Wind and Structures, Vol. 2, No. 2 (1999) 85-94
DOI: http://dx.doi.org/10.12989/was.1999.2.2.085

Parametric study based on synthetic realizations of EARPG(1)/UPS for simulation of extreme value statistics

Seung H. Seong†

School of Mechanical Engineering, Pusan National University, Pusan 609-735, Korea

Abstract. The EARPG(1)/UPS was first developed by Seong (1993) and has been tested for wind pressure time series simulations (Seong and Peterka 1993, 1997, 1998) to prove its excellent performance for generating non-Gaussian time series, in particular, with large amplitude sharp peaks. This paper presents a parametric study focused on simulation of extreme value statistics based on the synthetic realizations of the EARPG(1)/UPS. The method is shown to have a great capability to simulate a wide range of non-Gaussian statistic values and extreme value statistics with exact target sample power spectrum. The variation of skewed long tail in PDF and extreme value distribution are illustrated as a function of relevant parameters.

Key words: extreme value distribution; synthetic realizations; non-Gaussian statistics.

1. Introduction

Accurate prediction of extreme events is of primary importance in many engineering applications. The applications are usually environmental, for example, estimation of extreme wind speed for design wind load calculation, modeling annual maximum sea level to design offshore structures, and underwater acoustic signal detection (Grigoriu 1995, Wegman 1989). Also artificial time history generation with such extreme value characteristics are equally important for many engineering applications as well as the estimation study itself.

There have been many attempts for modeling and simulating non-Gaussian time series using stochastic models based on linear autoregressive process with Gaussian noise or non-Gaussian noise (Lechner 1993, Gaver 1980). They contain some critical limitations to simulating real non-Gaussian phenomena (Tong 1990). A FFT-based generation method, which has been widely used in engineering applications, are also limited basically to a Gaussian case (Shinozuka 1972). And for a non-Gaussian case, the method requires nonlinear transformation and an optimization procedure by iterations (Yamazaki 1988).

Even for some successful cases of non-Gaussian results, it is difficult to simulate the extreme statistics because they are very sensitive to extreme peak character of time series which is directly related to the tail behavior of parent distribution.

The EARPG(1)/UPS (First-order Exponential Auto Regressive Peak Generation Model and

† Assistant Professor

Uniform Phase Shift) has been developed and applied to wind pressure simulations. The method have a capability to simulate both spectral character and non-Gaussian character at the same time and independently without any optimization process as shown in Yamazaki (1988). The method, obtained by combining a stochastic model with non-Gaussian noise and the FFT method, has successfully overcome such limitations and difficulties that the traditional methods have. The success is mainly in the introduction of the preliminary stage where peak generation model produces preliminary signals and the use of unique property of Fourier phases that relax the requirement of both spectrum and non-Gaussian distribution matching with those of target time series. This study shows a capability of the method to simulate non-Gaussian time series focused on extreme value statistics by investigating the variation of probability density distribution as a function of associated parameters. First, the essential part of the digital generation of time series by EARPG(1)/UPS is reviewed.

2. Digital generation of parent pressure data

The desired pressure time series data can be generated by inverting the properly selected Fourier coefficients in the Fourier representation of time series $\{y_t,\ t=0,\ 1,\ \cdots,\ n-1\}$. A unit time interval is used for simplicity.

$$y_t = n^{-\frac{1}{2}} \sum_{k=0}^{n-1} \sqrt{I_k^m}\, e^{i\left(\frac{2\pi kt}{n} + \phi_k\right)} \tag{1}$$

The amplitude part of the coefficient $\sqrt{I_k^m}$ in Eq.(1) is obtained from I_k^m, the squared value of the ensemble-averaged periodogram of m records of time series of length n(even) $\{z_{tl},\ t=0,\ 1, 2,\ \cdots,\ n-1\}$, $l=1,\ \cdots,\ m$, from a target measured time series. In other words, I_k^m is a measure of the discrete amplitude in the power spectral density of the desired time series y_t.

