Generating random rough edges, surfaces, and volumes

Chris A. Mack

Lithoguru.com, 1605 Watchhill Rd., Austin, Texas 78703, USA (e-mail: chris@lithoguru.com)

Numerical methods of generating rough edges, surfaces, and volumes for subsequent simulations are commonly employed, but result in data with a variance that is downward biased from the desired value. Thus, it is highly desirable to quantify and to minimize this bias. Here, the degree of bias is determined through analytical derivations and numerical simulations as a function of the correlation length and the roughness exponent of several model power spectral density functions. The bias can be minimized by proper choice of grid size for a fixed number of data points, and this optimum grid size scales as the correlation length. The common approach of using a fixed grid size for such simulations leads to varying amounts of bias, which can easily be confounded with the physical effects being investigated. © 2013 Optical Society of America

OCIS codes: 290.5880, 120.6660, 000.4430, 000.5490.

1. Introduction

In many areas of modeling, there is often a need to numerically generate edges, surfaces, or volumes that are randomly rough with set statistical properties. In particular, one often desires a data set to represent a random edge, surface, or volume with a particular probability distribution (for example, Gaussian with a given mean and variance) and with a particular autocorrelation behavior. Rough edges are used for modeling the metrology of photoresist line-edge roughness [1,2] or as one-dimensional (1D) cuts of a rough surface [3,4]. Two-dimensional (2D) rough surfaces are often generated for use in optical and acoustical scattering simulations [5–7]. Three-dimensional (3D) rough volumes can be used for stochastic chemical reaction modeling and etching, including photoresist development [8,9].

While methods for generating random, correlated data sets (as described in the next section) are commonly used, these methods are statistically biased. They produce data with the desired variance (or standard deviation) only in the limit of very large data sets and small grid sizes. This statistical bias is a straightforward consequence of describing an infinite, continuous function by a discrete data set over a finite range. Further, the convergence of these methods to the desired power spectral density (PSD) is also a function of the number of points in the data set. This paper will determine the statistical biases present in one common method for generating random, correlated data sets. The resulting bias is a strong function of the parameters of the problem, such as the roughness exponent and correlation length, and can be minimized by proper choice of grid sizes. Further, the convergence properties of these numerical methods will also be investigated: when generating random, correlated data sets for use in simulation, how many iterations of such a simulation might be required? Without considering these statistical biases, it can be easy to confound the variation in data set bias (an artifact of the simulation) with the true physical phenomenon being investigated.

2. Generating a Random, Correlated Data Set

There are two main approaches for generating a random data set with a given probability distribution

and autocorrelation behavior. The *moving average* method was first described Naylor *et al.* in 1966 for the limited case of an exponential autocorrelation function, though they did not recognize how to generalize this approach to an arbitrary autocorrelation function [10]. The general moving average technique was first developed by Novarini and Caruthers, [5] though it has been reinvented several times since [6,11]. In this technique, the random, correlated data (z_i) are generated as a weighted sum of uncorrelated random numbers (η_{ij}) that have the desired probability distribution function:

$$z_i = \sum_{j=-M}^{M} w_j \eta_{ij}. \tag{1}$$

The weights (w_j) are determined by the Fourier transform of the square root of the desired PSD:

$$w_j = \text{FT}\{\sqrt{\text{PSD}(k)}\}. \tag{2}$$

This moving average method is often thought of as a process of spectral filtering or smoothing [11].

A second approach was developed by Thorsos [7], where the randomness is introduced into the PSD itself. This method has also been reinvented several times [12] and is sometimes referred to as the Monte Carlo spectral method. Since the Thorsos method will be the subject of this investigation, a more complete description of the method will be outlined here. Consider first the 1D case. Given N grid points with spacing Δx covering a distance $L = N\Delta x$, the data value at the point $x_n = n\Delta x$ is given by

$$z(x_n) = \mu_z + \frac{1}{L} \sum_{j=-N/2}^{N/2-1} F(f_j) e^{i2\pi f_j x_n}, \tag{3}$$

where this calculation is performed as the fast Fourier transform (FFT) of F on a grid of frequencies $f_j = j/L$. The target mean value of the random data is set to μ_z. The function F, in turn, is calculated from the amplitude of the PSD:

$$F(f_j) = \sqrt{\text{LPSD}(f_j)} \begin{cases} (\eta_1 + i\eta_2)/\sqrt{2}, & j \neq 0, \pm N/2 \\ \eta_1, & j = 0, \pm N/2 \end{cases}, \tag{4}$$

where η_1 and η_2 are two independent random numbers with mean of 0 and variance of 1 and with the desired probability distribution function. Since $z(x_n)$ must be real, Eq. (4) is used for $j \geq 0$ and the negative frequencies of F are obtained from a symmetry relationship: $F(f_{-j}) = F^*(f_j)$. Note also that the value of F at $j = -N/2$ is set to be real since the summation in Eq. (3) only goes to $N/2 - 1$ (corresponding to the N values needed for the FFT).

