A hybrid method for fast and efficient rough surface generation

Arthur Francisco, Noël Brunetière

A Hybrid Method For Fast and Efficient Rough Surface Generation

Arthur Francisco1)* and Noël Brunetière2)

1. Département Génie Mécanique et Systèmes Complexes, Institut Pprime, University of Poitiers, 4 avenue de Varsovie, 16021 Angoulême Cedex, France
2. Département Génie Mécanique et Systèmes Complexes, Institut Pprime, University of Poitiers, Téléport 2 Boulevard Marie et Pierre Curie, BP 30179 Futuroscope Chasseneuil Cedex, France

*Corresponding author: arthur.francisco@univ-poitiers.fr

Abstract

Since many years, the pioneering work of Hu and Tonder is used to generate rough surfaces with prescribed statistical moments (skewness and kurtosis) along with spatial properties (correlation lengths). The present work enlightens the drawbacks of this method and it proposes an original approach based on a hybrid analytical/numerical method.

Simulations are conducted on very different surface specimens and the method is validated over a wide range of statistical moments. The results are obtained with high accuracy (beyond what is usually needed) and very short computing times (the order of a second)

Keywords

roughness, autocorrelation, skewness, kurtosis, non-Gaussian rough surface

1. Introduction

Every engineering surface exhibits irregularities when they are examined at an enough small scale: peaks and valleys appear, revealing its roughness. Engineering surfaces are prepared with progressive finishing processes. Each machining process (or wear) leaves its own signature, with more or less alternating peaks and valleys. Hence, the superposition of all of the signatures leads to the final roughness, characterized by the surface height distribution and the asperities spacing.

As recalled by Bhushan [1], the properties of solid surfaces are crucial to surface interaction because surface properties affect real area of contact, friction, wear, and lubrication. Depending on the desired mechanical application and operating conditions, roughness can have significant effects on the mechanism behavior: shortened durability, higher friction coefficients, higher load capacity, …

In the particular case of lubricated contacts, if the surfaces are kept well separated – high Sommerfeld number – a safe full-film regime is established, and the roughness plays no significant role in the contact behavior. However engineers are brought to extend these operating conditions towards lower Sommerfeld numbers, hence approaching critical lubrication conditions, and creating contact between opposing asperities. Many reasons may motivate this choice. First, the better understanding and control of mixed lubrication – compared to decades ago – and second, for enhanced functioning purposes. As an example, because of lower viscosity and/or higher temperatures, automotive conrod big end bearings partly operate under mixed lubrication conditions, but the mechanism integrity is still guaranteed. In the field of sealing, mechanical seals can not prevent fluid leak, if the two components are not kept very close to each other.

To gain knowledge on surface roughness influence, experiments are obviously needed, but modeling is unavoidable: it is a fast and cheap means to carry out parametric studies. Thus, since almost four decades, rough surface generation is being studied, for which a state of the art can be found in Minet *et al.* [2].

A difficult problem arises: there are as many different roughness types as surfaces. However its has been proved that the surfaces exhibiting the same well-chosen “global characteristics” are equivalent from a tribological point of view [3]. Furthermore, it is well known that both spatial and spectral characteristics are needed as “global characteristics”. The spatial characteristics are well described by the four first statistical moments: mean μ, variance *Va* (square of the standard deviation σ), skewness *Sk* and kurtosis *Ku*. As for the spacing characteristics, the AutoCorrelation Function (*acf*) is widely used.

As described by Whitehouse [4], most of the conventional machining processes produce non-Gaussian surfaces Fig.1, ref. [4], and, thanks to its versatility, the beta-distribution can be used to model the height distribution, Fig.2.

However, modeling the surface height distribution is not self-sufficient: additional information is needed regarding the height spacing. Having both controlled, has lead to an abundant literature since the pioneering work of Patir [5] and since then, the researchers have perfected the rough surface generation for contact, lubrication, …, purposes.

Non-Gaussian rough surfaces were numerically generated for elastic/plastic contact analysis by Kim *et al.* [6] and Chilamakuri and Bhushan [7], with the Hu and Tonder procedure [8]. Both papers [6,7] recall that most of the common machining processes produce surfaces with non-Gaussian distribution, but only low skewness were used ($|Sk| \leq 1$) with short correlation lengths β ($\beta \leq$ surface_width/20).

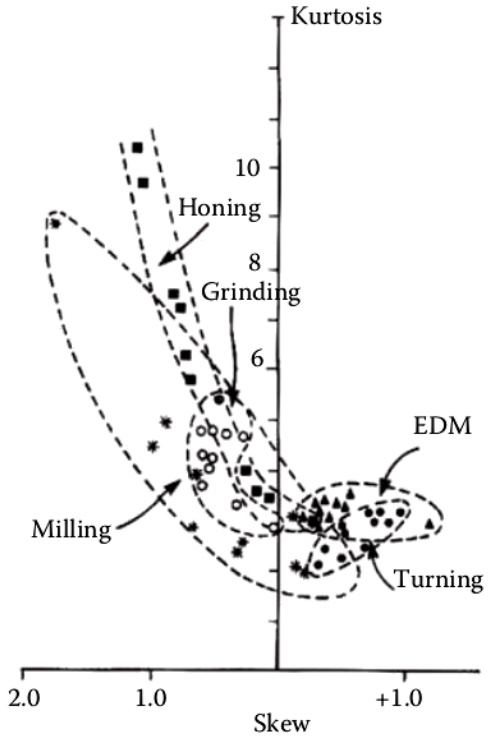


Figure 1: (*Sk, Ku*) for the most common machining processes, [4]

