Simulating Gaussian Random Processes with Specified Spectra *

Donald B. Percival
Applied Physics Laboratory, HN-10
University of Washington
Seattle, WA 98195

Abstract

Abstract—We discuss the problem of generating realizations of length N from a Gaussian stationary process $\{Y_t\}$ with a specified spectral density function $S_Y(\cdot)$. We review three methods for generating the required realizations and consider their relative merits. In particular, we discuss an approximate frequency domain technique that is evidently used frequently in practice, but that has some potential pitfalls. We discuss extensions to this technique that allow it to be used to generate realizations from a power-law process with spectral density function similar to $S(f) = |f|^\alpha$ for $\alpha < 0$.

I. Introduction

Let $\{Y_t\}$ be a real-valued Gaussian stationary process with spectral density function (sdf) $S_Y(\cdot)$, autocorrelation sequence (acvs) $\{s_{\tau,Y}\}$ and zero mean. If we define the sampling time between observations Y_t and Y_{t+1} to be unity so that the Nyquist frequency is $\frac{1}{2}$, then the acvs is related to the sdf via the usual relationship

$$s_{\tau,Y} = \int_{-\frac{1}{2}}^{\frac{1}{2}} S_Y(f) e^{i2\pi f\tau}\, df, \quad \text{where } i \equiv \sqrt{-1}. \qquad (1)$$

A problem of considerable practical interest is to generate a sample of length N of this process (i.e., a realization of $Y_0, \ldots, Y_{N-1}$) on a digital computer by suitably transforming samples of a zero mean, unit variance Gaussian white noise process $\{W_t\}$. In this paper we discuss three methods for doing this: an exact time domain method that is valid for all sdf's (Section II), an exact frequency domain method that is valid only for some sdf's (Section III), and an approximate frequency domain method that can be used for all sdf's (Section IV). We place particular emphasis on generating time series from stationary and nonstationary power-law (long memory) processes (Sections V and VI). (The three methods we discuss here are certainly not the only ones that have been advocated in the literature—one important omission is an approximate time domain method based upon the class of autoregressive-moving average models.)

* Except for a few minor corrections made here, this paper appeared in *Computing Science and Statistics*, **24** (1992), pp. 534–538.

II. An Exact Time Domain Method

If the acvs $\{s_{\tau,Y}\}$ is readily known out to lag $N-1$, there are well-known time domain techniques for generating samples of $\{Y_t\}$ (see, for example, Franklin, 1965). Typically these involve an lower-upper Cholesky factorization of the inverse of the N-th order Toeplitz covariance matrix for $Y_0, \ldots, Y_{N-1}$ (see Demeure and Scharf, 1987, for a good review). This factorization can be accomplished using the Levinson-Durbin recursions and then used to generate the desired samples, as follows. Let $W_0, \ldots, W_{N-1}$ be a set of N independent and identically distributed Gaussian random variables (rv's) with zero mean and unit variance. With $Y_0 = \sigma_0 W_0$, we generate the $N-1$ remaining samples recursively via

$$Y_t = \sum_{j=1}^{t} \phi_{j,t} Y_{t-j} + W_t \sigma_t, \qquad t = 1, \ldots, N-1.$$

The σ_t's and $\phi_{j,t}$ are obtained by first setting $\sigma_0^2 = s_{0,Y}$ and then recursively computing for $t = 1, \ldots, N-1$

$$\begin{aligned} \phi_{t,t} &= \frac{s_{t,Y} - \sum_{j=1}^{t-1} \phi_{j,t-1} s_{t-j,Y}}{\sigma_{t-1}^2} \\ \phi_{j,t} &= \phi_{j,t-1} - \phi_{t,t}\phi_{t-j,t-1}, \qquad 1 \le j \le t-1 \\ \sigma_t^2 &= \sigma_{t-1}^2 \left(1 - \phi_{t,t}^2\right) \end{aligned}$$

(for $t = 1$ the summation in the first equation is taken to be 0, and the second equation is skipped).