$$I_k^m = \begin{cases} [m^{-1} \sum_{l=1}^{m} c_{kl}]^2 & \text{if } k=0 \\ m^{-1} \sum_{l=1}^{m} |c_{kl}|^2 & \text{if } k \neq 0 \end{cases} \tag{2}$$

$$c_{kl} = n^{-\frac{1}{2}} \sum_{t=0}^{n-1} z_{tl}\, e^{-i\frac{2\pi kt}{n}} \tag{3}$$

The phase part of the coefficient, ϕ_k, is constructed by the Fourier phase, ζ_k, of the time series $\{x_t,\ t=0, 1,\ \cdots,\ n-1\}$ generated from the peak generation model followed by the phase shift operation, $\delta_k(d)$.

$$\phi_k = \begin{cases} (\bar{c}_{0l}/|\bar{c}_{0l}|) & \text{if } k=0 \\ \zeta_k + \delta_k(d) & \text{if } k \neq 0 \end{cases} \tag{4}$$

where $\bar{c}_{0l} = m^{-1} \sum_{l=1}^{m} c_{0l}$. The phase ϕ_0 is excluded to preserve the original sign of the mean during change of the fluctuating features by the phase shift operation.

$$\zeta_k = \arctan \frac{\sum_{t=0}^{n-1} X_t \cos\left(\frac{2\pi kt}{n}\right)}{\sum_{t=0}^{n-1} X_t \sin\left(\frac{2\pi kt}{n}\right)} \tag{5, 6}$$

where the time series $\{X_t,\ t=0,1,\ \cdots,\ n-1\}$ is observations from the random process EARPG(1) in Eq. (7)

$$X_t = aX_{t-1} + \begin{cases} 0 \text{ with probability } b \\ E_t \text{ with probability } 1-b \end{cases} \qquad t=0,\ 1,\ 2\cdots \tag{7}$$

where E_t is an exponential random variable. For the details of probabilistic structure of this model, refer to Seong (1993). The EARPG(1) produces peak time series, whose character can be controlled by the parameters in the model, and a similar autocorrelation dependency to that of the target time series. The uniform phase shift operation, $\delta_k(d)$, is added to ζ_k to modify peak shape of y_t. The modification is induced through the phase, ζ_k, by inter-playing with the amplitude part, $\sqrt{I_k^m}$.

$$\delta_k(d) = \begin{cases} d \text{ for } 1 \le k \le \frac{n}{2} - 1 \\ -d \text{ for } \frac{n}{2} + 1 \le k \le n-1 \end{cases} \tag{8}$$

Finally synthetic time series are generated record by record by inverting the coefficients with the specific $\sqrt{I_k^m}$ and continuous records of θ_k from the realizations of the peak generation model EARPG(1). The I_k^m can be obtained from measured time series data or specific spectrum data. Then records of y_t have a non-Gaussian spiky character induced by the peak generation model through the phase. Each record of y_t has the same mean, RMS value, and power spectrum as those of the m-record-averaged target signal.

3. Synthetic signal as a function of parameters

The characteristics of the synthetic signal depend on three components: target spectrum, peak generation model, and phase shift.

$$y_t = y_t\,(\text{Target spectrum, Peak generation model, Phase shift}) \tag{9}$$

For the given target spectrum, the y_t in the Fourier representation can be regarded as a function of the parameters of the Fourier phase associated with the peak generation model and the phase shift. Thus this technique has great flexibility in generating various fluctuating features depending on the selection of the peak generation model and the phase shift and to parametrically control the overall statistical properties without changing the specific spectral character.

For the EARPG(1)/UPS, for instance, the final signal is defined by three parameters;

$$\begin{aligned} y_t &= y_t\,(\sqrt{I_k^m},\ \text{EARPG}(1),\ \text{UPS}) \\ &= y_t(a, b, d) \text{ for the given } \sqrt{I_k^m} \text{ from target signal or spectrum} \end{aligned}$$

The associated parameters affect the distributional property of time series while the amplitudes from the target spectrum determine the spectral character regardless of the parameters.

