Original Article

Institution of MECHANICAL ENGINEERS

Journal of ENGINEERING TRIBOLOGY

Proc IMechE Part J:
J Engineering Tribology
0(0) 1–18
© IMechE 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1350650119885281
journals.sagepub.com/home/pij
SAGE

Metrology and numerical characterization of random rough surfaces—Data reduction via an effective filtering solution

Itzhak Green

Abstract

Random rough surfaces appear in measurements as noisy signals varying spatially. Mathematically, there is no theoretical difference between such and time-varying signals. Hence, the extensive array of methods and analysis tools that have been developed for signal processing are available also for rough surfaces characterization. In both, the objective is to reduce the vast amount of data to just a few meaningful parameters that allow the application of other physical concepts. Particularly in contact mechanics, it is well known that the Greenwood–Williamson model requires three parameters for the calculation of the elastic deformation of rough surface asperities. The parameters are the roughness standard deviation, the equivalent asperity radius, and the asperity density. These parameters are byproducts of the spectral moments. The spectral moments have been employed for decades in many fields of engineering and science. For rough surfaces, for example, the work by McCool outlines a mathematical blueprint procedure on how to straightforwardly reduce the entire roughness data into the said three spectral moments. It is commonly claimed, however, that the said procedure inherently suffers from resolution problems, that is, a given surface shall have much different spectral moments depending on the sampling rate (or spacing). To study these issues, synthetic surfaces are generated herein using a harmonic waveform precisely as McCool had done. However, here the signals are contaminated by a white noise process with various magnitudes. A signal-to-noise ratio is defined and used to assess the quality of the signal, and the spectral moments are evaluated for various magnitudes of the noise. Since closed-from solutions are available for the spectral moments of the uncontaminated signal, the contaminated signals are evaluated vis-à-vis the exact anticipated values, and the errors are calculated. It is shown that using the common techniques (such as those outlined by McCool) can lead to enormous and unacceptable errors. Resolution is studied as well; it is shown to have an effect only in the presence of noise, but by itself it has no independent influence on the spectral moments. The venerable Savitzky–Golay smoothing filter is used on the noisy signals, showing some improvements, but the resulting spectral moments predicted still contain objectionable errors. A generalized exponential smoothing filter, G-EXP, is constructed, and it is shown to markedly moderate the errors and reduce them to acceptable levels, while effectively restoring the underlying surface physical characteristics. Moreover, the filtered signals do not suffer from resolution problems, where results, in fact, improve with higher (i.e., finer) resolutions. Fractal-generated signals are likewise discussed.

Keywords

Surface roughness, surface metrology, spectral moments, white noise contamination, signal processing, signal-to-noise ratio, filter design, Savitzky–Golay filtering

Date received: 15 May 2019; accepted: 7 October 2019

Introduction

This work puts emphasis on data reduction regarding rough surfaces for the purpose of contact mechanics calculation, but the concepts herein apply equally well to the processing of any signal contaminated by a random noise.

The Greenwood–Williamson (GW)[1] approach to modeling the contact of two elastic rough surfaces has gained wide acceptance. The approach reduces the two rough surfaces into a single equivalent rough surface that is forced against a perfectly smooth and rigid flat. The common assumptions are that the equivalent surface has asperities that deform independently of the neighboring

GWW School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Corresponding author:
Itzhak Green, GWW School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA.
Email: itzhak.green@me.gatech.edu

asperities, all asperities have an equivalent radius, R, they have an areal density, η, and that they are distributed by some probability density function. At the time of its development (1966), in the absence of a closed-form solution for a Gaussian distribution, GW used an exponential distribution that could be solved analytically in closed-form, which led to some physical conclusions. However, even GW state and show that surfaces' height distribution tends to be Gaussian rather than exponential. It was not until 2011 that Jackson and Green[2] provided a closed-form solution to the GW model using an uncompromised Gaussian distribution. That work[2] has also demonstrated the resolution issue where the same measured data of real rough surfaces can provide very different spectral moments depending on the spacing (see Table 1 there).

The GW has gained acceptance in contact of rough surfaces even under elasto-plastic loading.[3–5] Lacking in the original GW work, however, is the mathematical process of reducing the two rough surface properties into a single rough one. The work by McCool[6] fills this gap by providing a complete mathematical blueprint on how two surfaces having three-dimensional (3D), orthotropic roughness, $z = z(x,y)$, can be converted into the desired single surface having a composite roughness. That work[6] allows, without loss of generality, to employ two-dimensional (2D) roughness quantities, and that is precisely what is considered herein, i.e. $z = z(x)$. As also summarized by McCool,[6,7] the three quantities m_0, m_2, and m_4, known as the spectral moment, are sufficient to completely define all the parameters needed in the GW model. These moments can be obtained in the spatial domain by

$$m_0 = \frac{1}{N}\sum_{i=1}^{N}(z)_i^2 \tag{1}$$

$$m_2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{dz}{dx}\right)_i^2 \tag{2}$$

$$m_4 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{d^2z}{dx^2}\right)_i^2 \tag{3}$$

where N is the total number of data points sampled on a surface along a generic coordinate x, while $z = z(x)$ is the asperity height measured from the mean surface. In the said works,[6,7] an alternative method is propounded by using the power spectrum $P(\omega)$ of the waveform $z(x)$ to yield the kth moments

$$m_k = \int_0^{\infty} \omega^k P(\omega)\mathrm{d}\omega \quad @\ k = 0,\ 2,\ 4 \tag{4}$$

Here, $\omega = 2\pi f = 2\pi/\lambda$, where ω is the circular frequency of f and λ is the wavelength (being equivalent to the period had $z = z(t)$ been a waveform in time, t). Note that for $k = 0$, equation (4) signifies Parseval's theorem. For additional information, see Sweitzer et al.;[8] Davidson and Loughlin;[9] Vogel;[10] and Brown.[11] Evidently, these moments are not specific to modeling surface roughness just in tribology, as they are central in the many fields of science and engineering, falling generally into the category of signal processing for which there is ample literature, see notably the excellent classical texts by Bendat and Piersol.[12–14] Non-tribological examples can range from geomechanics of rough wall fracture[11] to signal processing performed on the output from the pulsed laser photoacoustic instrument monitoring crude oil in water,[10] or in the analysis an optical telescope.[8] Specifically, in tribology though, these three

Table 1. Values of the exact spectral moments and their numerically calculated values with relative errors for various noise amplitudes, ΔA, and resolutions, δx.

	$m_0 = 1/2 = 0.5$		$m_2 = 2\pi^2 = 19.739$		$m_4 = 8\pi^4 = 779.27$			
At $\Delta A = 0$	Value	Error %	Value	Error %	Value	Error %	$\alpha = 1$	$SNR_{dB} = \infty$
$nfft = 9$, $\delta x = 0.012272$								
$\Delta A = 1\%$	0.503	0.5	19.87	0.7	16681.1	2.04E3	21.24	41.5
$\Delta A = 10\%$	0.508	1.7	42.04	113	1.591E6	2.04E5	457.8	21.5
$\Delta A = 30\%$	0.541	8.3	220.7	1018	1.432E7	1.84E6	159.3	12.2
$nfft = 11$, $\delta x = 0.003068$								
$\Delta A = 1\%$	0.503	0.5	22.9	16.23	3.830E6	4.91E5	3656.7	41.5
$\Delta A = 10\%$	0.505	1.1	349.3	1669	3.829E8	4.91E7	1586.0	21.5
$\Delta A = 30\%$	0.531	6.2	2986	1.5E4	3.446E9	4.42E8	205.22	12.2
$nfft = 12$, $\delta x = 0.001534$								
$\Delta A = 1\%$	0.503	0.5	32.5	64.52	5.846E7	7.50E6	27865	41.5
$\Delta A = 10\%$	0.506	1.3	1303.9	6505	5.846E9	7.50E8	1741.5	21.5
$\Delta A = 30\%$	0.534	6.8	11578	5.9E4	5.26E10	6.75E9	209.61	12.2

Notes: Given also are the bandwidth parameter, α, and the signal-to-noise ratio in (dB). For all cases, $A = 1$ m and $f = 1$ Hz.

moments are sufficient to execute the GW model, and they are focal in this work.

Similar coverage of the subject is recapped also by McCool,[7] where he *specifically* suggests that the calculation of the derivatives in equations (1) to (3) be done by finite difference approximations, stating that it offers a simpler approach to using equation (4). The reasoning offered is that that approach has computational speed advantages, it avoids "leakage" in the calculation of $P(\omega)$ that plagues the spectral estimation, and that there are other constraints. To prove his point, McCool[7] employs a pure sine waveform of amplitude, A, and frequency, f, given by

$$z(x) = A\sin(2\pi f x) \tag{5}$$

For this analytic waveform, the moments can be calculated exactly from equations (1) to (3), and McCool finds that the ratio of the approximated to the exact moments depends only upon the number of intervals per period, N, but not upon A or f. To prove his point, McCool varies the number of sample intervals per period from 3 to 50, and concludes that 8 intervals are sufficient to calculate m_2 with less than 5% underestimated error, and m_4 with a 3% overestimated error. (McCool continues to examine instrumentation and sampling relevant to that era, but that is mostly irrelevant for today's instrumentation.) Importantly though, McCool uses a forward difference algorithm for the first derivative, known to have a truncation error of $O(\delta x)$, and a central difference for the second derivative, known to have a truncation error of $O(\delta x^2)$. Here, δx, is the equidistant spacing between two adjacent sampled points, and is denoted as the resolution. For a pristine waveform such as in equation (5), a closed-form solution is possible for the spectral moment (as given below), and likewise these can be obtained with great accuracy using the finite difference approach suggested by McCool. It is emphasized that numerical differentiation to calculate equations (2) and (3) is commonplace in tribology and used by many, for example Jackson and Green;[2] Pawar et al.;[15] Xu and Jackson;[16] and Kalin.[17] Problems arise when signals are not quite as pristine, as clearly it is the case for data of real surfaces. That is the subject of this work. It is important to highlight before proceeding that even for the pristine signal of equation (5), the calculation of the spectral moments fails using equation (4), as it is detailed in Appendix A. That finding provides a more convincing argument for not using the spectral approach on the signal defined by equation (5) (rather than the "leakage" explanation mentioned by McCool). The stated problem that is outlined below escalates in difficulty as this work unfolds.