There are two potential drawbacks to this exact method. First, once we have computed the σ_t's and $\phi_{j,t}$'s, the number of floating point operations needed to generate a sample of length N is $O(N^2)$. There are ways, however, of reducing this number to $O(N)$ if in fact the Toeplitz matrix possesses enough special structure (this is the case if, for example, $\{Y_t\}$ is an autoregressive-moving average process—see Kay, 1981, for details). Second, if we are given the sdf $S_Y(\cdot)$ instead of the acvs, we must first obtain the required $s_{\tau,Y}$'s. In principle this can always be done via numerical integration, but in practice this approach can be error-prone and time consuming.

III. An Exact Frequency Domain Method

Davies and Harte (1987) recently outlined a frequency domain technique for simulating $\{Y_t\}$ that makes use of a fast Fourier transform (fft) algorithm and hence requires only $O(N \log(N))$ operations. As these authors noted, their method is not completely general in that it can fail to work for some processes. The situations for which their method is applicable are easily described by a nonnegativity constraint. Their method has in fact appeared previously in the literature in the context of

simulating Gaussian moving average processes of order q, for which the nonnegativity constraint holds as long as N is greater than q (see Davis, Hagan, and Borgman, 1981, and the discussion of their work in Ripley, 1987).

Let M be any even positive integer (typically a power of 2), and define $f_j = \frac{j}{M}$. Let

$$\mathcal{S}_j \equiv \sum_{\tau=-\left(\frac{M}{2}-1\right)}^{\frac{M}{2}} s_{\tau,Y} e^{-i2\pi f_j \tau}, \quad 0 \le j \le \frac{M}{2}. \qquad (2)$$

Note that we can rewrite the above as

$$\mathcal{S}_j = \left(\sum_{\tau=0}^{\frac{M}{2}} s_{\tau,Y} e^{-i2\pi f_j \tau} + \sum_{\tau=\frac{M}{2}+1}^{M-1} s_{M-\tau,Y} e^{-i2\pi f_j \tau} \right),$$

so we can obtain the $\mathcal{S}_j$'s via the discrete Fourier transform of the following sequence of length M:

$$s_{0,Y}, s_{1,Y}, \ldots, s_{\frac{M}{2}-1,Y}, s_{\frac{M}{2},Y}, s_{\frac{M}{2}-1,Y}, s_{\frac{M}{2}-2,Y}, \ldots, s_{1,Y}.$$

We can also reexpress Equation (2) as

$$\mathcal{S}_j = s_{\frac{M}{2},Y}(-1)^j + \int_{-\frac{1}{2}}^{\frac{1}{2}} W(f_j - f) S_Y(f)\, df,$$

where

$$W(f) \equiv \frac{\sin((M-1)\pi f)}{\sin(\pi f)}.$$

Because $W(\cdot)$ oscillates between positive and negative values and because $s_{\frac{M}{2},Y}(-1)^j$ can be negative, it is possible that some of the $\mathcal{S}_j$'s are negative. Note, however, that, if $\{Y_t\}$ were a moving average process of order $q \le \frac{M}{2} - 1$ so that $s_{\tau,Y} = 0$ for $|\tau| \ge \frac{M}{2}$, then we would have $\mathcal{S}_j = S_Y(f_j)$ so that $\mathcal{S}_j \ge 0$. In order for the simulation method to work, we must impose the nonnegativity constraint that $\mathcal{S}_j \ge 0$ for $0 \le j \le \frac{M}{2}$.

Let $W_0, \ldots, W_{M-1}$ be a set of M independent and identically distributed Gaussian rv's with zero mean and unit variance. Define

$$\mathcal{V}_j \equiv \begin{cases} \sqrt{\mathcal{S}_0} W_0, & j = 0; \\ \sqrt{\frac{1}{2}\mathcal{S}_j}\,(W_{2j-1} + iW_{2j}), & 1 \le j < \frac{M}{2}; \\ \sqrt{\mathcal{S}_{\frac{M}{2}}} W_{M-1}, & j = \frac{M}{2}; \\ \mathcal{V}^*_{M-j}, & \frac{M}{2} < j \le M-1 \end{cases}$$

(the asterisk denotes complex conjugation). Note that

$$cov\{\mathcal{V}_j, \mathcal{V}_k\} = E\{\mathcal{V}_j^* \mathcal{V}_k\} = \begin{cases} \mathcal{S}_j, & \text{if } j = k; \\ 0, & \text{otherwise.} \end{cases}$$