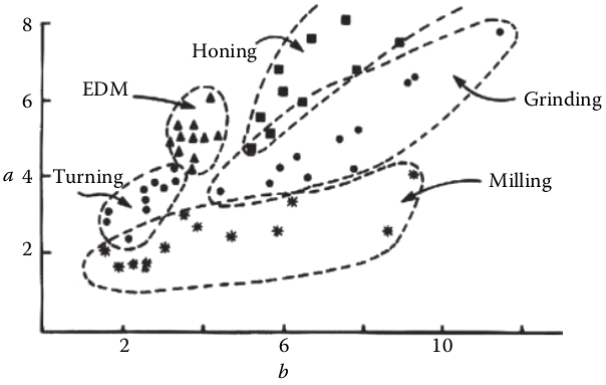


Figure 2: (*a, b*) beta function parameters for the most common processes, [4]

Wu [9] has improved the Hu and Tonder procedure, but his new method cannot generate surfaces with every skewness and kurtosis, especially when the skewness and kurtosis are large. The author argues that such surfaces do not exist, or maybe the limitation of the FFT use is reached. Using Wu's procedure, Reizer [10] has observed high divergence between the modeled and measured surface parameters, when the correlation length was larger than 0.15 of the evaluation length.

Luo *et al.* [11] have developed a roughness generation procedure, using Johnson or Pearson translators, regarding the desired final parameters. The chosen *Sk* range [-2,+2] and *Ku* range [0, 10] were justified regarding most common engineered surfaces, and more specifically, regarding Kim *et al.* work [6] on rough contact analysis. However as measured by Minet *et al.* [2] on mechanical seal faces, values of (*Sk*, *Ku*) up to (-6, 100) should be taken into account. Furthermore, Sedlacek *et al.* [12] have proved that the pair of parameters |*Sk*| and *Ku* not only influences the wear in dry conditions, but tends to lower friction when increasing, in mixed lubrication conditions.

Up to now, moderate ranges of (*Sk*, *Ku*) were used but, as confirmed by Minet *et al.* [2], larger ranges are sometimes needed. However, a more general numerical surface generation faces difficulties that make conventional procedures fail:

- large *Sk* and *Ku*, (some honed surfaces),
- *Ku* below 3, (some turned or milled surfaces),
- and, large autocorrelation lengths.

In the present work, an original and efficient method is proposed to deal with the aforementioned situations. In addition, it can be used to reproduce existing surfaces: instead of working with theoretical skewness and kurtosis values, one can get the heights from the real surface.

2. Surface global characteristics

The surface topography can be seen as a random process for which peaks randomly alternate with valleys. As for any random sequence, the peak and valley relative number and height is quantified by four statistics – the vertical roughness properties. The spatial distribution of the roughness can be described as the superposition of waves of different lengths, leading to smooth surfaces (predominant long wavelengths) or sharp surfaces (predominant short wavelengths).

2.1. Statistical properties

Let $\eta=(\eta_i)_{i=1,\dots,n}$ be a set of *n* **i**ndependent **i**dentically **d**istributed (*iid*) random variables. Under the assumption that it is fully characterized by its four statistical moments ($\mu_{n,\ n=1,2,3,4}$), Eq.(1), $(\eta_i)_{i=1,\dots,n}$ is said statistically equivalent to $(z_i)_{i=1,\dots,n}$ if both share the same moments.

$$\mu_1=\mu=\frac{1}{n}\sum_{i=1}^{n}\eta_i \ ;\ \mu_2=Va=\frac{1}{n}\sum_{i=1}^{n}(\eta_i-\mu)^2 \quad \sigma=\sqrt{Va}$$
$$\mu_3=Sk=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\eta_i-\mu}{\sigma}\right)^3 \ ;\ \mu_4=Ku=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\eta_i-\mu}{\sigma}\right)^4 \qquad (1)$$

(μ, *Va*, *Sk*, *Ku*) are known as, *resp*., mean, variance, skewness and kurtosis but *Va* is seldom used in the field of tribology, the standard deviation σ is preferred.

Fig.3 presents four surfaces measured with a white light interferometer device.

- Fig.3(1) 'Ech03' is a turned surface, which kurtosis is below 3: it is an uncommon case which is difficult to reproduce.
- Fig.3(2) 'Ech14' has been roughly milled; its statistics are quite common but the pattern is trouble for numerical generation.
- Fig.3(3) 'Fluorin' has been highly polished; it is an easy surface to generate because it is nearly Gaussian
- Fig.3(4) 'Rotor' is a worn rotor surface which statistics make the numerical generation difficult.