The problem

The procedure offered by McCool[7] is re-examined herein. First, suppose that the signal data are available in the spatial range $x \in [0, x_{\max}]$. The moments for a continuous waveform are calculated exactly using a continuous (integral) form of equations (1) to (3), namely

$$\begin{aligned} m_0 &= \frac{1}{x_{\max}} \int_0^{x_{\max}} [z(x)]^2 \mathrm{d}x \\ m_2 &= \frac{1}{x_{\max}} \int_0^{x_{\max}} [z'(x)]^2 \mathrm{d}x \\ m_4 &= \frac{1}{x_{\max}} \int_0^{x_{\max}} [z''(x)]^2 \mathrm{d}x \end{aligned} \tag{6}$$

Substituting equation (5) in equation (6), yields

$$\begin{aligned} m_0 &= \frac{A^2\left[\frac{\mathrm{x}_{\max}}{2} - \frac{\sin(4\pi f \mathrm{x}_{\max})}{8\pi f}\right]}{\mathrm{x}_{\max}} \\ m_2 &= \frac{\pi A^2 f[4\pi f \mathrm{x}_{\max} + \sin(4\pi f \mathrm{x}_{\max})]}{2\mathrm{x}_{\max}} \\ m_4 &= \frac{2\pi^3 A^2 f^3[4\pi f \mathrm{x}_{\max} - \sin(4\pi f\, \mathrm{x}_{\max})]}{\mathrm{x}_{\max}} \end{aligned} \tag{7}$$

For simplicity, suppose that $\sin(4\pi f x_{\max}) = 0$, i.e. $x_{\max}$ is a signal length that always renders complete cycles. In which case, the set $\{m_0,\ m_2,\ m_4\}$ is independent upon $x_{\max}$, giving exactly

$$m_t = \{m_0,\ m_2,\ m_4\} = \left\{A^2/2,\ 2(\pi A f)^2,\ 8A^2(\pi f)^4\right\} \tag{8}$$

A so-called bandwidth parameter is defined by Nayak[18] (and used, e.g. by McCool[6] and Pawar et al.[15])

$$\alpha = \frac{m_0 m_4}{m_2^2}$$

Where upon substitution of equation (8), $\alpha = 1$, as it ought to be for the single-frequency pristine signal of equation (5). Notably, this parameter is used to calculate the standard deviation of the asperities' summit height in the GW model.

Without repeating McCool's derivation (see McCool[7]), a finite difference scheme is employed to obtain numerically the approximated moments. Consistent with McCool's observation (as verified again), that for the said special case of $x_{\max}$, the ratio between the approximated to the exact values of m_k, $k = 0,\ 2,\ 4$, are indeed independent of A and f. So, arbitrarily, for the remainder of this work, select $A = 1$ m, and $f = 1$ Hz. Let ΔA be a *modulated noise amplitude*, which contaminates the signal expressed in equation (5). Hence, when $\Delta A = 0$, there is no noise, and the first row in Table 1 represents the results of the "pristine" or "ideal" signal. For that case, the moments are provided exactly, as well as by their numerical values. The noise amplitude ΔA shall be elaborated upon shortly, as it greatly affects the numerical values of the spectral moments, which are the objectives (i.e. target values) of this work. Also given are the bandwidth parameter as discussed above and the signal-to-noise ratio (SNR) (which is defined and discussed below).

So, the signal in equation (5) is an ideal (i.e. "pristine") sine wave, and that is the signal that McCool[7] focused upon. However, that is an unrealistic expectation for the behavior of real surfaces. Clearly, real surfaces shall always exhibit some noise in the measured signal (where in fact quite frequently, considerable noise should be expected[10,19,20]). Suppose that a random noise process of magnitude ΔA is superimposed upon the pure sine waveform of equation (5). The entire signal is constructed using the following Mathematica script (again for $A = 1\,\text{m}$ and $f = 1\,\text{Hz}$)

$$\begin{aligned}
&w = 2\pi; \quad nfft = 9;\\
&n = 2\,\hat{}\,nfft + 1; \quad delx = 2\pi/(n-1);\\
&x = Table[(i-1) * delx,\ \{i, 1, n\}];\\
&signal = Sin[w * x];\\
&SeedRandom[1234]; \quad \Delta A = 0 * 0.01;\\
&noise = \Delta A * Table[RandomReal[\{-1, 1\}],\ \{i, 1, n\}];\\
&z = signal + noise
\end{aligned} \tag{9}$$

The variable $z = z(x)$ in equation (9) is evidently composed of a pure sine waveform signal of equation (5) having an assigned circular frequency, w, and a white noise contamination having a uniform distribution in the range {–1, 1}, where the noise is modulated by an amplitude, ΔA. The arbitrary seed of *1234* guarantees that all noise cases analyzed herein shall always have the same white noise content throughout (By using the Mathematica script given in equation (9), along with the parameters in Table 1, all the results in this work can be straightforwardly replicated.). The lengths of the signals (the pure waveform and the noise) are set to be powers of *2* via the *nfft* exponent parameter. That is a convenience to help with a fast Fourier transform, when taken. Clearly, that parameter also decides the resolution, δx (as determined by equation (9), and given in Table 1).

To calculate the derivatives in this work, the first and second derivatives are calculated by a finite difference scheme using a *five-point* approximation (see Hildebrand,[21] p. 111). Corresponding to $nfft = \{9,11,12\}$, the number of points are $n = \{513, 2049, 4097\}$, and the truncation errors, for both first and second derivatives, are of order, $O(\delta x^4) = \{2.27*10^{-8}, 8.86*10^{-11}, 5.54*10^{-12}\}$, respectively. This is opposed to McCool's $n = 51$, having the finest truncation errors of orders, $O(\delta x) = 0.02$, for the first derivative, and $O(\delta x^2) = 4*10^{-4}$, for the second derivative. Hence, in the current work, the estimations for the derivatives are of truncation errors of *at least* four orders of magnitude smaller (better) than McCool's calculations. Also, six cycles are used herein (see Figure 1), where McCool uses only one cycle. Clearly, the numerical procedure used herein is considerably more accurate and robust than McCool's procedure.[7] To verify the validity of the current numerical procedure, the noise amplitude is set first to zero in equation (9), i.e. $\Delta A = 0$, and the numerical results obtained for the moments turn out to be identical to those of the exact predictions (with no error visible within the first six (6) significant digits). That is true for *any* resolution, δx. With that, the numerical derivative procedure used herein is verified. It is explicitly *emphasized* that in this work, only the five-point finite difference scheme of higher order $O(\delta x^4)$ is used to calculate the numerical derivatives.

Now, Table 1 summarizes the spectral moment values along with the deviation from the exact value for other values of ΔA and *nfft*, i.e. δx, which are varied in equation (9). The deviation, or relative error, is calculated according to "100%*abs(exact_value – numerical_value) / exact_value". The details are as follows.

To start off, a tiny random noise amplitude of 1% is tacked upon the signal (i.e. $\Delta A = 0.01\,\text{m}$). Figure 1(a) shows an ideal sine wave signal, equation (5), and its exact first and second derivatives in black color. The noisy signal and its respective numerical derivatives are shown in red color. When the pure sine waveform and the noisy signal are plotted together (left most plot in Figure 1(a)), it is nearly impossible to tell the difference between the two signals. However, the first derivative already shows some deviation from the exact solution near the inflection points, while the second derivative is hugely over-estimated throughout.

The zeroth moment m_0 contains no derivatives and, hence, it is predicted nearly exactly for any resolution, as shown in Table 1. The second moment m_2 is affected by the errors in the first derivative, and when averaging takes place by equation (2), the error varies from 0.7% to 16.23%, and then to 64.52% depending on the resolution. However, the fourth moment m_4 is affected significantly by the huge errors in the second derivative, and when averaging takes place by equation (3), the error varies from 2040% to 7.5×10^6%. That cannot be considered acceptable under any circumstances. The bandwidth parameter, α, which should have equaled unity, is also grossly overestimated, ranging from 21.24 to 27,865.

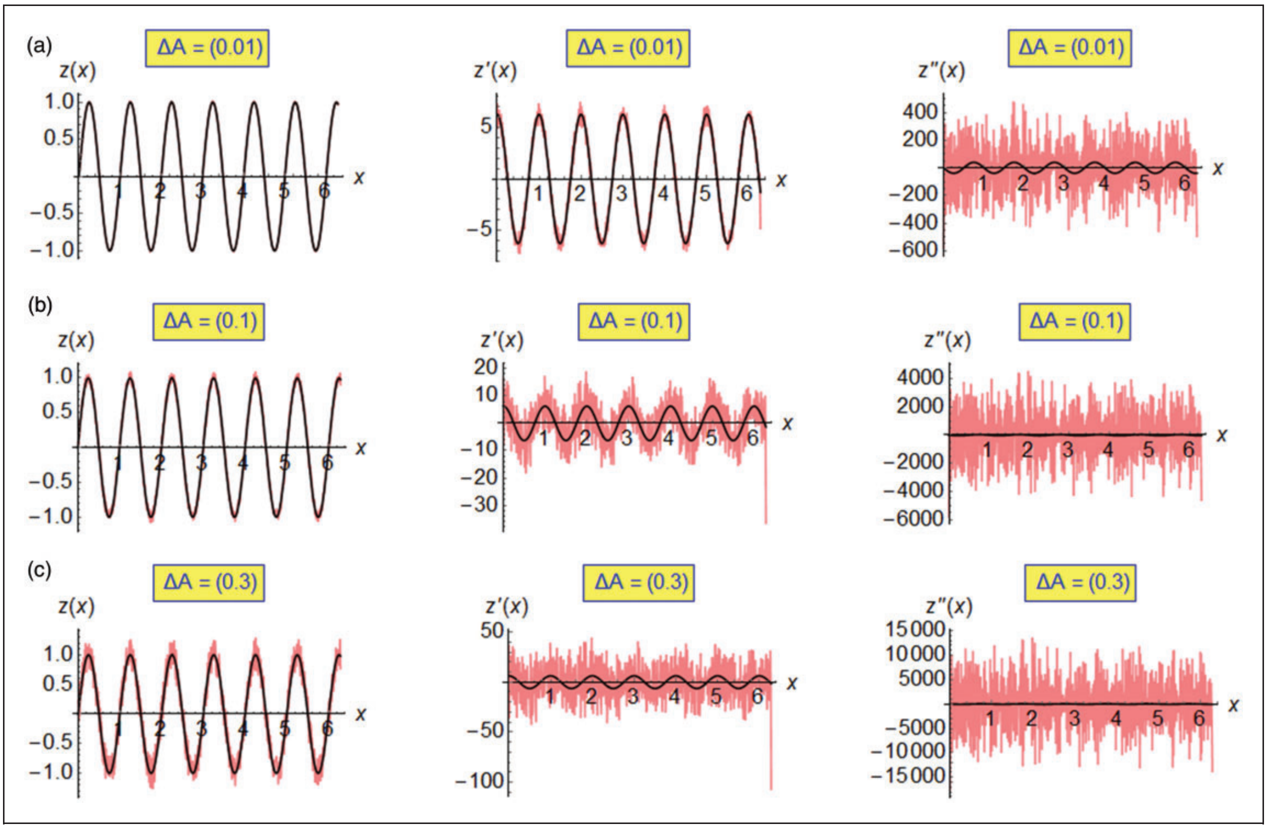

Figure 1. Signals, *z(x)*, first derivatives, z′(*x*), and second derivatives, z″(*x*), shown for three noise amplitudes, Δ*A* and *nfft* = 9.