We next define the process $\{V_t\}$ via

$$V_t \equiv \frac{1}{\sqrt{M}} \sum_{j=0}^{M-1} \mathcal{V}_j e^{-i2\pi f_j t}, \qquad t = 0, \ldots, M-1. \quad (3)$$

By construction, the process $\{V_t\}$ is real-valued. Because V_t is a linear combination of Gaussian rv's, the process is Gaussian. A straight-forward exercise shows that $\{V_t\}$ is a stationary process with zero mean and acvs $\{s_{\tau,V}\}$ given by

$$s_{\tau,V} = \frac{1}{M} \sum_{j=0}^{M-1} \mathcal{S}_j e^{i2\pi f_j \tau}. \qquad (4)$$

In contrast to $\{Y_t\}$, however, the stationary process $\{V_t\}$ is a harmonic process; i.e., it does not possess an sdf, but its spectral properties are given by an integrated spectrum that is a step function with steps at the $\pm f_j$'s. This fact implies that realizations of $\{V_t\}$ are periodic with period M, and hence so is its acvs $\{s_{\tau,V}\}$. Note that, if M is a power of 2, we can readily compute both $\{V_t\}$ and $\{s_{\tau,V}\}$ using a conventional fft algorithm.

By substituting the definition for $\mathcal{S}_j$ in Equation (2) into Equation (4) and interchanging the order of the two summations, we obtain

$$s_{\tau,V} = \frac{1}{M} \sum_{\rho=-\left(\frac{M}{2}-1\right)}^{\frac{M}{2}} s_{\rho,Y} \left(\sum_{j=0}^{M-1} e^{-i2\pi f_j (\tau - \rho)} \right).$$

From the result

$$\sum_{j=0}^{M-1} e^{-i2\pi f_j \eta} = \begin{cases} M, & \text{for } \eta = 0, \pm M, \pm 2M, \ldots; \\ 0, & \text{otherwise,} \end{cases}$$

we obtain

$$s_{\tau,V} = s_{\tau,Y} \text{ for all } |\tau| \le \frac{M}{2}.$$

Hence the statistical properties of $V_0, \ldots, V_{N-1}$ are identical to those of $Y_0, \ldots, Y_{N-1}$ if we set $N \le \frac{M}{2}$.

As is true for the exact time domain method, we might need to obtain the required $s_{\tau,Y}$'s via numerical integration if we are given the sdf $S_Y(\cdot)$ instead of the acvs. Also, because of the nonnegativity constraints on the $\mathcal{S}_j$'s, this method cannot be used for all stationary processes. As we noted previously, it does work for moving average processes of order $q \le \frac{M}{2} - 1$. Since every nonnegative lag window spectral estimate has an sdf corresponding to that of a high order moving average process, this exact frequency domain method is useful for simulating time series from such spectral estimates.

IV. An Approximate Frequency Domain Method

We consider here the construction of a zero mean Gaussian process $\{U_t\}$ whose acvs $\{s_{\tau,U}\}$ agrees—to a good approximation—with $\{s_{\tau,Y}\}$ out to lag $N-1$. To begin with, we make the assumption that the sdf $S_Y(\cdot)$ is continuous over $[-\frac{1}{2}, \frac{1}{2}]$ (we relax this restriction in the next section). Let M be any even integer greater than or equal to the desired sample size N. Let W_j, $j = 0, \ldots, M-1$, be a set of M independent and identically distributed Gaussian rv's with zero mean and unit variance. Let $f_j \equiv \frac{j}{M}$ as before. We define the process $\{U_t\}$ via

$$U_t \equiv \frac{1}{\sqrt{M}} \sum_{j=0}^{M-1} \mathcal{U}_j e^{-i2\pi f_j t}, \qquad t = 0, \ldots, M-1, \quad (5)$$

where

$$\mathcal{U}_j \equiv \begin{cases} \sqrt{S_Y(0)} W_0, & j = 0; \\ \sqrt{\frac{1}{2} S_Y(f_j)}\,(W_{2j-1} + iW_{2j}), & 1 \le j < \frac{M}{2}; \\ \sqrt{S_Y(\frac{1}{2})} W_{M-1}, & j = \frac{M}{2}; \\ \mathcal{U}^*_{M-j}, & \frac{M}{2} < j \le M-1. \end{cases}$$