4. Parametric variation of PDF and extreme value distribution

The parametric change of PDF(Probability Density Function) and extreme value distribution based on realizations of EARPG(1)/UPS are investigated. As a target signal, an experimental pressure data in the wind tunnel is used which exhibits strong non-Gaussian distribution together with sharp spikes. The parameter a is chosen by the estimation procedure considering the autocorrelation structure of the target signal so as to satisfy the general requirement of the preliminary signal. For the fixed a, the effect of the parameters b and d is investigated on the statistical property variation of the simulated time series characterizing the non-Gaussian distribution. The investigation includes the probability density function and the extreme value distribution, the skewness and kurtosis of the parent pressure data, and the mode and dispersion of the extreme values. The skewness and kurtosis are defined by $S_3/S_2^{3/2}$ and S_4/S_2^2 respectively where $S_k=n^{-1}\sum_{t=1}^{n}(y_t-\bar{y})^k$ and $\bar{y}$ is a sample mean of time series $\{y_t, t=1, 2, \cdots, n\}$.

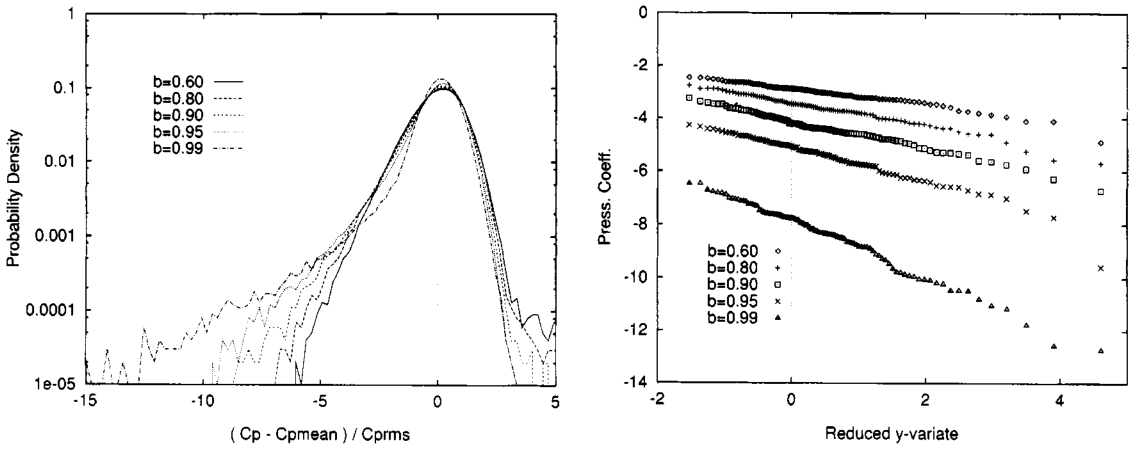
4.1. Effects of parameter b

The most dominant parameter which affects the overall probability distribution character in the simulated signal is the b parameter, which is the probability parameter in the model EARPG(1) to control the frequency of occurrence of large amplitude peaks in the preliminary signal.

In the synthetic signal, the parameter b controls not only the the frequency of occurrence of spikes but also their magnitude. This is due to the fact that the amplitudes of the Fourier coefficients remain unchanged while the phase part of the coefficients is variable through the parameters in the peak generation model. This fact results in the same RMS level of the final synthetic signal while the character of spikes and their frequency of occurrence are changed by the b parameter through the phase part of the coefficient. Thus the intensity of spikes is self-adjusted so that the final signal has the same RMS value.

Fig. 1(a) shows the variation of the PDF of the simulated signal using different b values. The large value of b corresponds to the highly intensive spike generation toward the negative direction. Consequently, the PDF has the narrow shape and a long tail in the negative direction. A significant deviation from normality is seen with increasing b values. This long tail in the PDF is similar in behavior to the actual pressure signal in the negative pressure zone near the corner or the roof edge of a building.