If all that happens just for a tiny noise of 1% that contaminates the signal, matters can only get worse with larger noise levels. Practically, there is always noise in the measuring equipment in addition to the fact that real surfaces are simply imperfect. It is also intuitively understood that such small noise levels should not have a meaningful effect when the surfaces are brought (loaded) into contact, because such small surface undulation would be structurally "weak," and they will be leveled (smashed) by the initial application of a normal load. However, with such large errors in m_2 and particularly m_4, the necessary GW parameters cannot be considered trustworthy even for a 1% noise level. Moreover, it seems that indeed, the errors exacerbate with the refinement of the resolution, a phenomenon that is analyzed in detail hereunder.

Next, larger contaminations are investigated. Suppose that the noise magnitude has a moderately larger value of 10% (i.e. $\Delta A = 0.1$ m, see Figure 1(b)). The errors in the prediction of the moments m_2 and m_4 are expected to worsen, and indeed they escalate rather significantly, as indicated in Table 1. And when the noise level is 30% (see Figure 1(c)), the predictions of m_2 and m_4 are practically useless. Likewise, the bandwidth parameter, α, is very much off from the ideal value of one unit, regardless of the magnitude of the noise or the resolution. Appendix A takes on the noisy signal case of $\Delta A = 0.3$ m, in an attempt to evaluate the moments by spectral means using equation (4). It is proven that on the subject problem, that method is *incapable* to produce exact or even satisfactory results. Hence, that leaves the differentiation method as the only option, but evidently, a remedy is sternly needed.

It is clear that if the differentiation method would ever render trustworthy spectral moments, then it necessary to reform the raw noisy signals to expose the underlying geometry before taking the derivatives, such that the moments would be principally unaffected by the imperfections. But before that objective is handled in a later section, a metric for signal "*goodness*" must be put forward. The metric of choice herein is the SNR. That metric is useful not only in assessing the "*goodness*" of the raw signal but it is also used to assess improvements in the proposed methods offered that are forthcoming.

The SNR

The SNR is a common measure that compares the level of a desired signal to the level of the background noise, and it is defined as the ratio of the signal power, P_{signal}, to the noise power, P_{noise}, often expressed in decibels. A ratio higher than one unit (greater than *0 dB*) indicates that the signal is more powerful than the noise. It can be shown that the SNR also equals to the ratio of the corresponding variances of the signal and noise. The following expressions are all equivalent

$$SNR_{dB} = 10Log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) = 10Log_{10}\left(\frac{\sigma^2_{signal}}{\sigma^2_{noise}}\right) = 20Log_{10}\left(\frac{RMS_{signal}}{RMS_{noise}}\right) \tag{10}$$

For the pure deterministic sine wave of equation (5), the root mean square (*RMS*) is $A/\sqrt{2}$, and since in this work $A = 1$ m, and $f = 1$ Hz, then the $RMS = 0.707$ is fixed for that pure signal. Since the noise is superimposed upon that pure signal (see equation (9)), its power and *RMS* values are calculated separately. Herein, only the *RMS* value is used, and it equals to the second central moment of the noise signal (see equation (9) for the definition of *noise*). In Mathematica's notation, two equivalent forms are given, an intrinsic function, and by its definition, hand-coded

$$\begin{aligned} RMS_{noise} &= CentralMoment[noise, 2] \\ &\quad (\ast\text{a Mathematica intrinsic function}\ast) \\ &= Sqrt[(noise - Total[noise]/n) \\ &\quad \cdot(noise - Total[noise]/n)/n] \end{aligned} \tag{11}$$

For the case when $\Delta A = 0$ (i.e. no noise), clearly the SNR is infinity, representing a perfect or ideal signal. For the other three cases in Table 1, the SNR_{dB} is decreasing with the increase in the noise amplitude, ΔA (As indicated, all white noise records herein use the same seed and procedure of equation (9). The only difference is that they are modulated by the amplitude ΔA. Thus, the *RMS* values and the noise amplitudes, ΔA, are proportional. For example, the *RMS* value for $\Delta A = 10\%$ is that of the 1% noise, multiplied by a factor of 10, etc.).

As a reference, the following is accepted amongst the telecommunication industry for wireless (cellular) networks:

1. When the SNR_{dB} is greater than *40 dB*, then the signal is excellent (five bars), and the connection is "*lightning*" fast;
2. When the SNR_{dB} is between *25–40 dB*, then the signal is very good (3–4 bars), with very fast connection;
3. When the SNR_{dB} is between *15–25 dB*, then the signal is of low or poor quality (two bars), but it may be acceptable if SNR_{dB} is still above *20 dB*;
4. When the SNR_{dB} is between 0–15 dB, then the signal is very poor (one bar), it is unreliable, and mostly there is a slow connection, if at all; and when SNR_{dB} is between 5–10 dB, connection is unlikely.

For the lack of a scale in tribology that is specific for "signal quality" of rough surfaces, suppose that the above ranges from the telecommunication industry can be adopted. Then, the cases herein for $\Delta A = \{0, 0.01, 0.1, \text{and } 0.3\}\ m$ render, respectively, signals with $SNR_{dB} = \{\infty, 41.5, 21.5, 12.2\}$ (see Table 1), going from perfect, and then degrading to excellent, good, and low. Reiterating and emphasizing that even for an "*excellent*" $SNR_{dB} = 41.5$ belonging to $\Delta A = 0.01$, the spectral moments, as seen above, *especially* m_4, cannot be trusted.

Resolution issues

A notion that is prevalent in the tribology research community is that the spectral moments are very sensitive to the sampling intervals, or to the resolution.[2,15,17,22] Observing the data in Table 1, *seemingly* that perception is "confirmed." In fact, that perception has prompted the development of other methods, for example, peak points, shoulders, neighboring asperities, etc.[17,22,23] The resolution's "*negative reputation*" truly deserves a much closer examination.

First, the resolution (spacing) of δx at $nfft = 11$ ($n = 2049$) is four (4) times finer than that with $nfft = 9$ ($n = 513$). The first observation from Table 1, is that for m_0, the errors for $nfft = 9$ and $nfft = 11$ are about the same for the same ΔA, with the trend that as ΔA increases so does the error, but very slightly. In other words, the resolution does not affect much the error in m_0. The reason that the error increases with ΔA is logical because the noise adds to the signal magnitude, and the larger the noise the larger the error. But, it is apparent that the errors for m_2, for the finer resolution of $nfft = 11$, are much larger than the corresponding value of $nfft = 9$, and are significantly larger for m_4. That may be counter intuitive because the truncation error in estimating the derivatives is much smaller for the case of $nfft = 11$, which is indeed so, but the truncation error is *not* the reason. The reason, as detailed in Appendix B, is that m_2 and m_4 depend on the derivatives of the signal, i.e. the differences between neighboring noisy points across a smaller (finer) δx. Hence, the theoretical error for m_2 at $nfft = 11$ should be $4^2 = 16$ higher than that for $nfft = 9$, and for m_4 it should be $(4^2)^2 = 256$ higher, respectively. Close examination in Table 1, say for $\Delta A = 0.1\,m$, confirms that finding with actual numerical values of *14.8* (vs *16* theoretically), and *241* (vs *256* theoretically), respectively. Roughly, that behavior holds for other noise levels. The theoretical error for m_2 at $nfft = 12$ should be $2^2 = 4$ higher than that for $nfft = 11$, and for m_4 it should be $(2^2)^2 = 16$ higher, respectively, and indeed that trend, by and large, is confirmed in Table 1.

Another observation that is apparent is that for each separated resolution, the error between $\Delta A = 0.01$ and *0.1* is nearly $10^2 = 100$ fold, and between $\Delta A = 0.1$ and *0.3* it is nearly $3^2 = 9$ fold, as that should be so because of the square powers in equations (1) to (3). It is, therefore, concluded that the resolution has an effect only in the presence of noise, but for a perfect signal with no noise, the resolution has absolutely no independent effect at all. In other words, the culprit is in the formulation that depends upon numerical differentiations that amplify the errors in inexact data. If that difficulty can be mitigated, then higher resolutions should actually provide results that are more dependable.

Data partitioning

The signals with *nfft* = *9* and *nfft* = *11*, besides being of different length (i.e. number of sampled points of *n* = *513* and *n* = *2049*, respectively) the white noises generated are not spread across the length the same. Only the first 513 values of the noise for *nfft* = *11* are the same as for *nfft* = *9*, but the rest are not—they are additional independent noise values (still, having the same statistics throughout).

So, it may be argued that an ‘‘equitable’’ comparison must use data from the same record. Hence, the signal of *nfft* = *11* is partitioned (apportioned) four times into four signals where the first signal is made of the values of 1, 5, 9, 13,..., the second takes on the values of 2, 6, 10, 14... etc. These four signals have a reduced *nfft* = *9*, with values taken from the original record of data (those from the signal of *nfft* = *11*). All four partitioned signals have the same resolution as for the case of *nfft* = *9*. For brevity, only the worst case of $\Delta A = 0.3$ is analyzed here. Those four signals are analyzed each individually, and the moments are averaged, where by doing so the statistical error is further reduced by a factor of $4^{1/2} = 2$, yielding $\{m_0, m_2, m_4\} = \{0.531, 222.5, 1.447E7\}$ with corresponding standard deviations of *{2%, 9.2%, 12.9%}* when normalized by the averaged values. Comparing these moments with those given in Table 1 for the case of *nfft* = *9*, and $\Delta A = 0.3$, shows nearly identical matches. Also the average $SNR_{dB} = 12.2$ for the four apportioned signals is almost identical to the said case in Table 1. Hence, the conclusion is that the results summarized in Table 1 may be regarded as a faithful representation for each one of the cases regardless of how data are apportioned.

Interpolation

Another approach to take numerical derivatives is to use interpolation functions. In fact, Mathematica does not contain built-in (intrinsic) finite difference numerical derivative functions (the ones mentioned above had been hand-coded). A tactic in Mathematica to calculate derivatives of discrete data is to fit interpolation functions to the data, and then take derivatives of the interpolation functions. Hermite polynomials and n-powered splines have been tried, and the general behavior shown in Figure 1 is repeated (and hence, for brevity, it is omitted). In other words, the interpolation approach has not produced better results than those reported in Table 1.

Fractals

In the work by Majumdar and Tien,[24] it is postulated that the Weierstrass–Mandelbrot (WM) function can be ‘‘used to simulate deterministically rough surfaces which exhibit statistical resemblance to real surfaces.’’ Following Majumdar and Tien[24] and Berry and Lewis,[25] the WM function is

$$z(x) = A^{(D-1)} \sum_{n=n_1}^{\infty} \frac{\cos 2\pi\gamma^n x}{\gamma^{(2-D)n}} \quad 1 < D < 2, \quad \gamma > 1 \tag{12}$$

where A is a scaling constant, D is a fractal dimension, and γ determines the density of the spectrum and the relative phase difference between spectral modes. As can be seen, equation (12) is made up by a sum of harmonics. Clearly the harmonic function in equation (5) can serve as a kernel to the sum of equation (12) having modulated frequencies and amplitudes, but the underlying mathematics is obviously the same. Appendix A details the mathematical difficulties to obtaining exact closed-form solutions for the spectral moments via equation (4) for the signal given in equation (5). Indeed, Berry and Lewis[25] ran into the same difficulties in formulating the Weierstrass spectrum. So they introduce a workaround by averaging the spectrum over a range of frequencies. That approximation is adopted by Majumdar and Tien,[24] who propose high and low cutoff frequencies (conjecturing physical reasoning), to replace the bounds of integration in equation (4). That approach had been tried herein too, with no success because the selection of such synthetic high and low cut-off frequencies, whether in fractal signals or those containing a white noise process, is not only subjective, it introduces biases which affect the spectral moment values *significantly*. Regardless of how this is looked at, from a purely mathematical point of view, the moments provided by Majumdar and Tien[24] (specifically equations (6) to (8) there) cannot be considered exact mathematical solutions.