The Thorsos algorithm can easily be extended to two and three dimensions, so long as care is taken to properly produce the boundary conditions (a purely real random number is used at the origin and at the outer edges of the volume) and the symmetry to result in a purely real z. In two dimensions, the frequency used in the PSD is $f = \sqrt{f_x^2 + f_y^2}$, and a real z is guaranteed when

$$\begin{aligned} \text{Re}\{F(f_x, f_y)\} &= \text{Re}\{F(-f_x, -f_y)\}, \\ \text{Re}\{F(-f_x, f_y)\} &= \text{Re}\{F(f_x, -f_y)\}, \\ \text{Im}\{F(f_x, f_y)\} &= -\text{Im}\{F(-f_x, -f_y)\}, \\ \text{Im}\{F(-f_x, f_y)\} &= -\text{Im}\{F(f_x, -f_y)\}. \end{aligned} \tag{5}$$

In three dimensions, $f = \sqrt{f_x^2 + f_y^2 + f_z^2}$ and

$$\begin{aligned} \text{Re}\{F(f_x, f_y, f_z)\} &= \text{Re}\{F(-f_x, -f_y, -f_z)\}, \\ \text{Re}\{F(-f_x, f_y, f_z)\} &= \text{Re}\{F(f_x, -f_y, -f_z)\}, \\ \text{Re}\{F(f_x, -f_y, f_z)\} &= \text{Re}\{F(-f_x, f_y, -f_z)\}, \\ \text{Re}\{F(f_x, f_y, -f_z)\} &= \text{Re}\{F(-f_x, -f_y, f_z)\}, \\ \text{Im}\{F(f_x, f_y, f_z)\} &= -\text{Im}\{F(-f_x, -f_y, -f_z)\}, \\ \text{Im}\{F(-f_x, f_y, f_z)\} &= -\text{Im}\{F(f_x, -f_y, -f_z)\}, \\ \text{Im}\{F(f_x, -f_y, f_z)\} &= -\text{Im}\{F(-f_x, f_y, -f_z)\}, \\ \text{Im}\{F(f_x, f_y, -f_z)\} &= -\text{Im}\{F(-f_x, -f_y, f_z)\}. \end{aligned} \tag{6}$$

3. Bias in the Thorsos Method

To determine if the Thorsos method produces a data set that, on average, has the desired properties, let us first rearrange Eqs. (3) and (4) to give

$$z(x_n) = \mu_z + \frac{1}{\sqrt{L}} \sum_{j=-N/2}^{N/2-1} \gamma_j \sqrt{\text{PSD}(f_j)} e^{i2\pi j(n/N)}, \tag{7}$$

where

$$\gamma_j = \begin{cases} (\eta_1 + i\eta_2)/\sqrt{2}, & j \neq 0, -N/2 \\ \eta_1, & j = 0, -N/2 \end{cases}. \tag{8}$$

Since each of the random numbers is independent with zero mean, it is easy to show that

$$\langle z \rangle = \mu_z, \tag{9}$$

where $\langle \ldots \rangle$ means an average over many trials. Thus, the mean is unbiased (and independent of the physical parameters of the PSD, such as the correlation length, as well as the numerical parameter Δx).

The data set variance can be obtained by considering

$$\langle (z(x_n) - \mu_z)^2 \rangle = \frac{1}{L} \sum_{j=-N/2}^{N/2-1} \sum_{k=-N/2}^{N/2-1} \langle \gamma_j \gamma_k \rangle$$
$$\times \sqrt{\mathrm{PSD}(f_k)\mathrm{PSD}(f_j)} e^{i2\pi(j+k)(n/N)}. \quad (10)$$

Since each of the random numbers is independent (with the exception of the symmetry requirements), $\langle \gamma_j \gamma_k \rangle = 0$ when $j \neq \pm k$. Taking the symmetry conditions carefully into account, the case of $j = \pm k$ can be evaluated, giving

$$\langle (z - \mu_z)^2 \rangle = \frac{1}{L} \sum_{j=-N/2}^{N/2-1} \mathrm{PSD}(f_j). \quad (11)$$

Likewise, for the 2D case, assuming an $N \times N$ array with $\Delta x = \Delta y$,

$$\langle (z - \mu_z)^2 \rangle = \frac{1}{L^2} \sum_{j=-N/2}^{N/2-1} \sum_{k=-N/2}^{N/2-1} \mathrm{PSD}(f_j, f_k). \quad (12)$$

The 3D case is obtained in the same way.

The average variance of a data set is not given by Eq. (11) or Eq. (12), however, since the μ_z will be replaced by $\bar{z}$, the data set mean. This is equivalent to losing information about the zero frequency of the PSD, since a numerical calculation of the PSD from a data set will give $\mathrm{PSD}(0) = 0$ unless the population mean is known *a priori*. Thus, the average of the data variance over many trials will be (for the 1D case)

$$\langle (z - \bar{z})^2 \rangle = \frac{1}{L} \sum_{j=-N/2}^{N/2-1} \mathrm{PSD}(f_j) - \frac{\mathrm{PSD}(0)}{L}. \quad (13)$$

Also, the standard formula for calculating the standard deviation from a sample will divide by $N - 1$ rather than N. Thus, the average variance for a sample will be

$$\langle \sigma_{\mathrm{sample}}^2 \rangle = \langle (z - \bar{z})^2 \rangle \left(\frac{N}{N-1} \right)$$
$$\approx \left(\frac{1}{L} \sum_{j=-N/2}^{N/2-1} \mathrm{PSD}(f_j) - \frac{\mathrm{PSD}(0)}{L} \right) \left(1 + \frac{1}{N} \right). \quad (14)$$

The true (or target) variance is the area under the PSD curve, by Parseval's theorem:

$$\sigma^2 = \int_{-\infty}^{\infty} \mathrm{PSD}(f) \mathrm{d}f. \quad (15)$$

Comparing Eqs. (14) and (15), there are four differences that will systematically bias the variance of the data set compared to the desired input variance:

the approximation of the continuous integral with a discrete sum, the loss of high-frequency information (above the Nyquist sampling frequency), the loss of the zero frequency information, and the $1/N$ term. Each of these factors will now be explored in turn.