Fig. 1(b) shows the extreme value distribution on a Gumbel plot for different sets of parameters which correspond to the parameters used in Fig. 1(a). For each b value and with



(a) Probability density functions (b) Extreme value distributions

Fig. 1 Variations of distribution depending on parameter b for the simulation using different $b = 0.6$, 0.8, 0.9, 0.95, 0.99 for fixed $a = 0.975$, $d = -\frac{2}{3}\pi$

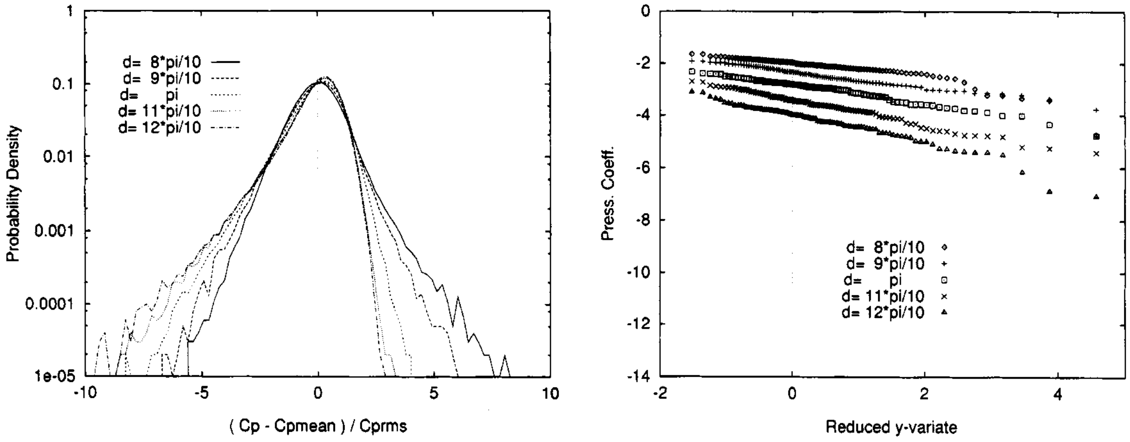
fixed other parameters, 96 records (96*10.2 seconds) of time series were produced. From each record (10.2 seconds, 4096 data), the smallest value was taken (the largest negative value) to obtain the distribution of extreme values. The mode of the extreme values, which approximately corresponds to the intercept at the zero of the reduced variate along the x-axis, can be gradually changed by increasing the b values. In addition, there is a slight dispersion change which corresponds to the slope of the line at the intercept. The large value of b, which produces large amplitudes and small numbers of peaks skewed in a negative direction, results in lowering the mode of the extreme values (largest negative values) and increasing the dispersion level.

4.2. Effect of parameter *d*

The parameter d, the amount of the uniform phase shift, controls the directionality of the time series. The preliminary signal was accompanied by unnatural shape distortions of the peaks. However, the proper uniform shift can produce a natural shape of spikes from the unnatural sawtooth shape peaks(Seong 1993).

Fig. 2(a) shows the variation of the probability density functions of the simulated signals using different d values for fixed values of other parameters. The gradual change of the parameter value produces a gradual shape change of the PDF. The cyclic change of d value across 2π produces one cycle of PDF shape change.

Fig. 2(b) shows the corresponding extreme value distribution in a Gumbel plot for a range of d values which produces spikes in the negative direction. The d parameter is less effective in changing the mode of extreme values than the b parameter. Its practical use is limited to a certain range within $0 \leq d \leq 2\pi$ because the phase shift d in a wide range distorts the peak shape and produces unnatural shapes of peaks.