Perhaps the most important observation about the WM function is that it has absolutely no randomness. This is because the three parameters *A, D,* and, γ define deterministically the signal. And because there is no randomness (i.e. there is no noise) the SNR_{dB} equals *infinity*. As such one may contemplate whether the WM function can truly represent random rough surfaces, genuinely pondering about the qualification of ‘‘statistical resemblance’’ made by Majumdar and Tien.[24] Moreover, according to Majumdar and Tien,[24] $\gamma^{n_1} = 1/L$, where L is the sample length, so that in equation (12), $n_1 = -\ln L/\ln\gamma$, somehow approximated to an integer. In numerical computation, an infinite sum cannot be accommodated, so the sum in equation (12) must be truncated. The selection of how many terms are retained in the sum is allegedly tied to the finest resolution (or highest cutoff frequency) of the measuring equipment. That adds another bias that affects the calculated spectral moments. Also Majumdar and Tien[24] state that "it is well known that the determination of the high cutoff frequency is fraught with difficulties." In summary, the fractal approach to rough surfaces representation is burdened with assumptions and approximations. Because of that, and the fact that the WM function contains no randomness, it is excluded from any further numerical investigation presented here, which strictly deals with random processes. Nevertheless, the power spectrum appearing in Appendix A is used as another test bench for the generalized exponential (G-EXP) filter that is forthcoming.

From all the attempts described above, none of the aforementioned methods had produced credible results for m_2 and m_4. The conclusion is that the spectral moments, m_2 and m_4, contain enormous errors and they cannot be trusted whatsoever. The next (fifth) method is offered as a plausible remedy.

Signal conditioning

It is a daunting proposition that the two techniques known to calculate the spectral moments fail wretchedly on the stated problem, which is simple: a harmonic signal that is contaminated by a tiny to a moderate noise. It is clear is that small undulations will mostly be smashed in contact, and should have little to no effect on the contact mechanics of the underlying geometry. Hence, the underlying geometry must be recovered and exposed. It is proposed herein that the raw signal must be conditioned, and the way to handle that is to filter out undulations that are not "natural" or not "significant" to the underlying geometry. While filtering, as it is well known, shall clearly introduce biases, it is a common practice in signal processing.

Modern packages such as Mathematica and Matlab are *rich* with filters for noisy signals (many are dedicated to filter audio and images). Similar filters are available in other procedural languages, e.g. Press[26] and IMSL.[27] A few filters, including Dirichlet, Sine, and Hann have been hand-coded and tried on the given problem. The Gaussian filter as implemented by Mathematica has also been tried. Other filters (e.g. Blackman, Nuttall, Hamming, Bartlett, Kaiser, Lanczos, and Parzen) have been considered but not implemented because under some conditions, they degenerate to those tried. It is emphasized that: (1) it is not the objective of this work to present an exhaustive comparison of the level of success of the multitude of windowed filters and (2) the fact that results are not reported for the said filters indicates that they have not produced meaningful improvements in the prediction of the spectral moments. Two filters stand out, and they are discussed below.

The first is the venerable Savitzky–Golay (SG)[26,28] filter as implemented in the Mathematica intrinsic library. It is found to provide a marked improvement upon the conditioning of the noisy signal. The SG filter is a digital filter that can be applied to a set of digital data points for the purpose of smoothing the data and can increase the precision or fidelity of the data without distorting the signal tendency. This is achieved by fitting successive subsets of adjacent data points with a low-degree polynomial by the method of linear least squares, or in a convolution process. In essence, the method uses a polynomial fit in the same way as a weighted moving average, where the coefficients of the smoothing procedure are predetermined and fixed. Moreover, the same algorithm can be used to calculate not only the smoothed signal, but also its first and second derivatives. So while this is the *best* filter tried out of the Mathematica library, on the given problem herein, it is *not* as *effective* as the G-EXP filter presented subsequently. Hence, these two linear filters are now explored in detail.

The SG filter

In fact, a recent work has already employed the SG filter for rough surfaces.[29] For brevity, the mathematical details are not repeated as numerous sources, including those already mentioned give many details and vast explanations. The SG filter is applied in Antón-Acedos et al.[29] on a series of machined probes producing similar results as the Gaussian or Splines filters in roughness parameters. The authors conclude that the SG filter is an interesting alternative to be applied in the study of surface finish. So, it has been decided to try that filter on the stated problem herein.

First, it should be noted that while there is flexibility in selecting polynomial order and the window size, the strict use of polynomials for smoothing hinders the effectiveness of the SG filter. Nevertheless, by trial-and-error, that filter may produce reasonable results. In this work, the smoothing polynomial is always quadratic (other orders had been tried but with worse outcomes). The window size is determined by a Fibonacci optimization algorithm to produce the smallest normalized norm between the smoothed signal spectral moments, $m_s = \{m_{s0}, m_{s2}, m_{s4}\}$ and the known theoretical values, $m_t = \{m_{t0}, m_{t2}, m_{t4}\}$, which are given by equation (8) for $A = 1\,m,$ and $f = 1\,\text{Hz}$. Procedurally, the optimization objective is given by

$$norm = \sqrt{\left(\frac{m_{s0} - m_{t0}}{m_{t0}}\right)^2 + \left(\frac{m_{s2} - m_{t2}}{m_{t2}}\right)^2 + \left(\frac{m_{s4} - m_{t4}}{m_{t4}}\right)^2} \\ \rightarrow minimum$$

Note that each term individually is normalized by its own theoretical value. That normalization guarantees that all moments are weighed equally (i.e. be of similar significance) in the optimization process.

So first, the noisy signal, *z*, is smoothed by the SG filter using the optimal window size, then derivatives are taken, and finally equations (1) to (3) are executed. Table 2 shows the same information as in Table 1, but now this is after the SG filter is applied. It should be noted that because the SG is a linear filter, the smoothed signal is subtracted from the raw noisy signal, leaving only the effective noise. Hence, the SNR_{dB} can likewise be calculated according to equations (10) and (11).

Figure 2 shows in red the smoothed signals for only the largest noise level case, $\Delta A = 0.3$, using the intrinsic Mathematica function, "SavitzkyGolayMatrix." The optimal window size found is 30. The smoothed signal (shown in red) is compared to the raw noisy signal (shown in gray), and the pure objective signal (shown in black). Visibly, the SG filter is quite effective in smoothing *z*(*x*). However, while the derivative, *z'*(*x*), generally seems to follow the objective, deviations from the target values are quite visible about the extremum points. These deviations when summed up, according to equations (1) to (3), are responsible for the errors in the spectral moment estimations. The biggest problem still remains with the second derivative, *z"(x),* as the deviations there are very significant, where the magnitudes are very far from the target. These two deviations, and particularly the latter, cannot fully restore the objective spectral moments. That is apparent by inspecting the moments in Table 2 where the errors in m_2 and m_4 are still quite large.

Two resolutions are also examined in Figure 2, *nfft* = *9*, and *nfft* = *12* (see also Table 2). It seems again that as the resolution gets finer, the errors get larger. The SG filtered signal and its derivatives (shown in red) have more difficulty following the desired signal and its derivative (shown in black). Particularly, when the resolution is finer, *nfft* = *12*, the deviations are still so enormous in the second derivative (at least an order of magnitude larger than the case for *nfft* = *9*), necessitating dropping it from the figure. Even the bandwidth parameter, α, is getting worse as the resolution gets finer. The fact that the resolution issue remains unresolved is definitely another reason why the SG filter is still falling short.

In summary, while the improvements compared to the results in Table 1 are indeed notable, however, the errors remain at a high level, making the resulting moments untrustworthy. Moreover, the resolution problem also remains unsolved because the errors increase with resolution refinement. In general, with the increase of ΔA, the errors increase, α is deviating quite significantly from the exact value of *1*, and SNR_{dB} decreases. It is again emphasized that this is the best filter tried out of the Mathematica library of intrinsic filters. So while the SG filter provides an improvement, clearly a better filter is needed. That filter is the G-EXP filter that is structured next.

The G-EXP filter

The G-EXP filter shares main characteristics with the SG filter or other filters. Namely: (1) "do no harm," i.e. do not introduce artifacts that are not present in the original data, and it should not distort the original tendency or underlying geometry, (2) apply the filter in a similar way, i.e. by convolving the signal with a windowed filter, (3) preserve linearity, i.e. if the signal is a total of two or more data sets, then the overall filtered effect can be obtained by superposition of the

Table 2. Results using the SG filter.

	$m_0 = 1/2 = 0.5$		$m_2 = 2\pi^2 = 19.739$		$m_4 = 8\pi^4 = 779.27$			
At $\Delta A = 0$	Value	Error%	Value	Error%	Value	Error%	$\alpha = 1$	$SNR_{dB} = \infty$
nfft = 9, $\delta x = 0.01227$								
$\Delta A = 1\%$	0.500	0.0	19.66	0.4	854.25	9.52	1.11	52.2
$\Delta A = 10\%$	0.474	5.2	18.93	4.11	1950.0	150.2	2.58	36.1
$\Delta A = 30\%$	0.417	16.6	16.75	15.2	5644.0	624.3	8.39	28.0
nfft = 11, $\delta x = 0.00307$								
$\Delta A = 1\%$	0.499	0.2	19.67	0.4	1505.0	93.2	1.94	57.25
$\Delta A = 10\%$	0.465	6.9	18.61	5.72	1.626E4	1987	21.85	40.28
$\Delta A = 30\%$	0.387	22.7	15.36	22.2	6.801E4	8628	111.39	32.37
nfft = 12, $\delta x = 0.001534$								
$\Delta A = 1\%$	0.500	0.0	19.65	0.45	3107.9	298.8	4.0	60.03
$\Delta A = 10\%$	0.450	9.6	18.06	8.52	4.958E4	6.26E3	68.75	44.0
$\Delta A = 30\%$	0.330	34.78	12.88	34.77	1.935E5	2.47E4	380.6	35.68

SG: Savitzky–Golay.

Notes: Values of the exact spectral moments and their numerically calculated values with relative errors for various noise amplitudes, ΔA, and resolutions, δx. Given also are the bandwidth parameter, α, and the signal to noise ratio in (dB). For all cases, $A = 1$ m and $f = 1$ Hz.

