Observations on non-Gaussian Karhunen-Loève expansions

L.B. Li*, S.T. Quek, K.K. Phoon

National University of Singapore, Civil Engineering Department, Singapore, 117576

Abstract

The non-Gaussian Karhunen-Loève (K-L) expansion has been used to generate a non-Gaussian process using an iterative scheme. Numerical results show that different non-Gaussian processes can be generated satisfying the same prescribed covariance function and marginal distribution by changing the assumed starting distribution of the K-L random variables. Non-Gaussian K-L processes produced by assuming an initial Gaussian distribution for the K-L random variables appear to be translation processes. When the K-L random variables were assigned a lognormal distribution before the iteration procedure, the resulting process is clearly non-translation. Hence, it would appear that translation processes form a subset of K-L processes. In other words, the class of non-Gaussian K-L processes is larger and potentially capable of providing better fit to observed data.

Keywords: Karhunen-Loève expansion; Modified Latin hypercube orthogonalization; Translation process; Multivariate Gaussianity; Up-crossing rate; Rank correlation

1. Introduction

The non-Gaussian Karhunen-Loève (K-L) expansion generates non-Gaussian processes based on prescribed covariance function and marginal distribution. The key feature of this technique is that the target covariance function is maintained, while the probability distributions of the K-L random variables are updated iteratively. In general, a non-Gaussian process cannot be defined by the first two moments uniquely. The well-known translation process [1] obtained via memoryless transformation of a Gaussian process may not produce a process that could match observed non-Gaussian data.

The non-Gaussian K-L expansion has the potential to simulate different non-Gaussian processes satisfying the same target covariance function and marginal distribution [2,3]. This is realized by assigning different distributions to the K-L random variables during the start of the iteration process (briefly described below). This paper examines the differences between non-Gaussian K-L and translation processes. The basic idea is to back-translate the K-L process such that its marginal distribution is Gaussian and to verify the following higher-order properties associated with a Gaussian process: (i) covariance, (ii) rank correlation, (iii) up-crossing rate, and (iv) multivariate Gaussianity.

* Corresponding author. Tel.: +65 68746347;
E-mail: g0306011.nus.edu.sg

2. Non-Gaussian Karhunen-Loève expansion

A random process $\varpi(x,\theta)$ defined on a probability space (Ω, A, B) and indexed on a bounded domain $x \in D$, having mean $\overline{\varpi}(x)$ and finite variance $\sigma^2(x)$, can be approximated using the following finite K-L expansion:

$$\varpi_M(x,\theta) = \overline{\varpi}(x) + \sum_{i=1}^{M} \sqrt{\lambda_i}\xi_i(\theta) f_i(x) \quad (1)$$

where λ_i and $f_i(x)$ are the eigenvalues and eigenfunctions of the covariance function $C(x_1,x_2)$, $\xi_i(\theta)$ is a set of uncorrelated K-L random variables with zero mean and unit variance, and M is the number of K-L terms.

If $\varpi(x,\theta)$ is a Gaussian process, then $\xi_i(\theta)$ is a vector of uncorrelated standard Gaussian random variables. For $\varpi(x,\theta)$ with an arbitrarily prescribed marginal distribution, the distributions of $\xi_i(\theta)$ are unknown. The following iterative steps were proposed to compute these unknown K-L distributions [4]:

1. Generate n sample functions of the non-Gaussian process:

Computational Fluid and Solid Mechanics 2005
K.J. Bathe (Editor)

$$\varpi_M^{(k)}(x,\theta_m) = \overline{\varpi}(x) + \sum_{i=1}^{M} \sqrt{\lambda_i}\xi_i^{(k)}(\theta_m) f_i(x), m = 1,2,\ldots n \tag{2}$$

where k = iteration number and m = sample number.

2. Estimate the empirical cumulative marginal distribution function as:

$$\widehat{F}_M^{(k)}(y|x) = \frac{1}{n}\sum_{m=1}^{n} I\left(\varpi_M^{(k)}(x,\theta_m) \le y\right) \tag{3}$$

where $I(event)$ = indicator function = 1 if event is true and 0 otherwise.