By construction, $\{U_t\}$ is a real-valued Gaussian process. A straight-forward exercise shows that $\{U_t\}$ is a stationary process with zero mean and acvs $\{s_{\tau,U}\}$ given by

$$s_{\tau,U} = \frac{1}{M} \sum_{j=0}^{M-1} S(f_j) e^{i2\pi f_j \tau} \quad (6)$$

(note that this acvs can readily be computed using an fft). In contrast to $\{Y_t\}$, however, the stationary process $\{U_t\}$ is a harmonic process (as was the case with the V_t's in Equation (3)).

Because the right-hand side of Equation (6) can be regarded as a Riemann sum approximation to the integral of Equation (1), we have $s_{\tau,U} \approx s_{\tau,Y}$ for possibly some values of τ, but certainly not all: whereas $\{s_{\tau,U}\}$ is periodic, $\{s_{\tau,Y}\}$ must damp down to 0. Note that, if we let M be a power of 2, Equation (5) can be quickly computed using a conventional fft algorithm—realizations of $\{U_t\}$ of length M can thus be readily generated. By making M large enough, we can make $s_{\tau,U}$ arbitrarily close to $s_{\tau,Y}$ for $\tau = 0, \ldots, N-1$, and hence the simulations of $U_0, \ldots, U_{N-1}$ should have statistical properties that closely match those of $Y_0, \ldots, Y_{N-1}$.

This scheme was evidently first proposed by Thompson (1973), who advocated just letting $M = N$. This formulation is in common use in the physical sciences, perhaps due to the following result. Consider the periodogram of $U_0, \ldots, U_{M-1}$:

$$\hat{S}^{(p)}_U(f) \equiv \frac{1}{M} \left| \sum_{t=0}^{M-1} U_t e^{-i2\pi f t} \right|^2. \quad (7)$$

If M is a power of 2, we can evaluate the periodogram quickly over the grid of Fourier frequencies f_j using an fft algorithm. When $M = N$, we have $E\{\hat{S}^{(p)}_U(f_j)\} = S_Y(f_j)$ by construction—hence realizations of $\{U_t\}$ have a periodogram in good apparent agreement with the target sdf $S_Y(\cdot)$ at the Fourier frequencies. Unfortunately, the periodogram for $Y_0, \ldots, Y_{M-1}$ can be a badly biased estimator of $S_Y(\cdot)$, so this intuitively pleasing agreement is misleading. Figure 1 illustrates this important point.

Mitchell and McPherson (1981) also discussed this approximate scheme. To avoid the "$M = N$" problem discussed above, they advocated that M should be made larger than N commensurate with the "correlation length" of $\{Y_t\}$, but—beyond this brief statement—they did not provide explicit guidelines for selecting M relative to N. We can do so by noting the following useful measure of how well the $s_{\tau,U}$'s approximate the $s_{\tau,Y}$'s for $|\tau| \le N-1$. Let $s^{(M)}_{\tau,U}$ denote the value of $s_{\tau,U}$ generated using Equation (6) for a particular value of M. Define $S^{(M)}_U(\cdot)$ as the function whose Fourier coefficients are equal to $s^{(M)}_{\tau,U}$ for $|\tau| \le N-1$ and equal to $s_{\tau,Y}$ for $|\tau| \ge N$. Parseval's theorem then tells us that

$$SS(M) \equiv \sum_{\tau=-(N-1)}^{N-1} \left| s^{(M)}_{\tau,U} - s_{\tau,Y} \right|^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \left| S^{(M)}_U(f) - S_Y(f) \right|^2 df.$$

$SS(M)$ can be interpreted as the average squared difference between the sdf $S_Y(\cdot)$ and the function $S^{(M)}_U(\cdot)$. As M gets large, $SS(M)$ will decrease to zero. If we know the $s_{\tau,Y}$'s, we can readily compute $SS(M)$ and find a value of M such that $S^{(M)}_U(\cdot)$ is sufficiently close to $S_Y(\cdot)$ in terms of average squared difference. If the $s_{\tau,Y}$'s cannot be readily computed, we can increase M by, say, factors of 2 until

$$\sum_{\tau=-(N-1)}^{N-1} \left| s^{(M)}_{\tau,U} - s^{(2M)}_{\tau,U} \right|^2$$

is small, indicating that the effect of doubling M is small in that the average squared difference between $S^{(M)}_U(\cdot)$ and $S^{(2M)}_U(\cdot)$ is small. Figure 2 illustrates how increasing M yields a better approximation to $\{s_{\tau,Y}\}$.