In order to quantify the bias in the data set variance, a specific form of the PSD will be used. A very common autocorrelation function ($\tilde{R}$) used in many simulation studies is the stretched/ compressed exponential:

$$\tilde{R}(x) = e^{-(x/\xi)^{2\alpha}}, \quad (16)$$

where α is called the roughness (or Hurst) exponent and the correlation length is ξ. One common value for the roughness exponent is $\alpha = 0.5$, corresponding to an exponential autocorrelation function. Another common autocorrelation function is the Gaussian, with a roughness exponent of $\alpha = 1.0$. For an autocovariance R these two functions are, along with their 1D PSDs [13],

$$\alpha = 0.5 : R(x) = \sigma^2 e^{-|x|/\xi}, \quad \mathrm{PSD}(f) = \frac{2\sigma^2 \xi}{1 + (2\pi f \xi)^2},$$
$$\alpha = 1.0 : R(x) = \sigma^2 e^{-(x/\xi)^2}, \quad \mathrm{PSD}(f) = \sqrt{\pi}\sigma^2 \xi e^{-(\pi f \xi)^2}, \quad (17)$$

where the PSD is calculated as the Fourier transform of the autocovariance (and the autocorrelation is simply the autocovariance normalized by dividing by the variance).

Consider first the 1D case with $\alpha = 0.5$. The error due to the discrete approximation to the integral (ε_d) will be

$$\varepsilon_d = \int_{-\infty}^{\infty} \mathrm{PSD}(f) \mathrm{d}f - \frac{1}{L} \sum_{j=-\infty}^{\infty} \mathrm{PSD}(f_j). \quad (18)$$

The infinite sum of this PSD has an analytical solution, giving a hyperbolic cotangent, so that

$$\varepsilon_d = \sigma^2 \left[1 - \coth\left(\frac{L}{2\xi} \right) \right]. \quad (19)$$

When $L \gg \xi$, the hyperbolic cotangent can be approximated with an asymptotic series about infinity to give

$$\varepsilon_d \approx -2\sigma^2 e^{-L/\xi}. \quad (20)$$

As we shall see below, this error component is small compared to the others.

The impact of the loss of the high-frequency terms from the summation can be estimated by defining the error term $\varepsilon_{\mathrm{hi}}$:

$$\varepsilon_{\text{hi}} = \int_{-\infty}^{\infty} \text{PSD}(f)df - \int_{-f_N}^{f_N} \text{PSD}(f)df$$

$$= 2\int_{f_N}^{\infty} \text{PSD}(f)df, \tag{21}$$

where the Nyquist frequency is $f_N = 1/(2\Delta x)$. For the 1D $\alpha = 0.5$ PSD being used here, this gives

$$\varepsilon_{\text{hi}} = \sigma^2\left[1 - \frac{2}{\pi}\tan^{-1}\left(\frac{\pi\xi}{\Delta x}\right)\right]. \tag{22}$$

When $\xi \gg \Delta x$, a Taylor-series expansion of the inverse tangent gives

$$\varepsilon_{\text{hi}} \approx \sigma^2\left(\frac{2}{\pi^2}\right)\left(\frac{\Delta x}{\xi}\right). \tag{23}$$

Finally, the error term caused by the loss of zero frequency information, ε_{lo}, will be

$$\varepsilon_{\text{lo}} = \frac{\text{PSD}(0)}{L} = \sigma^2\left(\frac{2\xi}{L}\right). \tag{24}$$

Combining these three error terms together, and including the $(1 + 1/N)$ multiplicative factor,

$$\langle\sigma_{\text{sample}}^2\rangle \approx \sigma^2\left[1 + 2e^{-L/\xi} - \left(\frac{2}{\pi^2}\right)\left(\frac{\Delta x}{\xi}\right) - 2\left(\frac{\xi}{L}\right) + \frac{1}{N}\right]. \tag{25}$$

In general, $-\varepsilon_d$ will be much smaller than ε_{lo} and can be ignored. Also, when $\Delta x < \xi$, the $1/N$ term will be small compared to ξ/L. Thus, the data set will have a variance that is biased lower than the desired value used in the Thorsos method to generate that data set. It is worth noting that the bias arises not from some deficiency in the Thorsos method *per se*, but from the act of representing a continuous function of infinite domain by a discrete function of finite domain.

The above derivation considered only the 1D case with $\alpha = 0.5$. Similar derivations can be made for the 2D and 3D cases, as well as for $\alpha = 1.0$. Defining the relative bias as

$$\varepsilon_{\text{rel}} = \frac{\sigma^2 - \langle\sigma_{\text{sample}}^2\rangle}{\sigma^2}, \tag{26}$$

the results of these derivations (ignoring the very small ε_d terms) are shown in Table 1, where

Table 1. Relative Bias in the Variance (ε_{rel}) Produced by the Thorsos Method

	$\alpha = 0.5$	$\alpha = 1.0$
1D	$\left(\frac{2}{\pi^2}\right)\left(\frac{\Delta x}{\xi}\right) + 2\left(\frac{\xi}{L}\right) - \frac{1}{N}$	$\left(\frac{2}{\pi^{3/2}}\right)\left(\frac{\Delta x}{\xi}\right)e^{-\left(\frac{\pi\xi}{2\Delta x}\right)^2} + \sqrt{\pi}\left(\frac{\xi}{L}\right) - \frac{1}{N}$
2D	$\left(\frac{2\sqrt{2}}{\pi^2}\right)\left(\frac{\Delta x}{\xi}\right) + 2\pi\left(\frac{\xi}{L}\right)^2 - \frac{1}{N^2}$	$\left(\frac{4}{\pi^{3/2}}\right)\left(\frac{\Delta x}{\xi}\right)e^{-\left(\frac{\pi\xi}{2\Delta x}\right)^2} + \pi\left(\frac{\xi}{L}\right)^2 - \frac{1}{N^2}$
3D	$\left(\frac{2+\sqrt{2}}{\pi^2}\right)\left(\frac{\Delta x}{\xi}\right) + 8\pi\left(\frac{\xi}{L}\right)^3 - \frac{1}{N^3}$	$\left(\frac{6}{\pi^{3/2}}\right)\left(\frac{\Delta x}{\xi}\right)e^{-\left(\frac{\pi\xi}{2\Delta x}\right)^2} + \pi^{3/2}\left(\frac{\xi}{L}\right)^3 - \frac{1}{N^3}$

approximations were made assuming $\Delta x < \xi < L$. Looking at all these expressions, it is clear that the bias can be minimized by choosing a grid size much smaller than the correlation length, and a domain size much larger than the correlation length. Thus, as one might expect, the correlation length provides the length scale with which the numerical impact of the artificial grid and domain sizes is compared.