3. Transform each sample function to match the target marginal distribution F:

$$\eta_M^{(k)}(x,\theta_m) = F^{-1}\widehat{F}_M^{(k)}\left[\varpi_M^{(k)}(x,\theta_m)\right] \tag{4}$$

4. Estimate the next generation of $\xi_i(\theta)$ as:

$$\xi_i^{(k+1)}(\theta_m) = \frac{1}{\sqrt{\lambda_i}} \int_D \left[\eta_M^{(k)}(x,\theta_m) - \bar{\eta}_M^{(k)}(x)\right] f_i(x)dx \tag{5}$$

5. Standardize $\xi_i^{k+1}(\theta)$ to unit variance. Note that $\xi_i^{k+1}(\theta)$ is a zero-mean vector by virtue of Eq. (5). A modified Latin hypercube orthogonalization technique [4] is applied to reduce the product-moment correlations between $\xi_i^{k+1}(\theta)$. This technique is described below, assuming that the realizations of $\xi_i(\theta)$ are stored in an $n \times M$ matrix $\boldsymbol{X}$:
   (i) Compute the $M \times M$ product-moment covariance matrix of $\boldsymbol{X}$:

$$\boldsymbol{T} = \frac{\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}}{n-1} - \frac{\boldsymbol{X}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}}{n(n-1)} \tag{6}$$

   where $\boldsymbol{U}$ is a $n \times 1$ vector containing ones.
   (ii) Obtain an uncorrelated realization matrix $\boldsymbol{X}'$ by:

$$\mathbf{X}' = \mathbf{X}\mathbf{Q}^{-1} \tag{7}$$

   where $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{T}$
   (iii) Re-order the realizations in each column of $\boldsymbol{X}$ to follow the ranking of realizations in each column of $\boldsymbol{X}'$.

6. Repeat steps (1) through (5) until the sample functions achieve the target marginal distribution.

3. Are non-Gaussian K-L processes 'translation'?

Although the translation method produces a large class of non-Gaussian processes, it will be shown in the next section that it is a special case of the K-L method. In other words, the class of non-Gaussian K-L processes is even larger and potentially capable of providing better fit to observed data.

To demonstrate the above, the basic idea is to back-translate the K-L process such that its marginal distribution is Gaussian and to verify the following higher-order properties associated with a Gaussian process: (i) covariance, (ii) rank correlation, (iii) up-crossing rate, and (iv) multivariate Gaussianity.

For illustration, consider a standard Gaussian process (ϖ_g) with the following covariance function:

$$\rho(x_1,x_2) = e^{-|x_1-x_2|} \tag{8}$$

A non-Gaussian process satisfying a prescribed marginal cumulative distribution function F can be constructed using the translation method [1]:

$$\varpi(x) = F^{-1}\boldsymbol{\Phi}[\varpi_g(x)] \tag{9}$$

where $\boldsymbol{\Phi}(\cdot)$ denotes the cumulative distribution of the standard Gaussian variate. A shifted exponential cumulative distribution function is selected for F:

$$F(y,\mu,\lambda) = 1 - e^{-\lambda(y-\mu)} \tag{10}$$

The distribution parameters $\lambda = 1$ and $\mu = -1$ are selected to produce zero mean and the covariance of the non-Gaussian translation process $C(x_1,x_2)$ is [5]:

$$C(x_1,x_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F^{-1}[\Phi(z_1)]F^{-1}[\Phi(z_2)]\phi[z_1,z_2;\rho(x_1,x_2)]dz_1dz_2 \tag{11}$$

where

$$\phi[z_1,z_2;\,\rho(x_1,x_2)] = \frac{1}{2\pi[1-\rho(x_1,x_2)^2]^{1/2}} \exp\left\{-\frac{z_1^2+z_2^2-2z_1z_2\rho(x_1,x_2)}{2[1-\rho(x_1,x_2)^2]}\right\} \tag{12}$$

It is possible to simulate non-Gaussian K-L processes satisfying F and C given by Eqs. (10) and (11), respectively. If these processes are translation, the back-translated process $\varpi_b(x) = \Phi^{-1}F[\varpi(x)]$ should produce the covariance function given in Eq. (8) because $\varpi_b(x) = \varpi_g(x)$. This is the most direct check. In this example, two K-L processes were simulated by assigning the initial distributions of $\xi_i(\theta)$ to be Gaussian in the first case and lognormal in the second case before applying the iterative procedure outlined in Section 2. Note that distributions of $\xi_i(\theta)$ are unknown *a priori* and should not be confused with the marginal distribution of the process itself, which is shifted exponential. The processes are denoted by $\varpi_i(x)$ for i = 1 and 2, and the corresponding back-translated processes are $\varpi_{bi}(x)$. Both processes are indexed over $x \in [0, 1]$. A finite K-L expansion with $M = 32$ terms is used. Eigenvalues and eigenfunctions were computed numerically using the wavelet-Galerkin approach, based on $2^5 = 32$ wavelet basis functions [6]. The sample size is $n = 10000$, and 40 iterative steps are used.