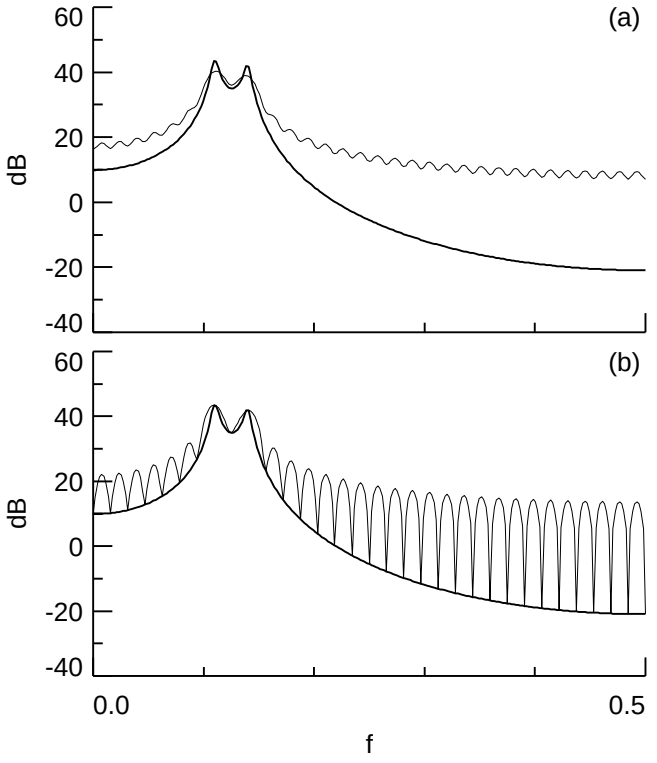

Figure 1. The thick curves in plots (a) and (b) show the true sdf $S_Y(\cdot)$ (on a decibel scale) for a particular stationary process $\{Y_t\}$. The thin bumpy curve in plot (a) shows the expected value of the periodogram for a sample of size $N = 64$ from this process. The thin bumpy curve in plot (b) shows the expected value of the periodogram $\hat{S}_U^{(p)}(\cdot)$ of Equation (7) for $M = N = 64$. If the statistical properties of $\{U_t\}$ closely matched those of $\{Y_t\}$, there would be good agreement between the two thin bumpy curves, but in fact they are substantially different. In particular, note that in plot (b) $E\{\hat{S}_U^{(p)}(f)\} = S_Y(f)$ for $f = f_j = \frac{j}{64}$ and $j = 0, \ldots, 32$, whereas $E\{\hat{S}_Y^{(p)}(\cdot)\}$ and $S_Y(\cdot)$ in plot (a) differ at some of these frequencies by more than 2 orders of magnitude (20 dB).