For the $\alpha = 1.0$ case (the Gaussian autocorrelation function), the error due to the discrete nature of the sum is even smaller than for $\alpha = 0.5$:

$$\varepsilon_d \approx -2\sigma^2 e^{-(L/\xi)^2}. \tag{27}$$

Also, the very fast falloff of this PSD with frequency means that the high-frequency error term is very small. For the 1D case,

$$\varepsilon_{\text{hi}} = \sigma^2\text{erfc}\left(\frac{\pi\xi}{2\Delta x}\right) \approx \left(\frac{2}{\pi^{3/2}}\right)\left(\frac{\Delta x}{\xi}\right)e^{-\left(\frac{\pi\xi}{2\Delta x}\right)^2}. \tag{28}$$

For $\Delta x = \xi/2$, we find that $\varepsilon_{\text{hi}}/\sigma^2$ is less than 10^{-5}. Thus, it is the low-frequency term that dominates the bias for the Gaussian autocorrelation function case ($\alpha = 1.0$).

To test the validity of these expressions, random correlated rough edges, surfaces, and volumes were generated with the $\alpha = 0.5$ PSD using the Thorsos method, and their variances were calculated directly from the data sets. For all the simulations used here, the sizes of the data sets were N, $N \times N$, or $N \times N \times N$. For a fixed grid size of 1 nm, the number of grids N and the correlation length ξ were varied and a data set variance was calculated for each trial. Variances for many trials were then averaged. The results are shown in Figs. 1–3, including comparisons with the derived expressions in Table 1.

As can be seen from the figures, the approximate expressions for the data set variance from Table 1 match very well to the numerical output of the
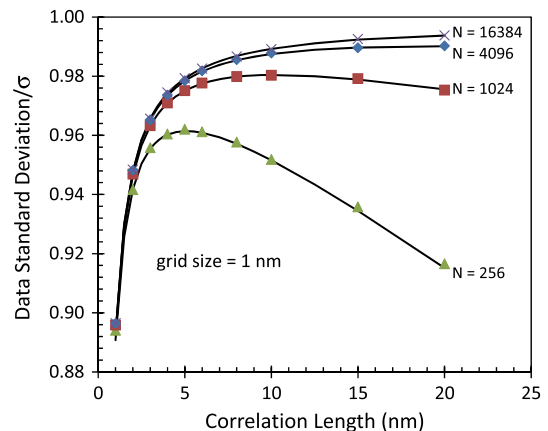

Fig. 1. (Color online) For the 1D case, $\alpha = 0.5$, comparison of numerical simulations using the Thorsos method (symbols) to the predictions of the relevant equation from Table 1 (solid curves).
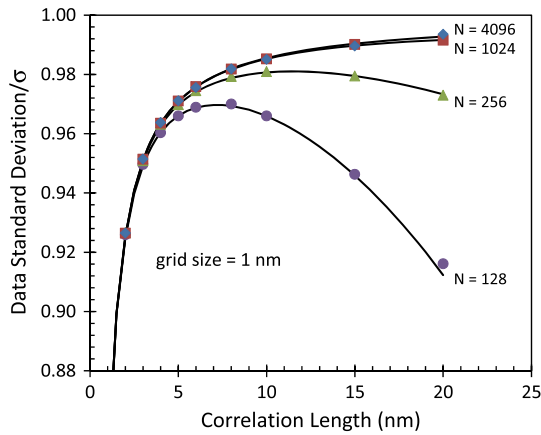
Fig. 2. (Color online) For the 2D case, $\alpha = 0.5$, comparison of numerical simulations using the Thorsos method (symbols) to the predictions of the relevant equation from Table 1 (solid curves).
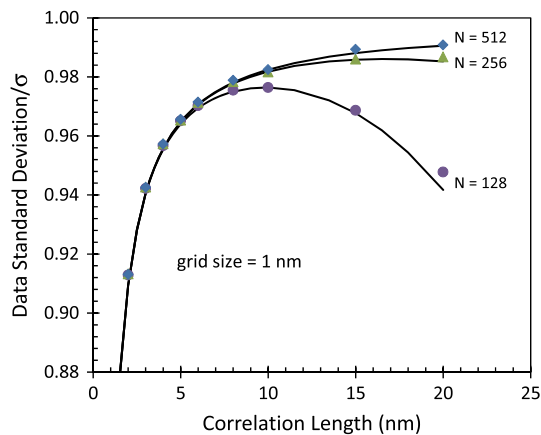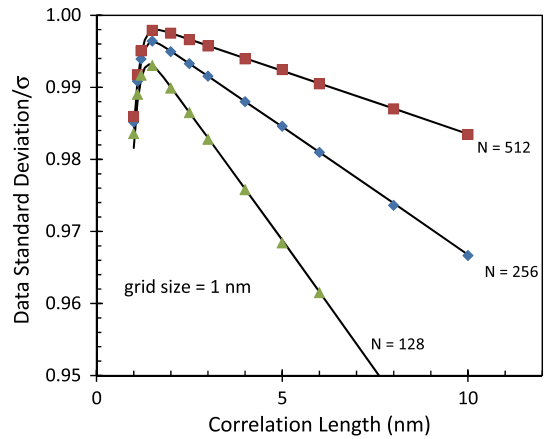

Fig. 4. (Color online) For the 1D case, $\alpha = 1.0$, comparison of numerical simulations using the Thorsos method (symbols) to the predictions of the relevant equation from Table 1 (solid curves).