Numerical results show that the covariance and marginal distribution of the simulated processes almost coincide with the respective targets, regardless of the initial distribution assumption for the K-L random variables (Gaussian or lognormal). This is to be expected if the iterative procedure is working correctly. The maximum relative error for the covariance defined by $\max_{x_1,x_2}[|\widehat{C}(x_1,x_2) - C(x_1,x_2)|/C(x_1,x_2)]$ (where $\widehat{C}(x_1,x_2)$ is the covariance of the simulated process) is 2.5×10^{-4} and 2.1×10^{-4} for $\varpi_1(x)$ and $\varpi_2(x)$, respectively. For the marginal distribution, the maximum deviation between F and empirical cumulative distribution function is 0.002 and 0.0029 for $\varpi_1(x)$ and $\varpi_2(x)$, respectively.

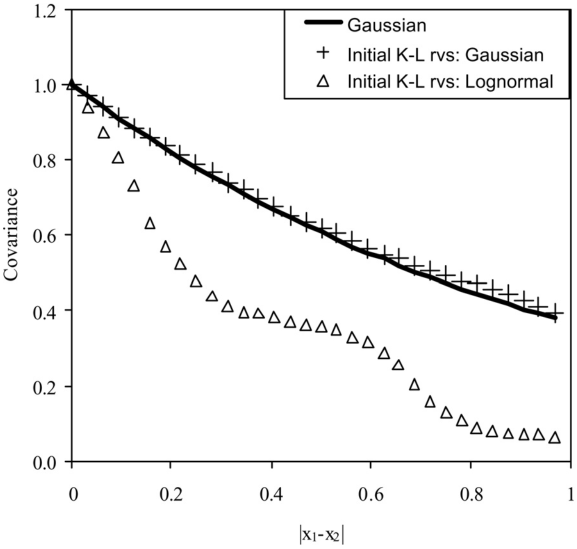

Fig. 1. Covariance comparison.

Figure 1 compares the covariance functions of the back-translated processes $\varpi_{b1}(x)$ and $\varpi_{b2}(x)$ with that of $\varpi_g(x)$. It is quite clear that $\varpi_1(x)$ is translation, but $\varpi_2(x)$ is not. A closed-form relationship between the rank (r) and product-moment correlation (ρ) exists for a bivariate Gaussian random vector [7]:

$$\rho = 2\sin\left(\frac{\pi}{6}r\right) \tag{13}$$

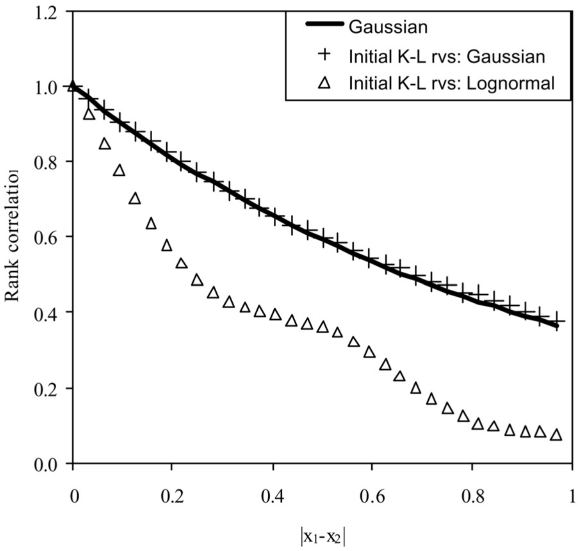

Fig. 2. Rank correlation comparison.

Figure 2 confirms that the rank correlations in $\varpi_1(x)$ do follow this relationship. The up-crossing rate shown in Fig. 3 further supports the hypothesis that $\varpi_{b1}(x)$ is translation.

4. Multivariate Gaussianity

The principal component method [8] based on the measure of skewness and kurtosis is used to test the multivariate normality for $\varpi_{b1}(x)$ and $\varpi_{b2}(x)$. For a

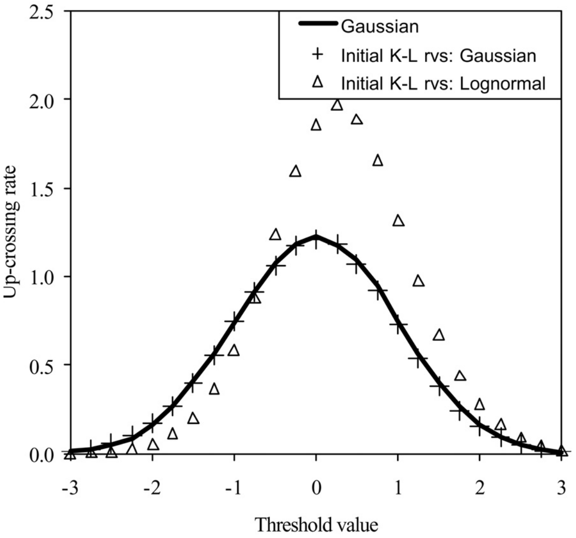

Fig. 3. Up-crossing rates comparison.