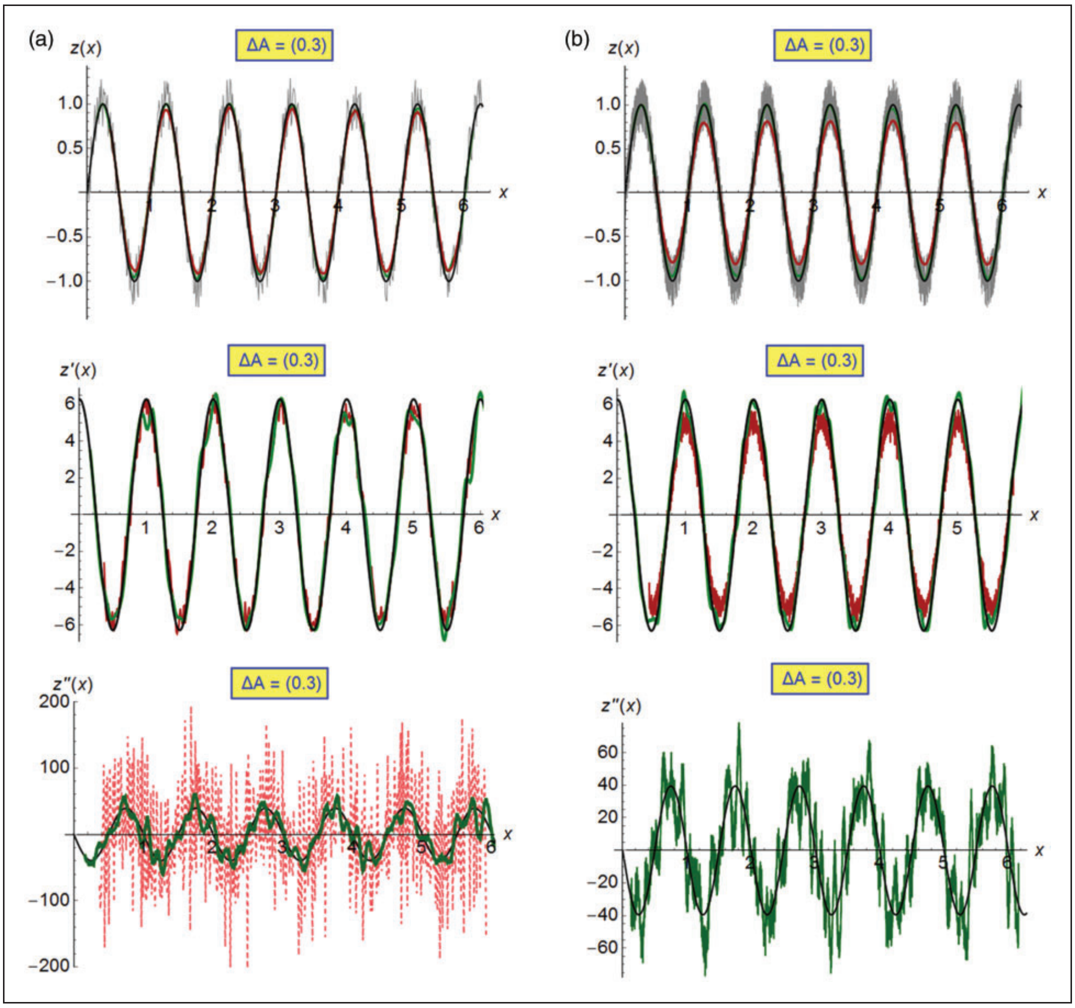

Figure 2. Signal, $z(x)$, first derivatives, $z'(x)$, and second derivatives, $z''(x)$, for two resolutions: (a) *nfft* = 9 and (b) *nfft* = 12. Color keys: gray = noisy signal generated by equation(9) with $\Delta A = 0.3$; black = pure signal using equation (9) with $\Delta A = 0$; red = SG filtered; and green = G-EXP filtered.
SG: Savitzky–Golay; G-EXP: generalized exponential.

data sets filtered individually; and, vice versa, subtraction a certain filtered data set from the filtered total, would reveal the remainder intact, and (4) similar to the aforementioned filters, it should contain coefficients or weights that are symmetric about the midpoint. That is a necessity as there should not be preference to neighboring points on either side of a point of interest undergoing smoothing.

On the other hand, the G-EXP filter is significantly different from the SG filter in some major ways. The SG filter uses polynomials of fixed order with fixed coefficients (i.e. fixed windowed weights), while the G-EXP filter is flexible using any positive real parameters, allowing the weights to assume whichever designed values.

The G-EXP filter can be thought of belonging to a general exponential form

$$g(x) = e^{-\beta|x|^{n}} \quad @\ x \in (-L, L) \tag{13}$$

This generalization allows for a three-parameter filter design where the power, n, can take on any positive real value. A changeable power, n, provides a whole family of exponential filters. The absolute of $|x|$ allows the power calculation, while preserving filter symmetry for any power, n. If $n = 0$, then $g(x)$ degenerates to a Dirichlet filter. When $n = 1$, then $g(x)$ is strictly exponential, and when $n = 2$, then $g(x)$ is related to the *standard normal distribution*,[30–32] retaining only the quadratic exponential form (The standard normal distribution is: $P(x) = \left(1/\sqrt{2\pi}\right)e^{-x^2/2}\ @x \in (-\infty, \infty)$. First, the leading coefficient of $1/\sqrt{2\pi}$ is of no consequence, because the G-EXP filter as given subsequently by equation (14) is normalized. Then, the parameter β is free to take on any positive real value (i.e. other than ½). Finally, the range $x \in (-L, L)$, i.e. the filter window size, is selectable or adjustable.), in which case the absolute of x can be omitted. The filter is actually completely stated by

$$g = e^{-\beta|Range[-L,L]|^{n}} / Total\left[e^{-\beta|Range[-L,L]|^{n}}\right] \tag{14}$$

In the current work, $n = 1$ and $n = 2$ are tried, and found to smooth the signals as intended. However, because of the superior results rendered by the filter with $n = 2$, other values of n had been bypassed. Hence, for the remainder of this work, the G-EXP is specific to the case of $n = 2$. In which case, equation (13) degenerates to

$$g(x) = e^{-\beta x^2} \quad @ \; x \in (-L, L) \qquad (13 - \mathrm{a})$$

So now, the G-EXP filter has two parameters left that can be adjusted. Specifically, the tuning parameter, β, is any real positive value, while the integer, L, decides the window size. Again, using Mathematica's syntax, the G-EXP filter, given in equation (14), is entirely expressed by a single statement

$$g = e^{-\beta Range[-L,L]^2} / Total\left[e^{-\beta Range[-L,L]^2}\right] \qquad (14 - \mathrm{a})$$

The advantage of the G-EXP filter is that L and β are generally not restricted, and in addition to n, they allow for a "three-degree of freedom" filter. The only restriction on L is that it is sufficiently smaller than the number of points in the signal to be smoothed, i.e. $L << N$, which should commonly be the case. Additional construction details of the G-EXP filter, along with examples, are given in Appendix C. For portability, the appendix provides also a Fortran 77 code for the construction and execution of the G-EXP filter.

We turn now to the outcomes of applying the G-EXP filter. As described above, similar to the application of the SG filter, a Fibonacci optimization algorithm is executed to assist in the parameter selection. The process is to try a set of window sizes, L, and then let the optimization algorithm determine the optimal β for each one of them. Of all sets of L and β, the one that produces the smallest normalized norm between the smoothed signal spectral moments and the known theoretical values is selected. Again, herein, that procedure can be executed because the target values are known a priori. Once the noisy signal, z, is smoothed by the G-EXP filter, derivatives are taken, and finally equations (1) to (3) are executed. Table 3 shows the same information as in Tables 1 and 2, but the results pertain to signals smoothed by the G-EXP filter. It should be noted that because the G-EXP is a linear filter, the smoothed signal is subtracted from the raw noisy signal, leaving only the effective noise. Hence, the SNR_{dB} can likewise be calculated according to equations (10) and (11).

Figure 2 depicts in green the smoothing results for the largest noise level case, $\Delta A = 0.3$, using the two-statement hand-coded filter as detailed in Appendix C using the Mathematica package. The smoothed signal (shown in green) is compared to the raw noisy signal (shown in gray) and the pure objective signal (shown in black). For comparison also the results of the SG filter are shown there (in red). It is quite clear that the G-EXP filter is very effective in smoothing $z(x)$. Moreover, the derivative, $z'(x)$, follows the objective derivative quite closely, with much less noise than the SG filtered signal. But the most startling finding is that the second derivative, $z''(x)$, while in itself is still somewhat noisy, the magnitude and behavior follow closely the target behavior and magnitude. It is emphasized that at this point, no more derivatives are taken, so that noisy behavior in the second derivative is of no consequence. Particularly, when equations (1) to (3) are executed, averaging takes place, which has an additional smoothing effects (much like a moving average).

Two resolutions are also examined in Figure 2, *nfft* = 9 and *nfft* = 12. The findings are again startling. Examining the results in Table 3, as the resolution gets finer, the errors get smaller. Even the bandwidth

Table 3. Results using the G-EXP filter ($n = 2$).

	$m_0 = 1/2 = 0.5$		$m_2 = 2\pi^2 = 19.739$		$m_4 = 8\pi^4 = 779.27$			
At $\Delta A = 0$	Value	Error%	Value	Error%	Value	Error%	$\alpha = 1$	$SNR_{dB} = \infty$
nfft = 9, $\delta x = 0.01227$								
$\Delta A = 1\%$	0.493	1.5	19.42	1.61	791.8	1.61	1.03	49.9
$\Delta A = 10\%$	0.472	5.5	18.65	5.52	822.2	5.51	1.12	33.5
$\Delta A = 30\%$	0.451	9.8	18.01	8.74	855.6	9.79	1.19	25.2
nfft = 11, $\delta x = 0.00307$								
$\Delta A = 1\%$	0.496	0.7	19.46	1.43	790.4	1.42	1.04	54.2
$\Delta A = 10\%$	0.480	3.94	18.88	4.33	812.9	4.31	1.09	37.0
$\Delta A = 30\%$	0.461	7.7	18.31	7.22	839.3	7.70	1.15	28.6
nfft = 12, $\delta x = 0.001534$								
$\Delta A = 1\%$	0.500	0	19.47	1.37	789.9	1.37	1.04	56.6
$\Delta A = 10\%$	0.480	3.24	19.03	3.58	807.0	3.56	1.08	39.2
$\Delta A = 30\%$	0.470	5.88	18.71	5.22	856.5	9.91	1.15	30.7

G-EXP: generalized exponential.

Note: Values of the exact spectral moments and their numerically calculated values with relative errors for various noise amplitudes, ΔA, and resolutions, δx. Given also are the bandwidth parameter, α, and the signal to noise ratio in (dB). For all cases, $A = 1\ m$ and $f = 1\ Hz$.

parameter, α, is getting better as the resolution gets finer, being very close to the ideal value of 1. The SNR_{dB} is also better compared to those reported in Tables 1 and 2.