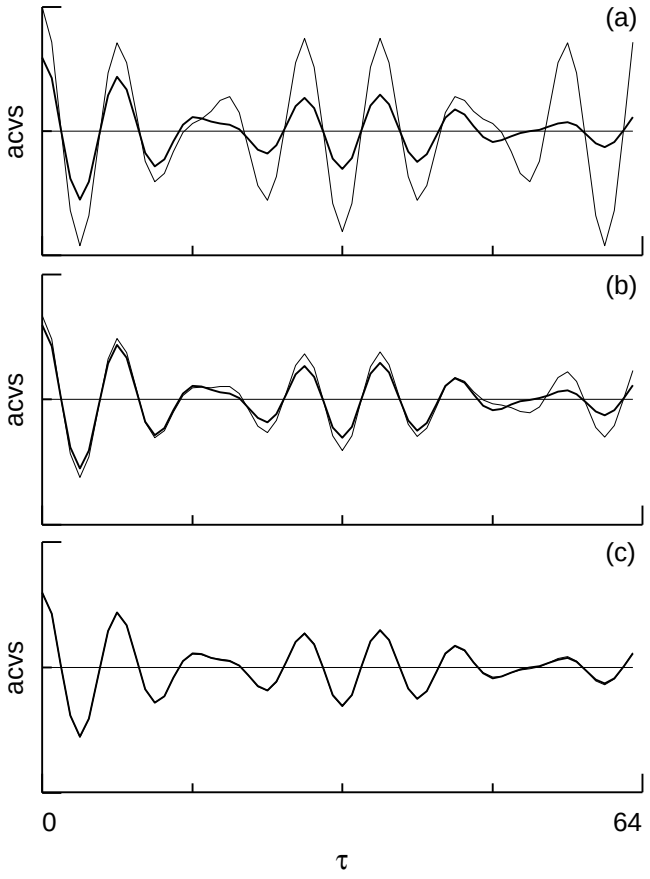

Figure 2. The thick curves in all three plot show the true acvs $\{s_{\tau,Y}\}$ for lags 0 to 63 for a particular stationary process $\{Y_t\}$ (in fact, the same process as was used in Figure 1). The thin curves show $\{s_{\tau,U}\}$ of Equation (6) for, from top to bottom, $M = 64, 128$ and 256—the thin curve in plot (c) agrees so well with the thick curve that they are visibly indistinguishable.

V. Stationary Power-Law Processes

Suppose now that $\{Y_t\}$ is a stationary power-law process, which—by definition—has an sdf given by

$$S_Y(f) = |f|^\alpha S_0(f), \qquad |f| \le \frac{1}{2}, \tag{8}$$

where $-1 < \alpha < 0$, and $S_0(\cdot)$ is strictly positive, continuous and has bounded variation. A specific example of such a process is a fractional difference process with sdf given by

$$S_Y(f) = (2|\sin(\pi f)|)^\alpha, \qquad |f| \le \frac{1}{2}.$$

This process also has an acvs $\{s_{\tau,Y}\}$ that can be computed recursively by a simple formula. Using these easily computed $s_{\tau,Y}$'s, Davies and Harte (1987) found that the exact frequency domain method can be used to simulate fractional difference processes. Since a similar simple formula for the acvs is not readily available for other stationary power-law processes, it is of some interest to see if we can modify the approximate frequency domain method for use here. The main difficulty is that this method requires that $S_Y(0)$ be finite, whereas Equation (8) tells us that $S_Y(0) = \infty$. We thus need to find a suitable replacement for $S_Y(0)$.

To do so, let us return momentarily to the setup of the previous section, namely, a process with an sdf that is continuous over $[-\frac{1}{2}, \frac{1}{2}]$. An easy exercise tells us that

$$\bar{U} \equiv \frac{1}{M} \sum_{t=0}^{M-1} U_t = \frac{\mathcal{U}_0}{\sqrt{M}} = \frac{\sqrt{S_Y(0)} W_0}{\sqrt{M}},$$

from which we obtain

$$var\{\bar{U}\} = \frac{S_Y(0)}{M} \approx var\{\bar{Y}\}, \text{ where } \bar{Y} = \frac{1}{M} \sum_{t=0}^{M-1} Y_t$$

(Priestley, 1981, p. 320). We can thus regard $S_Y(0)$ as an approximation to $M \cdot var\{\bar{Y}\}$. Künsch (1991) shows that, for a stationary power-law process $\{Y_t\}$,

$$var\{\bar{Y}\} \approx \frac{4S_0(0)\Gamma(1+\alpha)\sin(-\pi\alpha/2)}{(2\pi M)^{1+\alpha}\alpha(\alpha-1)} \equiv C_M.$$

This result suggests that we let $\mathcal{U}_0 = \sqrt{MC_M} W_0$ for stationary power-law processes. Limited tests to date indicate that this substitution is effective. More work needs to be done, however, to find the best $\mathcal{U}_0$ so that the statistical properties of $U_0, \ldots, U_{N-1}$ are as close as possible to those of $Y_0, \ldots, Y_{N-1}$.