random p-vector ν with mean vector μ and covariance matrix Σ, the measure of skewness (β_{1p}^2) and kurtosis (β_{2p}), respectively, can be represented by:

$$\beta_{1p}^2 = p^{-1} \sum_{i=1}^{p} \left\{ E\left[\gamma_i'(\nu - \mu)\right]^3 \Big/ \lambda_i^{3/2} \right\}^2 \tag{14}$$

$$\beta_{2p} = p^{-1} \sum_{i=1}^{p} \left\{ E\left[\gamma_i'(\nu - \mu)\right]^4 \Big/ \lambda_i^2 \right\} \tag{15}$$

where λ_i are the eigenvalues and γ_i are eigenvectors of Σ, $i = 1, 2, \ldots, p$. If ν is a multivariate Gaussian random vector, $\beta_{1p}^2 = 0$ and $\beta_{2p} = 3$. For large sample size n, $A = (np/6)\beta_{1p}^2$ is chi-square distributed with p degrees of freedom and $B = (np/24)^{1/2}(\beta_{2p} - 3)$ follows a standard Gaussian distribution.

For $\varpi_{b1}(x)$, $\beta_{1p}^2 = 0.0011$ and $\beta_{2p} = 3.0114$, producing $A = 58.8437$ and $B = 1.3153$. For $\varpi_{b2}(x)$, $\beta_{1p}^2 = 1.6111$ and $\beta_{2p} = 11.2453$, producing $A = 85928$, and $B = 952.08$. Using the chi-square distribution with 32 degrees of freedom for the skewness statistics, the p-value for $\varpi_{b1}(x)$ is 0.0053 and $\varpi_{b2}(x)$ is 0. The corresponding p-value with respect to the kurtosis statistic for $\varpi_{b1}(x)$ is 0.1884 and $\varpi_{b2}(x)$ is 0. Given the results shown in Figs. 1 to 3, it is not surprising that $\varpi_{b2}(x)$ is not a Gaussian process. However, it is quite surprising that the skewness statistic p-value for $\varpi_{b1}(x)$ is small.

5. Conclusion

The non-Gaussian K-L expansion has the potential to simulate different non-Gaussian processes satisfying the same target covariance function and marginal distribution. This is realized by assigning different distributions to the K-L random variables during the start of the iteration process. Numerical results indicate that non-Gaussian K-L processes produced by assuming an initial Gaussian distribution for the K-L random variables appear to be translation processes. When the K-L random variables were assigned a lognormal distribution before the iteration procedure, the resulting process is clearly non-translation. Hence, it would appear that translation processes form a subset of K-L processes. In other words, the class of non-Gaussian K-L processes is larger and potentially capable of providing a better fit to observed data.

References

[1] Grigoriu, M. Simulation of non-Gaussian translation processes. Journal of Engineering Mechanics, ASCE 1988; 124(2):121–126.

[2] Huang HW. Simulation of stochastic processes using Karhunen-Loeve expansion. MEng thesis, National University of Singapore, 2004.

[3] Phoon KK, Huang HW, Quek ST. Simulation of strongly non-Gaussian processes using Karhunen-Loeve expansion. Ninth ASCE Specialty Conference on Probabilistic Mechanics and Structural Reliability, Albuquerque, New Mexico, 26–28 July, 2004 (CD-ROM).

[4] Phoon KK, Huang HW, Quek ST. Simulation of strongly non-Gaussian processes using Karhunen-Loève expansion (under review).

[5] Grigoriu, M. Applied non-Gaussian processes: examples, theory, simulation, linear random vibration, and MATLAB solutions. Englewood Cliffs, NJ: PTR Prentice Hall, 1995.

[6] Phoon KK, Huang SP, Quek ST. Implementation of Karhunen-Loeve expansion for simulation using a wavelet-Galerkin scheme. Probabilistic Engineering Mechanics 2002;17(3):293–303.

[7] Fackler, PL. Modeling interdependence: An approach to simulation and elicitation. American Journal of Agricultural Economics 1991;73:1091–1097.

[8] Srivastava, MS. A measure of skewness and kurtosis and a graphical method for assessing multivariate normality. Statistics & Probability Letters 1984;2(5):263–267.