In summary, the G-EXP filter makes a remarkable discovery of the underlying geometry of the pure signal. The filter recovers almost to perfection the ideal spectral moments and the bandwidth parameter. There is no longer a resolution problem, in fact, as the resolution gets finer, the results get better. All this is true for tiny, moderate, and fairly large magnitudes of noise, ΔA. The SNR_{dB} increases to acceptable values. It can be concluded that the G-EXP filter provides a very effective solution for noisy signals.

In addition to that objective, the G-EXP filter is shown to be a general tool for smoothing out noisy signals not only in the time domain but also in the frequency domain. The latter capability is demonstrated by the green color line in Figure 3 in Appendix A, where the G-EXP filter is successfully used to smooth out also the power spectrum. Moreover, the G-EXP filter can be applied repeatedly in what is known as "passes." Each pass tends to remove even more noise, exposing more of the underlying signal. For some filters (e.g. the G-EXP), the process removes the noise very effectively even by using a single pass. But like other filters, additional passes may increase the possibility of signal distortion, loss of information, and more importantly, L points on each side of the signal are lost after each pass. That is, N must be much greater than L. In this work, only one pass is applied to test the various filters without the bias of repeated passes.

Application of the G-EXP filter on real and fractal rough surfaces

The challenge in calculating the spectral moments for real surfaces stems from the fact that the target values are not known a priori. A trial-and-error process is needed to find the filter parameters. The G-EXP filter, with $L=20$, and $\beta=5$, is used upon two real rough surfaces of a ceramic spherical indenter loaded against a multi-wall carbon nanotube counterpart.[20] The GW[1] model is subsequently executed. For brevity, details of that application and measuring techniques are spared. The important and relevant fact herein is that the roughness of each surface is *3D*, not homogeneous, and not isotropic. The said paper[20] details how such surface characteristics are handled by pursuing McCool's[6] procedure. That involves finding sets of two two-dimensional (*2D*) orthogonal directions of maximum and minimum spectral moments that are averaged either arithmetically or harmonically. Hence, the procedure developed herein for a *2D* case is specifically suitable to handle *3D* surface roughness. It is not the intent herein to repeat the calculation procedure other than to exhibit how the G-EXP is applied to such *2D* rough surfaces, as shown in Figure 3. As can be seen, some of the data of the original signal measurements are exceedingly far off from the "norm." These are either defects in the surfaces, voids or bumps in the material, or they are equipment related, producing erroneous measurements. Should the original data are taken "as is" to calculate the spectral moments according to equations (1) to (3), enormous errors would result, as it takes only one "bad" point, or a region of "bad" points to skew the results very profoundly. Filtering out those data points that "do not belong" is undoubtedly essential.

The G-EXP filter can be applied equally well also to surface having fractal roughness. That, however, is not necessary, as fractal roughness contains no noise. As discussed above, the signal is entirely deterministic for given A, D, and γ. Consider equation (12) modified to have a truncated sum having an upper summation bound, n_2

$$z(x)=A^{(D-1)}\sum_{n=n_1}^{n_2}\frac{\cos 2\pi\gamma^n x}{\gamma^{(2-D)n}} \quad 1<D<2,$$
$$\gamma>1, \quad n_2\geqslant n_1$$

The value of n_2 is related to the finest resolution (i.e. highest frequency) capability of the measuring equipment.[24] The signal would appear "*smooth*" or "*rough*" depending on how "*small*" or "*large*" is n_2, respectively. Hence, reducing the value of n_2 smoothens the signal mathematically without filtering. Fractal roughness deserves a completely different treatment that is beyond the scope of this work.

Conclusions

The upside of the GW model is that only three spectral moments are needed to execute it. That is also the downside of the GW model, as not too many parameters are available to work with. To estimate the spectral moments reliably, the surface roughness needs to undergo massive data reduction. So, the estimation must be quite good with almost zero room for error.

In the current work, a sine waveform is contaminated by a white noise process varying in magnitude from tiny to moderate. To recover the underlying geometry, these methods have been tested to calculate the spectral moment: (1) finite-difference derivatives using accurate five-points (instead of two- or three-points as suggested by McCool), (2) data partitioning and averaging (to reduce statistical errors), (3) calculation of derivatives via interpolation functions, and (4) calculation via the power spectrum. The following conclusions can be drawn:

1. The finite difference method as suggested by McCool to calculate the spectral moments fails dramatically on his own problem when the waveform is contaminated even by the tiniest noise, and matters become worse with moderate to higher noise levels.

2. The zeroth spectral moment does not contain derivatives in its definition and, hence, it is less sensitive to noises in the signal, i.e. its prediction by equation (1) is quite good and can be trusted.
3. The second spectral moment depends on the first derivative, and hence a noisy signal affects its accuracy because derivatives tend to accentuate errors.
4. The fourth spectral moment depends on the second derivative (theoretically, it is the derivative of the first derivative, which is already erroneous, although herein it is calculated directly by a five (5) finite difference, without resorting to a "derivative of the first derivative"). The errors in the calculation of the fourth spectral moment are much worse than even the second moment.
5. A finer resolution makes things even worse, but that is attributed to the presence of the noise. Once the noise is attenuated, a higher resolution produces better predictions for the spectral moments.
6. Signal conditioning is a necessity for the removal of noise in the data. Various filters have been tried, but most have not greatly improved upon the calculation of the spectral moments. The venerable SG filter is found to make progress in that calculation, but still the results are objectionable.
7. The only filter that is successful in restoring the underlying geometry and removing the noise effectively (even when it is of a relatively high magnitude) is the G-EXP filter. Upon smoothing, the predictions of the spectral moments and the bandwidth parameter are extremely close to the theoretical values. The SNR improves considerably upon the application of the G-EXP filter.
8. The G-EXP is applied upon real surfaces in eight different directions, and it is shown to smoothed out the enormous noise levels very effectively.
9. The G-EXP filter can be applied not only on the waveform, but also on the power spectrum.
10. The G-EXP filter is a linear filter and, hence, superposition can be applied and taken advantage of so that the noise can be isolated.
11. The G-EXP can be applied in successive passes, if deemed necessary.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Greenwood JA and Williamson JBP. Contact of nominally flat surfaces. *Proc R Soc London Ser A. Math Phys Sci* 1966; 295: 300–319.
2. Jackson RL and Green I. On the modeling of elastic contact between rough surfaces. *Tribol Trans* 2011; 54: 300–314.
3. Chang WR, Etsion I and Bogy DB. An elastic-plastic model for the contact of rough surfaces. *J Tribol* 1987; 109: 257.
4. Jackson RL and Green I. A statistical model of elasto-plastic asperity contact between rough surfaces. *Tribol Int* 2006; 39: 906–914.
5. Kogut L and Jackson RL. A comparison of contact modeling utilizing statistical and fractal approaches. *J Tribol* 2006; 128: 213.
6. McCool JI. Relating profile instrument measurements to the functional performance of rough surfaces. *J Tribol* 1987; 109: 264.
7. McCool JI. Finite difference spectral moment estimation for profiles the effect of sample spacing and quantization error. *Precis Eng* 1982; 4: 181–184.
8. Sweitzer K, Bishop N and Genberg V. Efficient Computation of Spectral Moments for Determination of Random Response Statistics. In: *Proceedings of ISMA 2004, Signal Processing and Instrumentation*, 2004, pp. 2677–2691.
9. Davidson KL and Loughlin PJ. Instantaneous spectral moments. *J Franklin Inst* 2000; 337: 421–436.
10. Vogel F. Spectral moments and linear models used for photoacoustic detection of crude oil in produced water, Department of Informatics, University of Oslo, Oslo, Norway, 2001.
11. Brown SR. Simple mathematical model of a rough fracture. *J Geophys Res: Solid Earth* 1995; 100: 5941–5952.
12. Bendat JS and Piersol AG. *Measurement and analysis of random data*. New York, NY: Wiley, 1966.
13. Bendat JS and Piersol AG. *Engineering applications of correlation and spectral analysis*. New York, NY: John Wiley and Sons, Inc., 1980.
14. Bendat JS and Piersol AG. *Random data: analysis and measurement procedures*. Hoboken, NJ: John Wiley & Sons, 2011.
15. Pawar G, Pawlus P, Etsion I, et al. The effect of determining topography parameters on analyzing elastic contact between isotropic rough surfaces. *J Tribol* 2012; 135: 011401.
16. Xu Y and Jackson RL. Statistical models of nearly complete elastic rough surface contact – comparison with numerical solutions. *Tribol Int* 2017; 105: 274–291.
17. Kalin M, Pogačnik A, Etsion I, et al. Comparing surface topography parameters of rough surfaces obtained with spectral moments and deterministic methods. *Tribol Int* 2016; 93: 137–141.
18. Nayak PR. Random process model of rough surfaces. *J Lubr Technol* 1971; 93: 398–407.
19. Bhushan B. *Surface roughness analysis and measurement techniques. Modern tribology handbook, two volume set*. Boca Raton, FL: CRC Press, 2000.
20. Reinert L, Green I, Gimmler S, et al. Tribological behavior of self-lubricating carbon nanoparticle reinforced metal matrix composites. *Wear* 2018; 408: 72–85.
21. Hildebrand FB. *Introduction to numerical analysis*. New York: McGraw-Hill, 1974.
22. Pogačnik A and Kalin M. How to determine the number of asperity peaks, their radii and their heights for engineering surfaces: a critical appraisal. *Wear* 2013; 300: 143–154.

23. Hariri A, Zu JW and Mrad RB. N-point asperity model for contact between nominally flat surfaces. *J Tribol* 2006; 128: 505–514.
24. Majumdar A and Tien CL. Fractal characterization and simulation of rough surfaces. *Wear* 1990; 136: 313–327.
25. Berry MV and Lewis ZV. On the Weierstrass–Mandelbrot fractal function. *Proc R Soc A* 1980; 370: 459–484.
26. Press WH, ed., *FORTRAN numerical recipes*. Cambridge, UK; New York, NY: Cambridge University Press, 1996.
27. IMSL. STAT/Library: FORTRAN subroutines for statistical analysis: user's manual. Houston, TX: IMSL, Inc., 1987.
28. Savitzky A and Golay MJE. Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem* 1964; 36: 1627–1639.
29. Antón-Acedos P, Sanz-Lobera A, López-Baos A, et al. Feasibility analysis of Savitzky-Golay filter implementation in surface texture filtering and measurement. *Procedia Manuf* 2017; 13: 503–510.
30. Kenney JF. *Mathematics of statistics*. New York, NY: Van Nostrand, 1964.
31. Weisstein E. Standard normal distribution – from Wolfram MathWorld (Online), http://mathworld.wolfram.com/StandardNormalDistribution.html (accessed 10 May 2019).
32. Abramowitz M and Stegun IA. *National Bureau of Standards: applied mathematics series*. vol. 55. Gaithersburg, MD: National Bureau of Standards, 1972, p.1060.