(a) Probability density functions

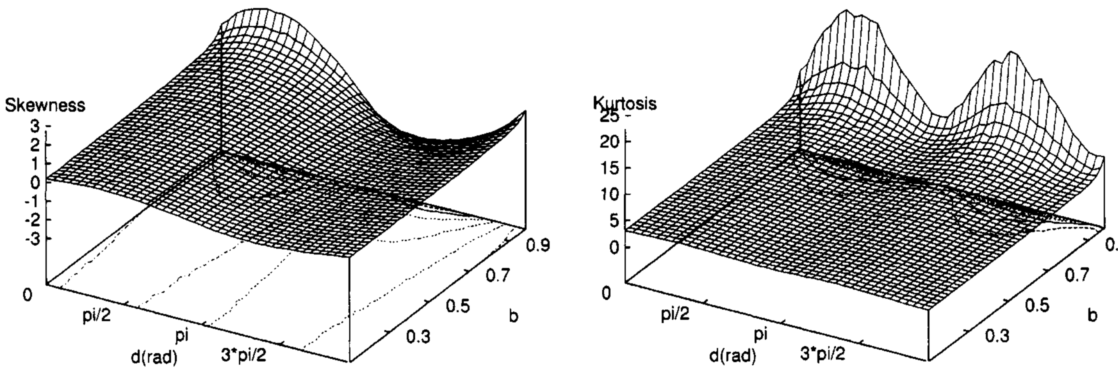
(b) Extreme value distributions

Fig. 2 Variations of distribution depending on parameter d for the simulation using $d=8\pi/10$, $9\pi/10$, π, $11\pi/10$, $12\pi/10$, for fixed $a=0.975$, $b=0.95$

4.3. Parametric variation of non-Gaussian statistics

Statistical quantities of the simulated pressure data are investigated as a function of the parameters b and d. The skewness and kurtosis of the parent pressure data are plotted in 3-dimensional plots as a function of the parameters b and d in Figs. 3(a), (b). The statistical quantity at each point on the surface is calculated by using 96 records of length 4096 of the simulated signal from EARPG(1) with a=0.975 and the different parameter values b and d. The figures exhibit the variation of the statistical quantities as a function of the parameters and further enable location of the useful range of the parameter value and the corresponding available range of variation of the statistics.

The skewness surface in Fig. 3(a) has a 2π-periodic and anti-symmetric shape about d=$17\pi/20$. The maximum positive or negative values of the skewness occurs approximately at the phase



(a) Skewness

(b) Kurtosis

Fig. 3 Variations of skewness and kurtosis of parent population data as a function of b and d

$d=7\pi/20$, $27\pi/20$, respectively, where most peaks are formed to produce maximum skewness. The phase angles to give the maximum skewness are dependent on the preliminary signal, specifically, on the peak shape within the signal. Roughly, the phase for the maximum skewness corresponds to the most natural shape of peaks. Once the optimum phase shift for the natural shape of peaks is fixed, the skewness depends on the b parameter which corresponds to a straight line parallel to the b parameter axis. Along this line, a wide variation of skewness is obtained by varying only the b parameter.

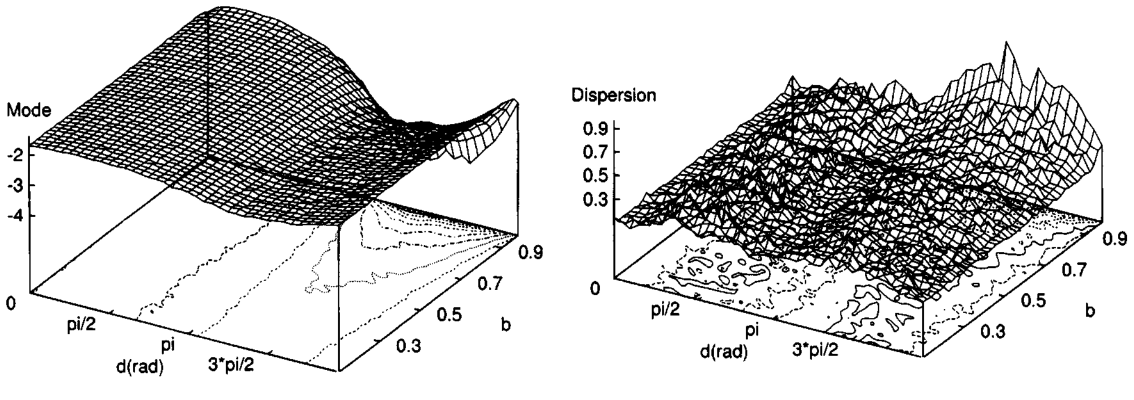
The kurtosis surface in Fig. 3(b) has a π periodicity and is approximately symmetric about $d=17\pi/20$, the same position as that of the skewness case. This is because the kurtosis is not related to the directionality of the skewness of a distribution. Thus the effect on the kurtosis of the uniform phase shift in the range $0<d<17\pi/20$ and $37\pi/20<d<2\pi$ is nearly the same as those in the range $17\pi/20<d<37\pi/20$. The maximum value of the kurtosis occurs approximately at the phase $d=7\pi/20$, $27\pi/20$ as for the case of the skewness. Similarly, the wide variation of the kurtosis value is obtained mainly by the b parameter.

