic in order to ensure that the compatibility of the Gaussian finite-dimensional distributions can be preserved. A stochastic process may not exist if its finite-dimensional distributions are not compatible in the sense required by Kolmogorov’s conditions (e.g. refer to Refs. [2,3]). Grigoriu [4] uses this to demonstrate that non-Gaussian translation processes do exist and are stationary if the underlying Gaussian process is stationary. At present, it is quite accurate to say that translation from an underlying Gaussian process is a practical necessity since higher-order finite-dimensional distributions are frequently unavailable and Kolmogorov’s compatibility requirements should be satisfied (if possible) to establish existence on a firmer basis. It may be noted in passing that transformations with memory are also capable of generating non-Gaussian processes, but this approach is not well-studied and appears to be less convenient from a practical point of view [5].

The challenge in simulation of translation processes is to match a target covariance function and non-Gaussian distribution function simultaneously. It is quite well known that both functions cannot be prescribed arbitrarily. Grigoriu [4] demonstrated that the correlation of the underlying Gaussian process and the translated non-Gaussian process matches at 0 and 1, but correlation of the latter has a lower bound greater than $-1$ unless the transform is an odd function. The matching problem is less difficult when lower moments are prescribed, rather than the entire distribution function. This problem has been studied rather extensively by Gurley [6]. Yamazaki and Shinozuka [7] were among the first to propose a simulation algorithm for stationary non-Gaussian translation processes. Their approach consisted of correcting the power spectral density (PSD) function (Fourier transform of covariance function) of the underlying Gaussian process iteratively until the PSD function of the non-Gaussian process is achieved. The key assumption is that the PSD of the Gaussian process between the next and current iterate bears the same ratio as the target PSD and the PSD of the non-Gaussian process in the current iterate. Deodatis and Micaletti [8] noted that there are some
difficulties in the above approach for highly skewed non-Gaussian processes. The main one is that the substrate Gaussian process is no longer Gaussian and stationary after the first iterate. Grigoriu [4] proposed a solution in time domain by applying the sampling theorem to match the prescribed covariance function exactly at discrete points. It is necessary to solve an integral equation to determine the product–moment correlation of the underlying Gaussian process.

Phoon et al. [9] suggested using the Karhunen–Loève (K–L) expansion with non-Gaussian random variables to produce the desired non-Gaussian process. The principal difficulty is that the distributions of the non-Gaussian K–L random variables (not identical) are not the same as the prescribed non-Gaussian distribution for the process and seems to require an iterative solution approach. In principle, this approach is very attractive because it can be extended readily to non-stationary and multi-dimensional fields in a unified way. However, the proposed algorithm is currently unable to reproduce tail probabilities accurately for strongly non-Gaussian marginal distributions. Sakamoto and Ghanem [10] expanded the non-Gaussian process at discrete points using classical polynomial chaos, which is based on Hermite polynomials in terms of Gaussian random variables. Xiu and Karniadakis [11] noted that this representation only achieves optimal convergence rate for Gaussian and near Gaussian random fields, although it will eventually converge to any distribution in the mean square sense (more accurately, any $L^2$ functionals in the random space). The authors suggested using a more general polynomial chaos framework called Askey chaos to describe different distribution types in an optimal way. While Sakamoto and Ghanem [10] postulated that arbitrary correlations in the non-Gaussian process can be specified using a suitably correlated Gaussian process, Puig et al. [12] believed that this is not always possible, partially because the underlying Gaussian correlation function has to be non-negative definite. However, the authors seemed to assume that it is always possible to construct a Gaussian correlation function that is ‘close’ enough and non-negative definite at the same time. An optimization scheme was proposed to achieve this task.