Thorsos method. Similar results were also obtained for the $\alpha = 1.0$ case (as shown in Fig. 4 for the 1D case). Next, these expressions will be used to find parameter settings that minimize the bias inherent in this method.

4. Optimizing Simulation Parameters for Minimum Bias

Bias in the data set variance occurs when the correlation length is too close to the grid size (Δx) or too close to the simulation domain size (L). Thus, there will be an optimum setting, corresponding to the peak of the curves in Figs. 1–4, where bias is minimized. Clearly, N should be made as large as possible to minimize bias. But N is generally limited by computational constraints: available computer memory, or the time one is willing to wait to complete the simulations. Thus, I will assume that N is fixed at its maximum practical value. (I will also assume that N is a power of two, due to the ubiquitous use of the FFT in such calculations.) Further, the correlation length ξ will be set by the problem need. Thus, the

question will be, what grid size should be used? Note that the choice of grid size also determines the simulation domain size since $L = N\Delta x$.

The error terms in Table 1 are minimized when the grid size is chosen according to the values specified in Table 2. Thus, given values for ξ and N, picking a simulation grid size for each problem equal to the values calculated from Table 2 will result in a minimum bias in the resulting variance of the rough edge, surface, or volume. The resulting minimum biases are shown in Table 3 (ignoring the $1/N$ terms). Some example calculations are presented in Table 4.

The results from Tables 2–4 show the important role of the roughness exponent α on bias in the Thorsos method. For $\alpha = 0.5$, the high-frequency error component is significant, requiring a grid size much smaller than the correlation length. For example, for a 1D simulation with $N = 1024$, the optimum grid size is about $\xi/10$. But for the equivalent $\alpha = 1.0$ case, the optimum grid size is closer to

Fig. 3. (Color online) For the 3D case, $\alpha = 0.5$, comparison of numerical simulations using the Thorsos method (symbols) to the predictions of the relevant equation from Table 1 (solid curves).

Table 2. Simulation Grid Size that Minimizes the Relative Bias in the Variance Produced by the Thorsos Method, for a Given ξ and N

	$\alpha = 0.5$	$\alpha = 1.0$
1D	$\Delta x = \dfrac{\pi\xi}{\sqrt{N}}$	$\Delta x \approx \dfrac{\pi\xi}{2\sqrt{\ln(N)}}$
2D	$\Delta x = 2^{1/6}\dfrac{\pi\xi}{N^{2/3}} \approx 1.1\dfrac{\pi\xi}{N^{2/3}}$	$\Delta x \approx \dfrac{\pi\xi}{2\sqrt{2\ln(N)}}$
3D	$\Delta x = \sqrt[4]{\dfrac{24}{\pi(2+\sqrt{2})}}\dfrac{\pi\xi}{N^{3/4}} \approx 1.2\dfrac{\pi\xi}{N^{3/4}}$	$\Delta x \approx \dfrac{\pi\xi}{2\sqrt{3\ln(N)}}$

Table 3. (Approximate) Minimum Possible Relative Bias in the Variance Produced by the Thorsos Method, for a Given N, When the Grid size is Set as in Table 2

	$\alpha = 0.5$	$\alpha = 1.0$
1D	$\left(\dfrac{4}{\pi}\right)\dfrac{1}{\sqrt{N}} \approx \dfrac{1.27}{\sqrt{N}}$	$\dfrac{1.2\sqrt{\ln(N)}-1}{N}$
2D	$\left(\dfrac{4\sqrt[3]{2}}{\pi}\right)\dfrac{1}{N^{2/3}} \approx \dfrac{1.60}{N^{2/3}}$	$\dfrac{2.5\ln(N)-1}{N^2}$
3D	$\dfrac{1.77}{N^{3/4}}$	$\dfrac{6[\ln(N)]^{3/2}-1}{N^3}$

$\xi/2$. This larger grid size results in a larger simulation domain size (L), and thus smaller low-frequency error. The overall bias is thus much lower. Overall, when $\alpha = 1.0$, setting $\Delta x = \xi/2$ results is very low bias in the resulting data set variance under most conditions. When $\alpha = 0.5$, much more careful attention to grid size is required. Interestingly, the lessons learned here apply to evaluation of experimental roughness data as well [14].

5. Alternate PSD Formulation

The results above show the importance of the roughness exponent α on the biases inherent in the Thorsos method of generating rough edges, surfaces, and volumes. The definition of the roughness exponent used in Eq. (16), however, is not the only one. Palasantzas [15] suggested a PSD that has been widely used (shown here in the slightly modified form that has become more common):

$$\text{PSD}(f) = \frac{\text{PSD}(0)}{[1 + (2\pi f \xi)^2]^{H+d/2}}, \tag{29}$$

where H plays the role of the Hurst (roughness) exponent, d is the dimensionality of the problem, and PSD(0) is adjusted to give the desired variance [for example, via Eq. (15) for $d = 1$]:

$$1D: \text{PSD}(0) = 2\sigma^2 \xi \left(\frac{\sqrt{\pi}\Gamma\left(H + \frac{1}{2}\right)}{\Gamma(H)} \right),$$

$$2D: \text{PSD}(0) = 2\pi\sigma^2 \xi^2 (2H),$$

$$3D: \text{PSD}(0) = 8\pi\sigma^2 \xi^3 \left(\frac{\sqrt{\pi}\Gamma\left(H + \frac{3}{2}\right)}{\Gamma(H)} \right). \tag{30}$$

It is important to note that when $H = 0.5$, the resulting PSDs are identical to the $\alpha = 0.5$ PSDs [for example, shown in Eq. (17) for the 1D case]. However, the $H = 1.0$ case does not correspond to the $\alpha = 1.0$ case.