Appendix A. The spectral moments via the power spectrum

If the procedure for obtaining the spectral moments is problematic using numerical differentiation, perhaps using equation (4) on noisy signals can produce a desirable solution. So, the next step is to obtain the power spectrum of the signal investigated herein. It is noted the noise is superimposed upon the waveform. Thus, according to equation (9), we have a signal of

$$z(x) = [A\sin(2\pi f x) + noise]_{A=1,f=1} = \sin(wx) + noise \tag{15}$$

where *noise* is a random process as discussed above. Note that $w = 2\pi$, is the specific frequency of the waveform of equation (5). The power spectrum is defined by the Fourier transform for a continuous function

$$Z(\omega) = \int_{-\infty}^{\infty} z(x) e^{-2\pi i x\omega}\,\mathrm{d}x \tag{16}$$

where here, and in equation (4), ω is a general frequency in the entire spectrum, i.e. $\omega \in [-\infty, \infty]$. Since the Fourier transform is a linear operator, and had $z(x)$ in equation (12) been a continuous function then

$$Z(\omega) = Z(wx) + Z(noise) \tag{17}$$

However, the noise herein is a *white noise process* with a uniform distribution. By definition, its Fourier transform is "*flat*" and equals to zero, i.e. $Z(noise) = 0$ throughout the entire spectrum regardless of the amplitude ΔA (see equation (9)). That leaves

$$Z(\omega) = Z(wx) = i\sqrt{\pi/2}[\delta(\omega - w) - \delta(\omega + w)] \tag{18}$$

where $\delta(*)$ is the Dirac delta function. The power of the signal is

$$\begin{aligned} P(\omega) = |Z(\omega)|^2 &= (\pi/2)[\delta(\omega - w) - \delta(\omega + w)]^2 \\ &= (\pi/2)\big[\delta^2(\omega - w) - 2\delta(\omega - w)\delta(\omega + w) \\ &\quad + \delta^2(\omega + w)\big] \end{aligned} \tag{19}$$

Indeed, the discussion can be limited to just positive frequencies $\omega \in [0, \infty]$. This result for the power needs to be substituted in equation (4) to calculate the spectral moments $k = 0, 2, 4$. However, when equation (19) is substituted in equation (4), an exact mathematical solution for the latter is elusive. That mathematical difficulty compelled the workarounds and approximations made by Majumdar and Tien.[24,25] Those workarounds contain biases and the approximated moments cannot be regarded exact solutions.

To further illustrate the difficulty with the spectrum approach, a numerical fast Fourier transform (FFT) is executed upon the discrete values of the contaminated signal of $z(x)$ (given by equation (9)). The power spectrum is obtained, and the spectral moments (following equation (4)) are calculated numerically using either a trapezoidal or the Simpson rule (both giving very similar results). That approach is tried for the noise level case of $\Delta A = 30\%$, and $nfft = 9$, where the power spectrum is shown in Figure 4. While the spectrum captures well the specific dominant frequency of the waveform (see the peak at $f = 1$ or $\omega = w = 2\pi$), the final results for $\{m_0, m_2, m_4\} = \{0.5012, 149.2, 7.466E6\}$ are inconsequential. First, the power spectrum shown in Figure 3 cannot represent the exact solution that equation (19) commands. Second, results for m_2 and m_4 are in great error and intolerable (similar or even worse than those appearing in Table 1).

Appendix B. Numerical derivatives

For simplicity, a three-point central derivative will be used to show the reason for the increasing error in the computation of m_2 and m_4. Starting off with the first derivative (omitting the truncation error),

$$z'(x) = \frac{z(x + \delta x) - z(x - \delta x)}{2\delta x} \tag{20}$$

Theoretically, for $A = 1$ and $f = 1$, at $x = k/4$ for $k = 1, 3, 5\ldots$, the derivative of equation (5) equals

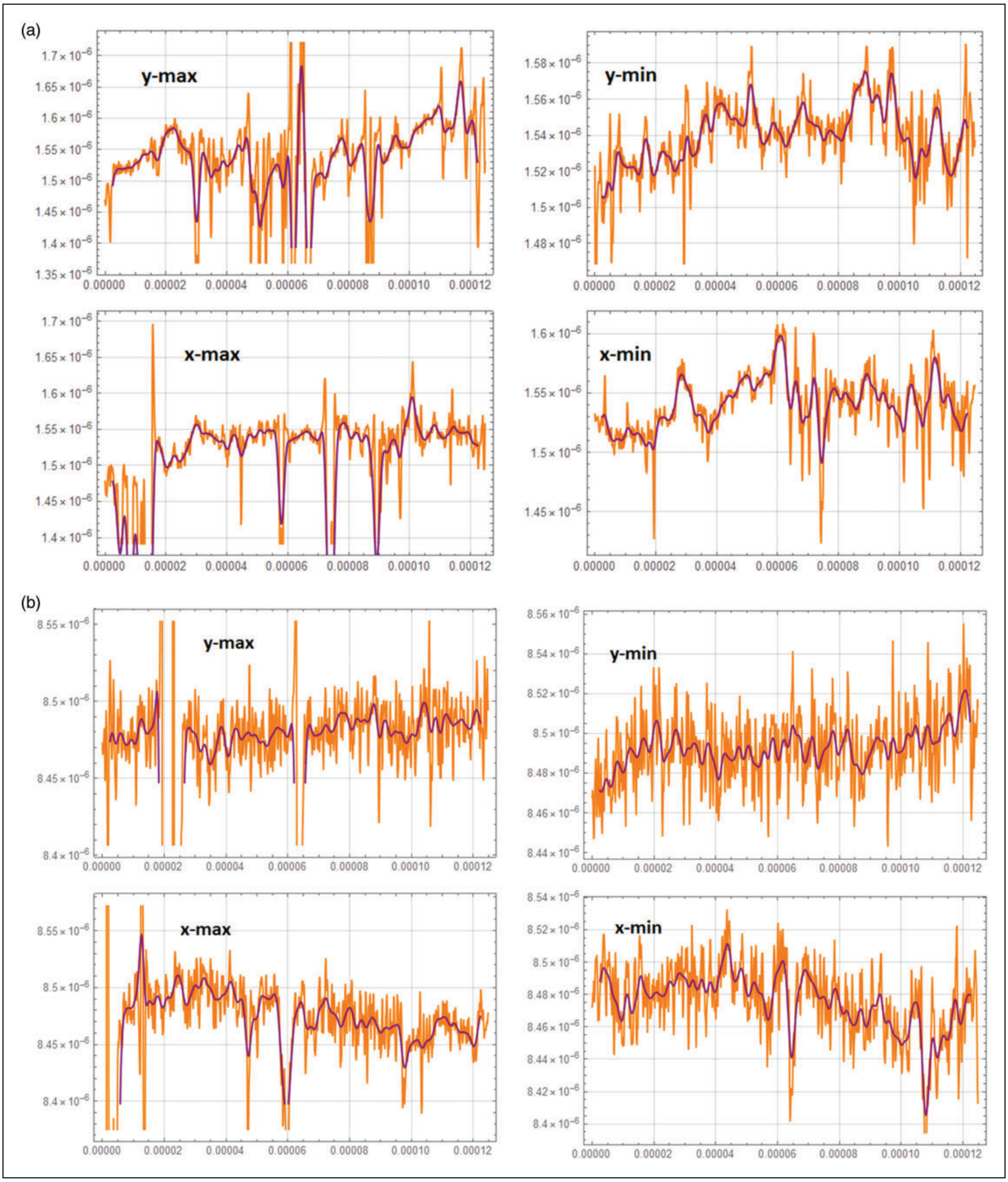

Figure 3. Real rough surfaces from Reinert et al.[20] Abscissa and ordinate are shown in *m*. (a) Composite carbon nanotube and (b) Ceramic ball counterpart.

Color key: Orange = the original rough surfaces, purple = the G-EXP smoothed surfaces, with $L = 20$ and $\beta = 5$.

Source: reproduced with permission from Reinert et al. 2018.[20]

G-EXP: generalized exponential.

zero. Indeed, when the pure equation (5) is digitized discretely, $z(x - \delta x) = z(x + \delta x)$ at the said points, and the numerical derivative turns out the correct zero result. However, using equation (20) on the noisy $z(x)$ at those points, $z(x - \delta x) \neq z(x + \delta x)$ because of the added random noise, thus clearly rendering a result that is *not* zero. That phenomenon happens not only at the selected x values, but actually at any value of $x \in \{0, x_{max}\}$ along the signal. As noise is modulated by ΔA, then when the sum is employed on $(z'(x))^2$ according to equation (2), the accumulation of the error only intensifies.

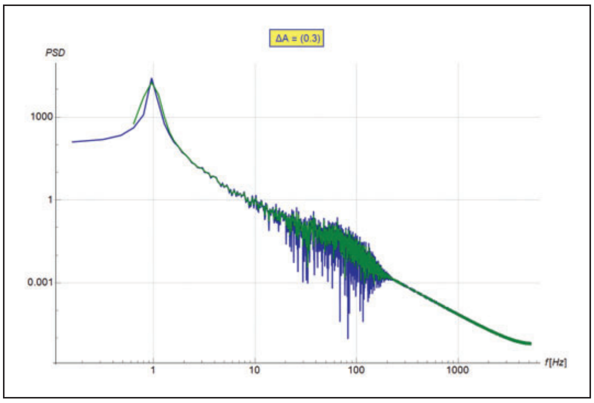

Figure 4. Power spectrum for a sine waveform contaminated by a white noise level of $\Delta A = 30\%$, shown in blue color. A smoothed spectrum is shown in green.

The second central numerical derivative is

$$\begin{aligned} z''(x) &= \frac{z(x-\delta x) - 2z(x) + z(x+\delta x)}{\delta x^2} \\ &= \frac{[z(x+\delta x) - z(x)] - [z(x) - z(x-\delta x)]}{\delta x^2} \\ &= \frac{[z(x+\delta x) - z(x)]/\delta x - [z(x) - z(x-\delta x)]/\delta x}{\delta x} \end{aligned} \tag{21}$$

The second form of equation (21) emphasizes that the second derivative (like the first derivative) is based on the difference of the differences between the neighboring z(x) values, while the third form of the equation renders the expected result that the second derivative is, of course, a derivative of the first derivative. Hence, if the first derivative contains errors clearly, the second derivative must be erroneous too.

So again, theoretically, for $A = 1$ and $f = 1$, at $x = k/4$ for $k = 0, 2, 4 \ldots N$ when the digitized pure signal of equation (5) is used, the numerical second derivative equals identically zero, as it should be. However, when the noisy $z(x)$ values are substituted into equation (21), the results are *not* zero. And as explained above, that behavior happens actually at any value of $x \in \{0, x_{max}\}$ along the signal. And as the error is modulated by ΔA, then when the sum is executed on $(z''(x))^2$ according to equation (3), the accumulation of the error escalates dramatically.