The objective of this paper is to present a simple and practical method for generating non-Gaussian processes consistently and accurately. The difficulties noted previously are circumvented by prescribing a fractile correlation function, rather than the usual product–moment correlation function. Notice that a fractile correlation function can be obtained from empirical observations as easily as product–moment correlation. This fractile correlation can be related to the product–moment correlation of a Gaussian process analytically. A Gaussian process with the requisite product–moment correlation can be simulated optimally using the K–L expansion and transformed to satisfy any arbitrary marginal distribution using the usual CDF mapping. The K–L expansion is ‘optimal’ for Gaussian process in the sense that: (a) it provides a unified theoretical framework for higher dimensional and non-stationary fields defined over finite domains, (b) it has a bi-orthogonal property (both the deterministic basis functions and the corresponding random coefficients are orthogonal) that allows for optimal encapsulation of information contained in the random process into a set of discrete uncorrelated random variables, (c) it is able to represent highly correlated process using a relatively short expansion because only the first few eigenvalues are significant. Runtime savings in computing a short expansion is considerable because simulation typically involves hundreds of thousands of realisations, and (d) it can be accurate and computationally competitive if the wavelet-Galerkin approach is used for eigensolution [13].

The fractile correlation of the non-Gaussian process will be identical to that of the underlying Gaussian process because it is invariant to monotone transforms. The proposed approach is illustrated numerically using a stationary squared exponential and a non-stationary Brown–Bridge fractile correlation function in conjunction with a shifted lognormal and a shifted exponential marginal distribution. In summary, the theoretical advantage of the proposed approach is that a non-Gaussian process can be translated from a Gaussian one without iteration or solution of an integral equation, once the fractile correlation function is evaluated from observed time series. The practical disadvantage is that existing product–moment correlation functions (or the equivalent PSD functions) cannot be converted easily to fractile correlation functions. Note that this difficulty is fundamentally related to the definition of the product–moment correlation and its resulting coupling to the marginal probability distribution.

2. Method

Consider a Gaussian random process $\sigma(x, \theta)$ defined on a probability space $(\Omega, \mathcal{A}, P)$ and indexed on a bounded domain $D$. Assume that the Gaussian process has zero mean and unit variance. This standard Gaussian process can be represented using the K–L expansion as [14]

$$
\sigma(x, \theta) = \sum_{k=1}^{\infty} \sqrt{\lambda_k} \xi_k(\theta) f_k(x)
$$

(1)

where $\lambda_k$ and $f_k(x)$ are the eigenvalues and eigenfunctions of the Gaussian covariance (also correlation) function $\rho_{\sigma}(x_1,x_2)$, and $\xi_k(\theta)$ are uncorrelated standard Gaussian variates. It is possible to simulate Eq. (1) for the Gaussian case efficiently using a wavelet-Galerkin scheme [13].

This standard Gaussian process can be translated to a non-Gaussian process with marginal cumulative distributions $F_i = F_i(x_i)$ by the following memoryless transform:

$$
\xi(x_i) = F_i^{-1}[\sigma(x_i)]
$$

(2)
However, the product–moment correlation function of $\zeta$, $\rho_\zeta(x_1,x_2)$, is different from $\rho_{\mu}(x_1,x_2)$, which is the source of many difficulties. To convert $\rho_{\mu}$ to $\rho_{\psi}$ or vice-versa, it is necessary to iterate [7–9] or solve an integral equation [4].

In this paper, this problem is circumvented by assuming that the fractile correlation (sample analog being the Pearson or rank correlation) is prescribed, rather than the product–moment correlation (sample analog being the Spearman correlation). The fractile is defined as

$$U_i = F_{\zeta}[\zeta(x_i)]$$

(3)

which is uniformly distributed between 0 and 1 with mean $E(U_i) = 1/2$ and variance $E(U_i^2) = 1/12$.

By definition, the fractile correlation is given by:

$$r_{ij} = \frac{E(U_iU_j) - E(U_i)E(U_j)}{\sqrt{E(U_i^2)E(U_j^2)}} = 12E(U_iU_j) - 3$$

(4)

It is evident from definition that the fractile correlation is invariant under monotone transforms, specifically of the type given by Eq. (2):

$$U_i = F_{\zeta}[\zeta(x_i)] = F_{\Phi^{-1}}[\Phi(\zeta(x_i))] = \Phi(\zeta(x_i))$$

(5)

Hence, a non-Gaussian process generated using Eq. (2) will match the fractile correlation function of the underlying Gaussian process exactly.