5. Parametric variation of extreme value statistics

Variation of extreme values statistics for simulated pressure data are investigated as a function of the parameters b and d. The mode and dispersion of the selected extreme pressure are plotted in 3-dimensional plots as a function of the parameters b and d in Figs. 4(a) and 4(b). The surface for the mode of extreme values in Fig. 4(a) has a 2π periodicity but has no symmetry in the 2π range. This is because only the extreme values in one direction, e.g., the largest negative values in this case, are considered. Each point on the surface is calculated using 96 extremes where each one is selected from one record of length 4096. The mode value is estimated using the relations in Eqs. (11) and (12) under the assumption of the type-I extreme value distribution by

$$P(X<x)=F_X(x)=\exp\{-\exp[-(x-U)/s]\} \tag{10}$$

where s and U are the dispersion and mode of the extreme values.



(a) Mode (b) Dispersion

Fig. 4 Variations of extremes value statistics as a function of b and d

$$\mu = U + 0.5772s \tag{11}$$

$$\sigma = \frac{\pi}{\sqrt{6}} s \tag{12}$$

where μ and σ are the mean and standard deviation of the extreme values respectively. The maximum mode value occurs approximately at the phase d=$27\pi/20$ which gives the maximum skewness and kurtosis.

Fig. 4(b) for the dispersion of extreme values shows that the variation range of the dispersion value is small comparing the variation of the mode value in the same range of the parameter values and the large variance in the dispersion value, as the rough surface illustrates. However, the same trend of variation is seen, such as the existence of the phase position which gives maximum dispersion and the increasing dispersion value with increasing b.

In summary, the most useful range of the parameter values to simulate a wide range of the statistics appears to be the area bounded by $0.7<b<1$ and $4\pi/5<d<8\pi/5$. It is noted that, practically, the d parameter range is limited to a narrow range to avoid an unnatural shape of peaks. The position and useful range of d depends on the shape of peaks in the preliminary signal as noted earlier.

6. Discussion

6.1. Selection of parameter values

A parameter value selection for non-Gaussian statistics
is made based on simulation data and according to a target property of simulation. For the fixed value of the a parameter which is determined considering the autocorrelation of the target signal, the values of b and d, which are related to the non-Gaussian property, are selected by using the 3-dimensional surfaces in Figs. 3 and 4. The surface shape depends on the type of preliminary signal and the given target spectral density. In case of EARPG(1)/UPS, first the useful range of d is determined which gives the most natural shape of peaks. The b value is selected according to the desired major target character to be simulated. For instance, if the major target non-Gaussian property is skewness, the parameter b can be selected by investigating the variation of the skewness as in Fig. 3(a).

6.2. Repeatability of the simulation

Final comments on the simulation concern the repeatability of simulation, that is whether or not the selected parameter values can repeat the same quality of the signal by using available random number sources in the computer which has a pseudo-random sequence.

The most vulnerable statistical quantity to variation in the random source is the extreme value statistics since it is sensitive to the intensity of largest peaks and their frequency of occurrence, which are in turn controlled mainly by the probability parameter b. Thus even for the same value of the parameters, the statistics of extreme values, e.g., mode and dispersion of extreme values, may not converge rapidly to the same values within an allowed error limit. Thus an investigation of the convergence of the statistical quantities of the simulated signal