Using this PSD, the bias in variance for the Thorsos method can be determined, just as in the previous sections:

$$1D: \varepsilon_{\text{rel}} \approx \left(\frac{\Gamma(H + \frac{1}{2})}{\sqrt{\pi}\Gamma(H + 1)} \right) \left(\frac{\Delta x}{\pi \xi} \right)^{2H} + 2 \left(\frac{\sqrt{\pi}\Gamma\left(H + \frac{1}{2}\right)}{\Gamma(H)} \right)$$

$$\times \left(\frac{\xi}{L} \right) - \frac{1}{N}$$

$$2D: \varepsilon_{\text{rel}} \approx \left(\frac{2\sqrt{2}}{\pi} \right) \left(\frac{\Delta x}{\pi \xi} \right)^{2H} + 2\pi(2H) \left(\frac{\xi}{L} \right)^2 - \frac{1}{N^2}$$

$$3D: \varepsilon_{\text{rel}} \approx \left(\frac{2 + \sqrt{2}}{\pi} \right) \left(\frac{\Delta x}{\pi \xi} \right)^{2H} + 8\pi \left(\frac{\sqrt{\pi}\Gamma\left(H + \frac{3}{2}\right)}{\Gamma(H)} \right)$$

$$\times \left(\frac{\xi}{L} \right)^3 - \frac{1}{N^3}. \tag{31}$$

For the 1D case, the minimum bias occurs when

$$\Delta x \approx \frac{\pi \xi}{N^{1/(1+2H)}}. \tag{32}$$

At this optimum grid size, the minimum bias in the variance in 1D is

$$\varepsilon_{\text{rel,min}} \approx 2 \left(\frac{\Gamma\left(H + \frac{3}{2}\right)}{\sqrt{\pi}\Gamma(H + 1)} \right) \left(\frac{1}{N^{2H/(1+2H)}} \right). \tag{33}$$

Similar results can be derived for the 2D and 3D cases as well. Figure 5 compares the predictions of Eq. (31) to 1D simulations for various values of the roughness exponent H.

6. Convergence of the Data Set PSD

While the discussion above shows that the Thorsos method produces a data set with a variance that is downward biased, the PSD of the data set is unbiased. Using the standard FFT approach to computing the PSD of a data set, Fig. 6(a) shows a typical
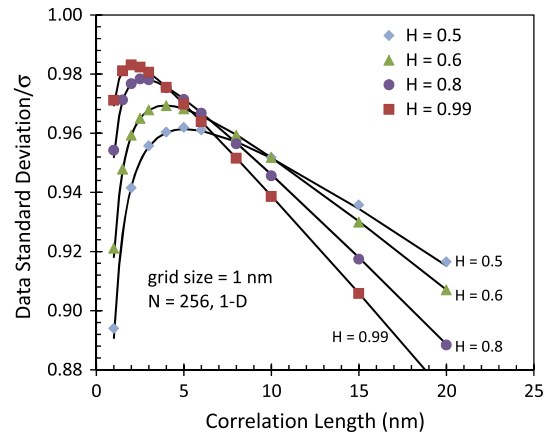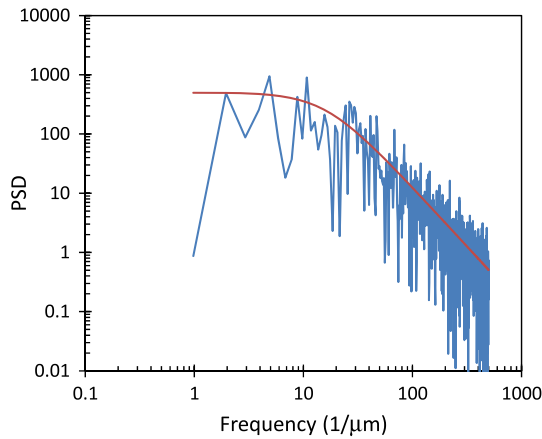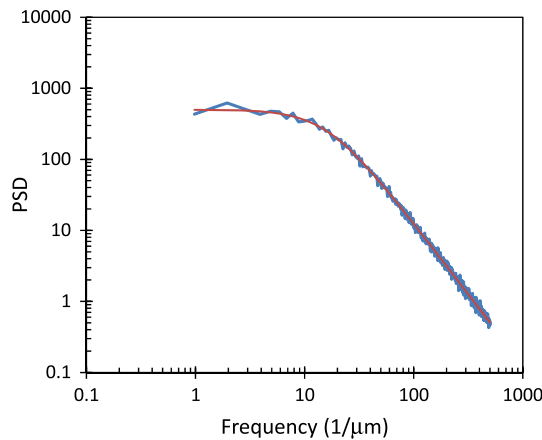
Fig. 5. (Color online) For the 1D case, $N = 256$, comparison of numerical simulations using the Thorsos method (symbols) to the predictions of Eq. (31) (solid curves) for various values of the roughness exponent H.

(a)

(b)

Fig. 6. (Color online) Typical PSD taken from a generated rough edge ($\sigma = 5$ nm, $\xi = 10$ nm, $\Delta x = 1$ nm, and $N = 1024$): (a) one trial and (b) average of 100 trials. The input PSD is shown as the smooth red curve.

result (for the 1D case). As is well known, this periodogram estimate of the power spectrum produces a PSD with a relative standard deviation of 1 at each frequency (that is, the standard deviation of the PSD is equal to the expectation value of the PSD), independent of N [16]. To verify that this numerical PSD converges to the input PSD, numerous trials of the PSD should be averaged together [Fig. 6(b)].