The remaining problem is to determine the relationship between $r$ and the product–moment correlation $\rho_{\psi}$ of the underlying Gaussian process. A closed-form solution for Gaussian random variables does exist [15]

$$\rho_{\psi} = 2 \sin \left( \frac{\pi}{6} r \right)$$

(6)

The difference $\rho_{\psi} - r$ is small with a maximum deviation of only 0.018 as depicted in Fig. 1. This difference must exist because the product–moment correlation $\rho_{\psi}$ is based on cross-products of Gaussian realisations (95% of the values lying between $\pm 1.96$ although larger values are theoretically possible), while the fractile correlation $r$ is based on cross-products of uniform realisations (values strictly lying between 0 and 1). Note that the mapping between $\rho_{\psi}$ and $r$ extends over the full range $[-1,1]$. It has been recognised that Eq. (6) only produces a valid correlation function if it is non-negative definite, i.e. the property of non-negative definiteness is an additional requirement [4].

The proposed method can be briefly summarized as follows:

1. Prescribe a fractile correlation function $r$ and the marginal cumulative distributions $F_i$ (not identical unless the process is stationary).
2. Convert the fractile correlation function to the product–moment correlation function $\rho_{\psi}$ using Eq. (6).
3. Compute eigenvalues and eigenfunctions of $\rho_{\psi}$ and verify that the minimum eigenvalue is larger or equal to zero, i.e. $\rho_{\psi}$ is non-negative definite.
4. Simulate realisations of the standard stationary or non-stationary Gaussian process using Eq. (1) and eigensolutions from step (3).
5. Translate the Gaussian process to follow the prescribed non-Gaussian distributions $F_i$ using Eq. (2).

It is clear that the non-Gaussian process generated using the above algorithm will match the prescribed fractile correlation and marginal distributions exactly. In addition, this process exists in the Kolmogorov sense because it is a simple monotonic memoryless transform of a Gaussian process (Eq. (2)). Fackler [16] and Iman and Conover [17] have applied similar procedures to induce rank correlations in random vectors. The former paper used fractile correlations rather than ranks. The latter paper obtained the desired correlated vector by re-arranging a previously generated uncorrelated non-Gaussian vector using ranks derived from Cholesky decomposition of the rank correlation matrix. This is essentially a Latin Hypercube sampling approach. Both papers recognised that the number of multi-variate uniform distributions with a specified correlation matrix is infinite, but recommended using the multi-variate Gaussian distribution to generate the vector of correlated ranks or fractiles

$$(U_1,\ldots,U_N) = (\Phi(Y_1),\ldots,\Phi(Y_N))$$

(7)

where $Y_i$ are correlated standard normal variates. Iman and Conover [17] observed that the use of normal variates will produce ‘natural elliptical’ correlation scatter plots while Fackler [16] felt that it is a convenient choice among many other possibilities.

For extension to continuous stochastic processes (separable type [3]), the use of Eq. (7) is probably more of a necessity than convenience because it is not easy to construct finite-dimensional distributions that are consistent and invariant to arbitrary permutations of the indices over every finite set of points as required by Kolmogorov’s conditions. In addition, the relationship between fractile correlation and product–moment correlation is available in the Gaussian case and both are almost equal (Fig. 1).
3. Numerical results

In this section, 10,000 realisations of the standard Gaussian process are simulated using a truncated K–L expansion consisting of 32 terms. The eigenvalues and eigenfunctions were evaluated numerically using the wavelet-Galerkin scheme described by Phoon et al. [13]. Two target fractile correlation models were considered:

1. Stationary squared exponential model defined over \([-1,1]\]
   \[ r(x_1, x_2) = e^{-(x_1-x_2)^2} \] (8)

2. Non-stationary Brown–Bridge model defined over \((0,1]\)
   \[ r(x_1, x_2) = \frac{\min(x_1, x_2) - x_1 x_2}{\sqrt{(x_1 - x_1^2)(x_2 - x_2^2)}} \] (9)

The marginal distributions are assumed to follow a shifted lognormal or shifted exponential CDF with zero mean and unit variance as follows:

1. Shifted lognormal
   \[ F(y; \mu, \sigma, \delta) = \Phi \left( \frac{\log(y - \delta) - \mu}{\sigma} \right) \] (10)
   where \( \mu = -0.7707, \sigma = 1, \) and \( \delta = -0.7628. \)

2. Shifted exponential
   \[ F(y; \lambda, \mu) = 1 - e^{-\lambda(y-\mu)} \] (11)
   where \( \lambda = 1 \) and \( \mu = -1. \)

Although the marginal distributions are assumed to be invariant under translation, it is easy to see that the distribution parameters or even the CDF can be defined as a function of the coordinate \( x \) without any difficulty.

Fig. 2 compares the simulated fractile and product–moment correlations for the stationary squared exponential correlation model with shifted lognormal and shifted exponential distributions. The simulated fractile and product–moment correlations are computed, respectively, from \( N \) realisations using:

\[ \hat{r} = 12 \frac{1}{N} \sum_{k=1}^{N} \{ \phi'[\phi(x_i, \theta_k)] \phi'[\phi(x_j, \theta_k)] \} - 3 \] (12)

\[ \hat{\rho}_z = \frac{1}{N} \sum_{k=1}^{N} [\zeta(x_i, \theta_k) \zeta(x_j, \theta_k)] \] (13)

As to be expected, the simulated fractile correlations are independent of the marginal distribution and in agreement with the target values. It is not surprising that there are differences between fractile and product–moment correlations, but it is reassuring that the latter shape bears strong similarity with the former (whose shape is typical of observed product–moment correlations). For stationary processes, the reason for this similarity between \( \rho_z \) and \( r \) can be partially explained by observing that: \( r \approx \rho_\sigma \)

Fig. 2. Simulated fractile and product–moment correlations for stationary non-Gaussian processes with squared exponential fractile correlation model and (a) shifted lognormal distribution, (b) shifted exponential distribution.

Fig. 3. Eigenvalues of simulated product–moment correlations for stationary non-Gaussian processes with squared exponential fractile correlation model.
(Fig. 1), (b) $|\rho_\xi| \leq |\rho_\sigma|$ [18] and (c) $\rho_\xi = 0$ and 1 if and only if $\rho_\sigma = 0$ and 1, respectively [4]. For positive correlations, the above broad theoretical results provide some constraints on the difference between $\rho_\xi$ and $\rho$. For negative correlations, the anchor point given by $\rho_\xi = -1$ and $\rho_\sigma = -1$ is not available unless the non-Gaussian cumulative distribution function $F$ is an odd function [4]. It may be noted in passing that another non-parametric measure of bivariate association, Kendall’s $\tau$, can also be related to the Gaussian product–moment correlation analytically using $\rho_\xi = \sin(\pi \tau/2)$ [19]. However, the difference between $\tau$ and $\rho_\sigma$ is much larger (maximum absolute difference = 0.21), potentially leading to more distortion between the prescribed non-parametric measure of association (Kendall’s $\tau$) and the resulting non-Gaussian product–moment correlation function.

The eigenvalues for the squared exponential fractile correlation function (Eq. (8)) are larger or equal to zero. However, the eigenvalues of the corresponding Gaussian product–moment correlation $\rho_\sigma$ between the sixth and ninth modes are slightly negative of order $10^{-4}$ or less. Hence, $\rho_\sigma$ is not an admissible correlation function. This implies that there is no Gaussian process that can produce this particular prescribed fractile correlation. The need to impose non-negative definiteness as an additional requirement for methods based on mapping of a bivariate measure of association (e.g. Eq. (6)) has been explained by Grigoriu [4]. A plausible method to circumvent this problem is to set these slightly negative eigenvalues to zero. The validity of this approximation is shown by the agreement between simulated and target fractile correlations in Fig. 2. The non-negative definiteness of the product–moment correlation functions for the desired non-Gaussian processes are shown in Fig. 3. For a variety of other common stationary correlation model such as the triangular, single exponential, linear exponential, and sine functions studied by Huang et al. [20], the corresponding Gaussian product–moment correlation functions are non-negative definite. Hence, the squared exponential correlation model is exceptional in this sense.