VI. Nonstationary Power-Law Processes

Suppose now that $\{Y_t\}$ is a nonstationary power-law process with an "sdf" given by Equation (8) with $\alpha \leq -1$ (because $S_Y(\cdot)$ integrates to ∞, it is not a proper sdf). Processes such as these are common models in the physical sciences—some typical values for α are -1 ("flicker" noise), $-\frac{5}{3}$ (certain types of turbulence) and -2 ("random walk" noise). Yaglom (1958) developed a rigorous interpretation for the sdf $S_Y(\cdot)$ in terms of finite differences of $\{Y_t\}$. For example, if $-3 < \alpha \leq -1$, the first difference process $X_t \equiv Y_t - Y_{t-1}$ is a stationary process with sdf

$$S_X(f) = 4\sin^2(\pi f) S_Y(f), \qquad |f| \leq \frac{1}{2},$$

an equation which is used to *define* $S_Y(\cdot)$ (for $\alpha \leq -3$, we define $S_Y(\cdot)$ using an appropriate higher order difference).

Because we define $\{Y_t\}$ for $-3 < \alpha \leq -1$ in terms of its first difference process $\{X_t\}$, we can simulate $\{X_t\}$ and form cumulative sums to simulate $\{Y_t\}$. Here we merely note that we can use $S_Y(\cdot)$ directly in the approximate frequency domain method because

$$U_t - U_{t-1} = \frac{-i}{\sqrt{M}} \sum_{j=0}^{M-1} 2\sin(\pi f_j) \mathcal{U}_j i e^{-i2\pi f_j (t-\frac{1}{2})},$$

approximates $Y_t - Y_{t-1}$ properly (again, care must be taken at $f = 0$).

VII. Concluding Comments

We have described three methods for simulating a stationary Gaussian process $\{Y_t\}$ with a specified sdf $S_Y(\cdot)$. If the $s_{\tau,Y}$'s for the process are readily available, then the best choice is the exact frequency domain method *if* the $\mathcal{S}_j$'s of Equation (2) are in fact nonnegative. If the $s_{\tau,Y}$'s are not readily available or if one or more of the $\mathcal{S}_j$'s are negative, then—with care—the approximate frequency domain method can be useful. However, the "$M = N$" formulation of this method that is sometimes used should be avoided.

Acknowledgement

The author thanks the Office of Naval Research for support under contract number N00014–81–K–0095.

References

Davies, R. B. and Harte, D. S. (1987) Tests for Hurst Effect. *Biometrika*, 74, 95–101.

Davis, B. M., Hagan, R. and Borgman, L. E. (1981) A Program for the Finite Fourier Transform Simulation of Realizations from a One-Dimensional Random Function with Known Covariance. *Comp. and Geosci.*, 7, 199–206.

Demeure, C. J. and Scharf, L. L. (1987) Linear Statistical Models for Stationary Sequences and Related Algorithms for Cholesky Factorization of Toeplitz Matrices. *IEEE Trans. Acoust., Speech, Signal Processing*, 35, 29–42.

Franklin, J. N. (1965) Numerical Simulation of Stationary and Nonstationary Gaussian Random Processes. *SIAM Review*, 7, 68–80.

Kay, S. M. (1981) Efficient Generation of Colored Noise. *Proc. IEEE*, 69, 480–481.

Künsch, H. R. (1991) Dependence Among Observations: Consequences and Methods to Deal with It. In *Directions in Robust Statistics and Diagnostics, Part I*, edited by W. Stahel and S. Weisberg, New York: Springer-Verlag, 131–140.

Mitchell, R. L. and McPherson, D. A. (1981) Generating Nonstationary Random Sequences. *IEEE Trans. Aero., Elec. Sys.*, 17, 553–560.

Priestley, M. B. (1981) *Spectral Analysis and Time Series.* London: Academic Press.

Ripley, B. D. (1987) *Stochastic Simulation.* New York: John Wiley & Sons.

Thompson, R. (1973) Generation of Stochastic Processes with Given Spectrum. *Util. Math.*, 3, 127–137.

Yaglom, A. M. (1958) Correlation Theory of Processes with Random Stationary nth Increments. *Translations Amer. Math. Soc.*, 8, 87–141.