The analysis above holds for any central difference *of any order*, including order five (5) that is used herein. The conclusion is that for m_0, which does not include derivatives in its definition, the estimation by equation (1) can be considered "*correct*" or "*sufficiently accurate*" as the noise contaminates the magnitude of the signal, but if the noise amplitude is relatively small, then m_0 shall have only a corresponding small error. However, m_2 and m_4 depend on the derivatives, which happen on the differences between $z(x)$ values (so the relative magnitudes of $z(x)$ values themselves are irrelevant). Hence, the errors in the derivatives are directly proportional to the magnitude of the noise, and that cannot be mitigated. In other words, m_0 hinges on "macro" or "global" quantities where the noise effects are less significant, while m_2 and m_4 hinge on "micro" or "local" quantities where the noise effects are very significant.

Appendix C. The construction of the G-EXP filter

The filter is best explained along with an example. The filter is defined by equation (13)

$$g(x) = e^{-\beta|x|^n} \quad @\ x \in (-L, L); \quad x = Range[-L, L] \tag{13}$$

Suppose that $L = 3$, then using a Mathematica statement: x = *Range [–L, L]* results in a vector $\{x\} = \{-3, -2, -1, 0, 1, 2, 3\}$. Here, the vector $\{x\}$ has no physical meaning; it is a dummy list (or a service vector) of length of $2L + 1$, and it is used for illustration only. If also $n = 2$, and $\beta = 0.5$, then upon substitution into equation (13), we have

$$\{\mathrm{g}\} = \{0.0111,\ 0.1353,\ 0.6065,\ 1.,\ 0.6065,\ 0.1353,\ 0.0111\}$$

Clearly, by definition, the vector $\{g\}$ also has an odd length, $2L + 1$. The values it contains are symmetric about the center point. These values have the role of weights, where the center weight has the largest value. The next step is to ensure that the smoothed signal does not overshoot, i.e. worsen the signal. Hence, the vector {g} is normalized by its total, so that the largest central weight equals to 1/Total[g]. Hence, issuing the computer assignment

$$\mathrm{g} = \mathrm{g}/\mathrm{Total}[\mathrm{g}] \tag{22}$$

achieves that goal. Instead of the two computer assignments expressed by equations (13) and (22), the filter design is reduced to a single Mathematica statement, as given by equation (14)

$$g = e^{-\beta|Range[-L,L]|^n} / Total\left[e^{-\beta|Range[-L,L]|^n}\right] \tag{14}$$

The numerator of equation (14) contains a list, which is normalized by its own total, and it is always symmetric about the origin. Clearly, the dummy variable "x" disappeared, because it is immaterial for the formal filter design, but it has utility for illustration purposes. Therefore, the filter given by equation (14) contains a list of weights that total to the value of one unit. On the said example, we have

$$\begin{aligned} \{\mathrm{g}\} &= \{\mathrm{g}\}/\mathrm{Total}[\mathrm{g}] \\ &= 10^{-3} \times \{4.43305, 54.0056, 242.036, 399.05, 242.036, 54.0056, 4.43305\} \end{aligned}$$

The filter is shown in Figure 5, for $\beta = 1/2$ *and* $L = 3$, but for various values of n. Note that $n = \pi$ is used to highlight the fact that n can take on rational or irrational values. It is specifically emphasized that the filter $\{g\}$ contains just the list of values marked by the bold points only (the continuous line-plots are shown for illustration only).

Even though values in the range $0 < n < 1$ are admissible, it is improbable that they would produce effective filtering. The power of $n = 1$ had been tried herein in addition to $n = 2$. While both filter cases perform as intended in smoothing the noisy signals, on the current problem statement, the filter with $n = 2$, always produces superior results compared to a filter with $n = 1$. Trial-and-error or Fibonacci optimization processes can be employed to determine "*better*" or "*best*" exponents, n.

The execution of the filter at this point proceeds by convolving the signal with the filter. Procedurally, filtering is done as follows. Suppose that the vector $\{z_s\}$, $s = 1, 2, \ldots, N$ contains the equidistant values of $z(x)$. Suppose that $\{g_r\}$, $r = 1, 2, \ldots, 2L+1$, contain the weight values of a G-EXP filter, then the convolution is executed by

$$b_s = \sum_{r=1}^{2L+1} g_r z_{s+r-1} \quad \textit{for all } s = 1, 2, \ldots, N - 2L \tag{23}$$

resulting in the smoothed signal, b_s. As an example, suppose that

$$\{g\} = \{g2,\ g1,\ g0,\ g1,\ g2\}$$
$$\{z\} = \{z1,\ z2,\ z3,\ z4,\ z5,\ z6,\ z7\}$$

Here, $L = 2$ and $N = 7$. Note that $\{g\}$ has symmetric values about the center value (as discussed above). Equation (23) can be easily coded in any desired programing language (e.g. Fortran or C++) in a nested do-loop. In Mathematica, equation (23) is implemented by

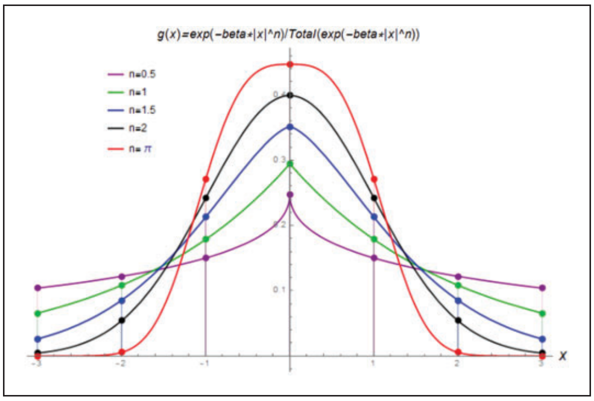

Figure 5. G-EXP filter for $\beta = 1/2$, $L = 3$, at various n. G-EXP: generalized exponential.

$$bs = Table[Sum[g[[r]] * z[[s + r - 1]],\ \{r, 1, 2 * L + 1\}],\ \{s, 1, N - 2 * L\}] \tag{24}$$

which results in the vector b_s

$$\{bs\} = \begin{Bmatrix} g2z1 + g1z2 + g0z3 + g1z4 + g2z5 \\ g2z2 + g1z3 + g0z4 + g1z5 + g2z6 \\ g2z3 + g1z4 + g0z5 + g1z6 + g2z7 \end{Bmatrix} \tag{25}$$

Taking advantage of the symmetry of $\{g\}$, an intrinsic function in Mathematica can conveniently be used instead to yield the same result as above

$$bs = ListConvolve[g,\ z] \tag{24 – a}$$

So, to smooth an equidistant noisy signal, $\{z\}$, construct first the filter using equation (14), and then apply it using equation (24). For portability, a simple Fortran 77 code is also given in Table 4 at the end of this Appendix showing the filter construction and the convolution unfolding.

The spacing (i.e. resolution) of the smoothed vector {bs} is not affected by the filter as the filter operates on the magnitudes of $\{z\}$ alone; hence, spacing is identical to that of the original vector $\{z\}$. However, the length of $\{bs\}$ is reduced to N-$2L$, compared to the length of $\{z\}$, which is N. This is common also to other filters, such as the SG. Clearly, in an actual case, N is considerably larger than L, such that only a few values at the two ends are missing. This normally does not hinder the usefulness of filtering in general. But even that problem can be overcome by padding, or imposing cyclic behavior of z (for brevity, these approaches are omitted, as they are secondary in this development).

Now the focus turns to β. Consider a certain data point s in the signal, which will be weighted the most by the center weight *(g0)*, while the neighboring data points $s - 1$ and $s + 1$ will be multiplied by a reduced neighboring weight *(g1)*, and so on. Had $\{g\}$ been a vector with a length of *1*, then the point of interest in the signal would be multiplied by a weight of one unit, i.e. no smoothing takes place. So the normalization guarantees that when more weights are used the new smoothed points do not overshoot. As indicated, the filter consists of a list of values that can be regarded as weights or coefficients. Noteworthy, these coefficients are not fixed (contrary to those in the SG or other filters). Varying β can produce any weights desired. To appreciate that, examine Figure 6, where $n = 2$, $L = 10$, while β takes on three different values.

It is seen that for a given n, L decides the window length, while β decides the sharpness. Note, however, that large β can lessen the effective window length (as seen with $\beta = 0.1$, the effective window length reduces

Table 4. A Fortran 77 code to produce and execute the G-EXP filter for the example above.

```
parameter (n = 7, L = 2, ng = 2*L + 1, nb = n-2*L)
dimension g(ng), z(n), bs(nb)
data z/1.,2.,3.,4.,5.,6.,7./
data beta/0.5/

write(6,*) “Construct the filter as in equation (13)”
sum = 0.0
do i = 1,ng
  x = i-(ng + 1)/2
  g(i) = exp(-beta*x**2)
  sum = sum + g(i)
  write(6,*) i,’ ‘,x, ‘ ‘, g(i)
enddo

write(6,*) “Normalize the filter as in equation (22)”
do i = 1,ng
  g(i) = g(i)/sum
  write(6,*) i,’ ‘,g(i)
enddo

write(6,*) “Execute the convolution given
in equation (24) or (24-a)”
do is = 1,nb
  bs(is) = 0.0
    do ir = 1,2*L + 1
      bs(is) = bs(is) + g(ir)*z(is + ir-1)
    enddo
  write(6,*) is,’ ‘,bs(is)
enddo

write(6,*)’Verify results expected (in paper): for L = 2 and any
n’
do is = 1,nb
i = is-1
write(6,*) is,’ ‘,
g(1)*z(1 + i) + g(2)*z(2 + i) + g(3)*z(3 + i) +
g(4)*z(4 + i) + g(5)*z(5 + i)
enddo
end
```

G-EXP: generalized exponential.

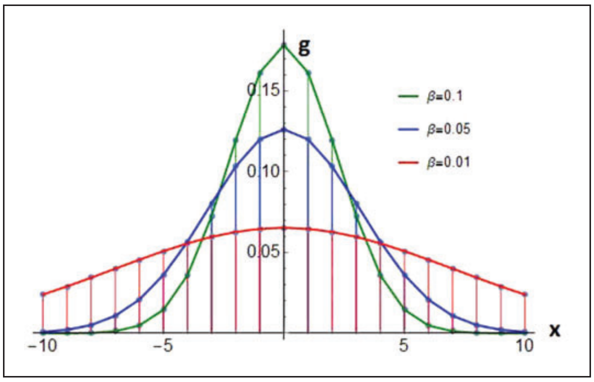

Figure 6. The G-EXP filter for $n = 2$, $L = 10$, and three values of β.
G-EXP: generalized exponential.

to 6 because at point 7 and above, the {g} values approach zero on both ends). So while the two parameters L and β provide great flexibility in the filter design, a meticulous trial-and-error process is normally entailed in their selection. The general trends are: as L gets larger, more neighboring points participate in the smoothing, where a larger β puts more weight on the closest neighbors and the extent of smoothing is reduced (and vice versa for smaller values of L and β). Clearly L should be sufficiently smaller than the number of data points, N, to be smoothed.