It is also interesting to examine the results under the assumption that the fractile and product–moment correlations for the Gaussian case are the same, i.e. ignore Eq. (6). The fractile and product–moment correlations under this assumption differ only slightly from the solutions generated using Eq. (6) as shown in Table 1.

![Fig. 4. Simulated fractile and product–moment correlations for non-stationary non-Gaussian processes with Brown–Bridge fractile correlation model and (a) shifted lognormal distribution, (b) shifted exponential distribution.](image-url)
In this example, the product–moment correlation \( r_{4} \) model.
stationary non-Gaussian processes with Brown–Bridge fractile correlation
Fig. 5. Eigenvalues of simulated product–moment correlations for non-
correlation measure. This permits the K–L expansion to be
marginal distribution because it is a non-parametric
fractile correlation can be prescribed independently of the
a more robust estimator for the same amount of data. The
empirical observations as easily as product–moment
monotone transforms.
non-Gaussian process because it is invariant to
non-Gaussian process will be identical to that of the
underlying Gaussian process is non-negative definite and it
therefore, not surprising that the product–moment
correlation functions of the desired non-Gaussian processes
are also non-negative definite as shown in Fig. 5.

4. Conclusions
The difficulties of simulating non-Gaussian stochastic
processes to follow arbitrary product–moment covariance
models and arbitrary non-Gaussian marginal distributions
are well known. This paper proposes to circumvent these
difficulties by prescribing a fractile correlation function,
rather than the usual product–moment covariance function.
This fractile correlation can be related to the product–
moment correlation of a Gaussian process analytically. A
Gaussian process with the requisite product–moment
correlation can be simulated using the K–L expansion and
transformed to satisfy any arbitrary marginal distribution
using the usual CDF mapping. The fractile correlation of the
non-Gaussian process will be identical to that of the
underlying Gaussian process because it is invariant to
monotone transforms.
The common use of product–moment correlation to
describe dependency is mostly a matter of convention than
convenience. A fractile correlation can be obtained from
empirical observations as easily as product–moment
and possibly the Pearson correlation statistic is a
more robust estimator for the same amount of data. The
fractile correlation can be prescribed independently of the
marginal distribution because it is a non-parametric
correlation measure. This permits the K–L expansion to be
extended in a very general way to any second-order
non-Gaussian processes. Note that the expansion is already
capable of handling non-stationary covariance functions and
multi-dimensional fields for the Gaussian case in an efficient
and unified way. Finally, the existence of such processes can
be rigorously established from Kolmogorov’s conditions by
observing that they are monotonic translations from a
standard Gaussian process. The simplicity of the proposed
approach is illustrated numerically using a stationary
squared exponential and a non-stationary Brown–Bridge
fractile correlation function in conjunction with a shifted
lognormal and a shifted exponential marginal distribution.

References
[3] Soong TT, Grigoriu M. Random vibration of mechanical and
Thesis, Department of Civil Engineering and Geological Science,
University of Notre Dame; 1997.
[9] Phoon KK, Huang SP, Quek ST. Simulation of second-order
80(12):1049–60. special issue on probabilistic mechanics and
structural reliability.
[10] Sakamoto S, Ghanem R. Polynomial chaos decomposition for the
simulation of non-Gaussian non-stationary stochastic processes.
619–44.
polynomial expansion: convergences and algorithms. Probab Engng
[13] Phoon KK, Huang SP, Quek ST. Implementation of Karhunen–Loeve
expansion for simulation using a wavelet-Galerkin scheme. Probab
[16] Fackler PL. Modeling interdependence: an approach to simulation and
[17] Iman RL, Conover WJ. A distribution-free approach to inducing rank
[18] Grigoriu M. Applied non-gaussian processes: examples, theory,
simulation, linear random vibration, and MATLAB solutions.
[20] Huang SP, Quek ST, Phoon KK. Convergence study of the truncated
Karhunen–Loeve expansion for simulation of stochastic processes.