By comparing the PSD of the generated data to the input PSD, the RMS relative difference can be calculated. Figure 7 shows the convergence of the generated PSD as a function of the number of trials being averaged (M), showing that it follows the expected $1/\sqrt{M}$ trend. Various values of N and ξ do not change this result. Further, the uncertainty in each RMS calculation is equal to $1/\sqrt{NM}$. Thus, the data points in Fig. 7 are the average of numerous RMS calculations.

For 2D and 3D data, extracting a radially symmetric PSD from the 2D or 3D FFT requires interpolating the rectangular data onto a radial grid. Further, the radial symmetry means that many different PSD points can be averaged together for each radial frequency. Consider the symmetric 2D case of
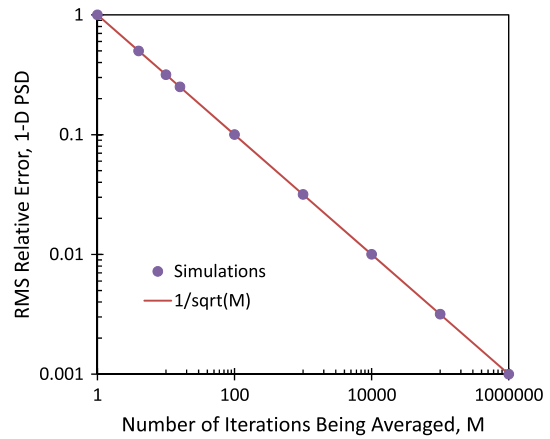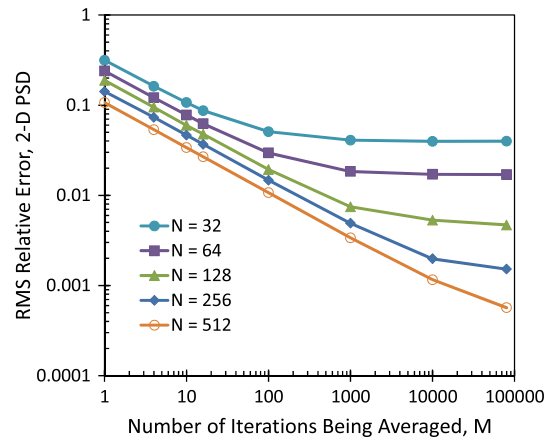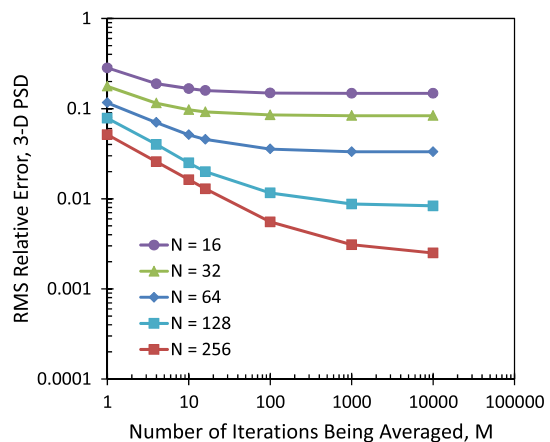

Fig. 7. (Color online) Convergence of the numerically generated 1D PSD to the input PSD function as a function of the number of trials being averaged together ($\sigma = 5$ nm, $\xi = 10$ nm, $\Delta x = 1$ nm, and $N = 1,024$). The standard $1/\sqrt{M}$ convergence trend is shown as the solid line, with simulations shown as the symbols.

an $N \times N$ data set with $\Delta x = \Delta y$. If the radial grid is set to $\Delta r = \Delta x$, then at a radial position of $n\Delta r$ there will be about $(n+1)\pi/2$ different x-y PSD values

(a)

(b)

Fig. 8. (Color online) Convergence of the numerically generated PSD to the input PSD function as a function of the number of trials being averaged together (M) and the number of points per side (N) for $\xi = 10$ nm and $\Delta x = 1$ nm: (a) 2D case and (b) 3D case.

averaged together to produce one radial PSD value (up to $n = N$). For the 3D case, the nth radial PSD value will average together about $(n + 1)^2 \pi/2$ rectangular PSD values. The result of this internal averaging will be a radial PSD with lower RMS difference.

The use of an interpolation algorithm, however, adds interpolation error. Here, a linear interpolation in two or three dimensions was used [taking care not to interpolate using the invalid PSD(0) data point]. Since the PSD shape is decidedly not linear, the use of a linear interpolation adds a systematic error that will decrease with increasing N. Figure 8 shows the results. As can be seen, increasing N results in significant internal averaging when computing the radial PSD and a much lower RMS difference even when the number of trials being averaged together (M) is one. As M increases, however, the RMS difference quickly saturates at a value determined by the interpolation error. Unless N is sufficiently large, errors in the PSD can be dominated by this linear interpolation error (Fig. 9). In such cases, quadratic or other more sophisticated interpolation schemes could be used to reduce this error.

A short discussion of random numbers is now in order. Pseudo-random number generators (PRNGs) are used to generate the "random numbers" used in the Thorsos method. An important but often underappreciated aspect of such PRNGs is their period—how many random numbers can be generated before the sequence repeats. Consider one of the data points in Fig. 7: $N = 16,384$, $M = 1,000,000$. Two random numbers are needed per data set point, and 10 separate runs were used to estimate the uncertainty in the RMS value plotted in the figure. Thus, a string of 3.3×10^{11} independent random numbers was required. A typical 24 bit random number generator, the default in many computer language compilers, can only generate $2^{24} - 1$ (16 million) independent random numbers—obviously inadequate here. A 32 bit PRNG is also insufficient (period of 4 billion). In this work, a Mersenne Twister

PRNG was used, with a period of $2^{19937} - 1$ [17]. The Box–Muller algorithm was used to convert the uniform distribution of the PRNG into a Gaussian distribution.