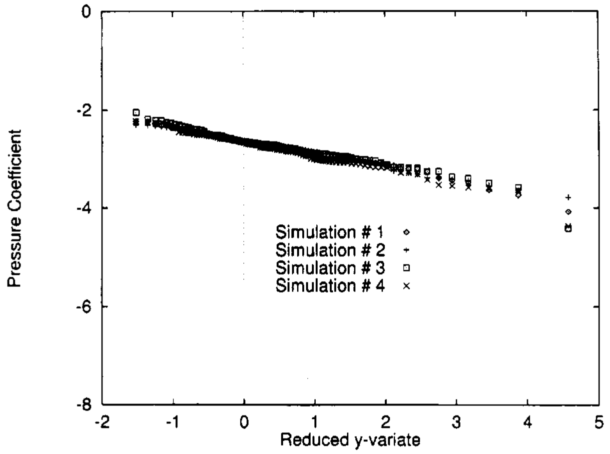


Fig. 5 Extreme value distributions from the four different sets of simulated pressure peaks using the same model parameter values; $a = 0.98$, $b = 0.9$, $d = -2\pi/3$

and its improvement, if necessary, are essential for practical applications.

Fig. 5 shows the extreme value distribution of the simulated signals in a Gumbel plot using four different sets of pseudo-random numbers and the same values of the associated parameters. The good convergence of the ordered extreme values in each of the four different simulations to the single straight line implies a good repeatability of the simulation. The small slope variability between simulations in Fig. 5 is consistent with the roughness in the dispersion surface of Fig. 4(b).

7. Conclusions

The EARPG(1)/UPS has been shown to have a good performance in simulating a wide range of variations in non-Gaussian statistics and consequently in extreme value statistics while preserving the specified spectral density characteristics. It is interesting to note that all the parent simulation data set showing various non-Gaussian distribution and extreme value character controlled by model parameters have identical spectral characteristics. This interesting fact implies that the spectral character can be separated from non-Gaussian time series synthesis by utilizing the Fourier representation and the unique property of the Fourier phase. The recomposition of the Fourier phase is a central idea of the method. Even if several interesting topics remain for investigation, which may be considered for future research, this method could be used as a most powerful tool for non-Gaussian signal generation and extreme value study encountered in practical engineering applications.

References

Gaver, D.P. and Lewis, P.A.W. (1980), "First-order autoregressive gamma sequences and point processes", *Advances on Applied Probability*, **12**, 727-745.

Grigoriu, M. (1995), *Applied Non-Gaussian Processes*, Prentice Hall PTR, Englewood Cliffs, N.J.

Lechner, J.A., Simiu, E., and Heckert, N.A. (1993), "Assessment of 'peaks over threshold' methods for estimating extreme value distribution tails", *Structural Safety*, **12**, 305-314.

Seong, S.H. (1993), "Digital synthesis of wind pressure fluctuations on building surfaces", *Ph.D Thesis*, Fluid Mechanics and Wind Engineering Program, Department of Civil Engineering, Colorado State University, Fort Collins, Colorado.

Seong, S.H. and Peterka, J.A. (1993), "Computer simulation of non-Gaussian wind pressure fluctuations", *Proc. of Seventh U.S. National Conf. on Wind Engineering*, Los Angeles, California, 623-632.

Seong, S.H. and Peterka, J.A. (1997), "Computer simulation of non-Gaussian multiple time series", *J. Wind Eng. Ind. Aerodyn.*, **72**, 95-105.

Seong, S.H. and Peterka, J.A. (1998), "Digital generation of surface-pressure fluctuations with spiky features", *J. Wind Eng. Ind. Aerodyn.*, **73**, 181-192.

Shinozuka, M. and Jan, C.-M. (1972), "Digital simulation of random processes and its applications", *Journal of Sound and Vibration*, **25**, 111-128.

Tong, H. (1990), *Non-Linear Time Series: A Dynamical System Approach*, Clarendon, Oxford.

Wegman, E.J., Schwartz, S.C., and Thomas, J.B. (Editors) (1989), *Topics in Non-Gaussian Signal Processing*, Springer-Verlag.

Yamazaki, F. and Shinozuka, M. (1988), "Digital generation of non-Gaussian stochastic fields", *J. Engrg. Mech.*, ASCE, **114**(7).

(Communicated by Chang-Koon Choi)