7. Conclusions

The Thorsos method for generating random rough edges, surfaces, and volumes is both simple and powerful. It produces data sets whose means are without bias, and has the desired spectral characteristics PSD without bias. The process of modeling a continuous function of infinite extent using a discrete set of data of finite extent, however, inevitably results in data with a variance (and standard deviation) that is downward biased. Further, the amount of bias is a strong function of not only the number of data points but also the parameters of the model: the data parameters of grid size and data length, and the PSD parameters of correlation length and roughness exponent. Under common modeling conditions, the bias in the standard deviation can be small (less than 1%) or large (greater than 10%), depending on the parameter settings.

A common modeling approach is to fix the data size (N) and the grid size (Δx), and then generate random data sets with varying standard deviation (σ), correlation length (ξ), and roughness exponent $(\alpha$ or $H)$. As the above analysis shows, however, the resulting data sets will have varying degrees of bias. Thus, it may be difficult to distinguish a true effect of varying roughness versus an apparent effect of varying data bias. A better approach might be to select the grid size that always minimizes the bias. Then, at least for a fixed roughness exponent, the relative bias will remain constant. For a varying roughness exponent, this optimum grid size will at least minimize the bias, though it will not keep it constant. The significance of these effects is a strong function of the roughness exponent, with $\alpha = H = 0.5$ providing the worst case.

References

1. P. Naulleau and J. Cain, "Experimental and model-based study of the robustness of line-edge roughness metric extraction in the presence of noise," J. Vac. Sci. Technol. B. **25**, 1647–1657 (2007).
2. V. Constantoudis, G. Patsis, L. Leunissen, and E. Gogolides, "Line edge roughness and critical dimension variation: fractal characterization and comparison using model functions," J. Vac. Sci. Technol. B. **22**, 1974–1981 (2004).
3. Z. Chen, L. Dai, and C. Jiang, "Nonlinear response of a silicon waveguide enhanced by a metal grating," Appl. Opt. **51**, 5752–5757 (2012).
4. A.-Q. Wang, L.-X. Guo, and C. Chai, "Fast numerical method for electromagnetic scattering from an object above a large-scale layered rough surface at large incident angle: vertical polarization," Appl. Opt. **50**, 500–508 (2011).
5. J. C. Novarini and J. W. Caruthers, "Numerical modelling of acoustic wave scattering from randomly rough surfaces: an image model," J. Acoust. Soc. Am. **53**, 876–884 (1973).
6. N. Garcia and E. Stoll, "Monte Carlo calculation for electromagnetic-wave scattering from random rough surfaces," Phys. Rev. Lett. **52**, 1798–1801 (1984).
7. E. I. Thorsos, "The validity of the Kirchhoff approximation for rough surface scattering using a Gaussian roughness spectrum," J. Acoust. Soc. Am. **83**, 78–92 (1988).
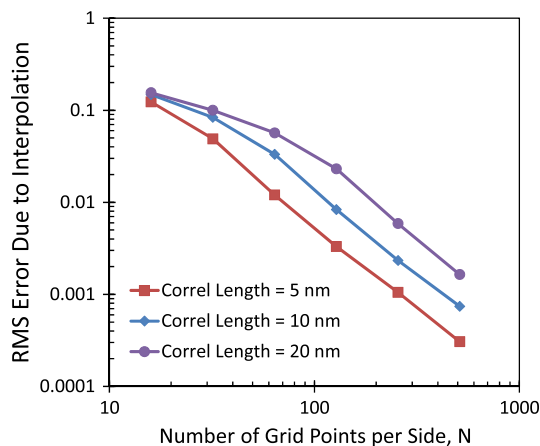
Fig. 9. (Color online) Linear interpolation error (relative RMS error of output PSD compared to the input PSD) when converting a 3D PSD onto a radial grid ($\Delta x = 1$ nm).

8. C. A. Mack, "Reaction-diffusion power spectral density," J. Microlithogr. Microfabr. Microsyst. **11**, 043007 (2012).

9. C. A. Mack, "Correlated surface roughening during photoresist development," Proc. SPIE **8325**, 83250I (2012).

10. T. H. Naylor, J. L. Balintfy, D. S. Burdick, and K. Chu, *Computer Simulation Techniques* (Wiley, 1966), pp. 118–121.

11. Y. Z. Hu and K. Tonder, "Simulation of 3-D random rough surface by 2-D digital filter and Fourier analysis," Int. J. Mach. Tools Manuf. **32**, 83–90 (1992).

12. L. Tsang, J. A. Kong, K.-H. Ding, and C. Ao, *Scattering of Electromagnetic Waves—Numerical Simulations* (Wiley, 2000).

13. C. A. Mack, "Analytic form for the power spectral density in one, two, and three dimensions," J. Microlithogr. Microfabr. Microsyst. **10**, 040501 (2011).

14. J. A. Ogilvy and J. R. Foster, "Rough surfaces: Gaussian or exponential statistics?," J. Phys. D. **22**, 1243–1251 (1989).

15. G. Palasantzas, "Roughness spectrum and surface width of self-affine fractal surfaces via the K-correlation model," Phys. Rev. B. **48**, 14472–14478 (1993).

16. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran 77: The Art of Scientific Computing*, 2nd ed. (Cambridge University, 1992), p. 545.

17. M. Matsumoto and T. Nishimura, "Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator," ACM Trans. Model. Comput. Simul. **8**, 3–30 (1998).