The specimen are less than 1mm x 1mm, along *x* and *y* axes (horizontal and vertical *resp*.) In order to compare the surface statistical properties, each surface height distribution is normalized with regard to its standard deviation. Fig.4 presents the resulting curves, and some remarks arise:

- A surface distribution can appear to be bimodal 'Ech03', because of the superposition of two distributions (often due to two machining processes).
- The skewness is non-zero when the roughness

distribution is asymmetric. When positive, the skewed distribution illustrates cases where the peaks are statistically greater than the valleys. Conversely, a negative skewed surface, 'Rotor', has deeper valleys, as if the heights of a peaked surface have been reversed.

- The 'Fluorin' height distribution is nearly Gaussian,
- 'Ech14' has a smooth slightly skewed distribution despite a highly patterned appearance.

The kurtosis – to be distinguished from *excess of kurtosis* – is a positive statistics and measures the "peakedness" of the height distribution; for a Gaussian distribution Ku =3. For a symmetric peaked distribution – Sk =0, Ku >3 – there are more middle heights, but some positive and negative heights are greater.

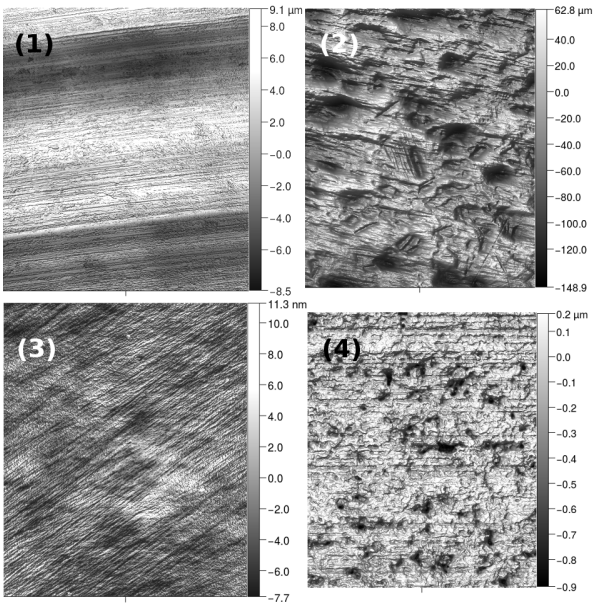


Figure 3: Real worn surfaces – (0,0) coordinate is top left
(1) Ech03, 0.9mmx0.9mm, (Sk, Ku) = (-0.13, 1.87)
(2) Ech14, 0.9mmx0.9mm, (Sk, Ku) = (-0.86, 3.94)
(3) Fluorin, 0.37mmx0.37mm, (Sk, Ku) = (0.09, 2.77)
(4) Rotor, 0.37mmx0.37mm, (Sk, Ku) = (-4.45, 29.01)

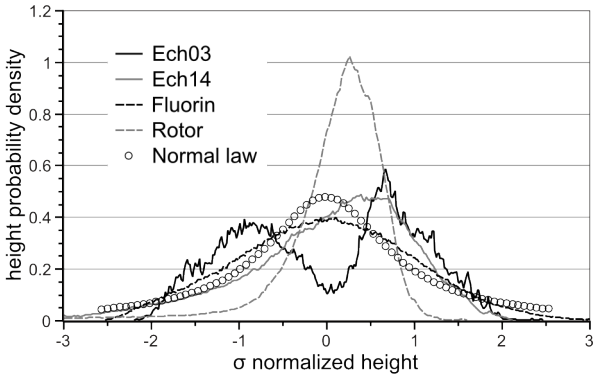


Figure 4: Height probability density of the four surfaces and Gaussian curve (circles)

It is to be reminded for further developments that there exists a relationship between Sk and Ku, $Ku \geqslant Sk^2+1$. which is sometimes referred as Pearson's inequality [13]. As suggested in ref.14, it can be proved as follows:

- $E[(\eta^2-a\eta-1)^2] \geqslant 0$, because the distribution $(\eta^2-a\eta-1)^2$ is positive; E stands for the mathematical expectancy,
- if E(η)=0 and σ(η)=1 then, the previous relationship writes $a^2-2Sk\,a+(Ku-1)\geqslant 0$, where Sk=E(η^3) and Ku=E(η^4)
- and, the condition is verified for $Sk^2-(Ku-1)\leqslant 0$ (negative discriminant to ensure that the second order polynomial remains positive with regard to a)

The strict equality holds for binary distributions.

2.2. Spatial properties

2.2.a. Generality

The statistical parameters are related to how the heights are vertically distributed. The spatial properties represent how the heights η(x,y) are organized in the 2D-space. A practical means to quantify the spatial properties is to compare the surface to a copy shifted of

the quantity $\Delta x,\Delta y$; the more similar, the smoother the surface. The comparison is usually performed thanks to the normalized autocorrelation $\eta\otimes\eta$, Eq.(2).

$$\eta\otimes\eta = acf(\Delta x,\Delta y) = \frac{1}{\sigma^2}\iint \eta(x,y)\eta(x+\Delta x, y+\Delta y)\,dxdy \qquad (2)$$

As information is given on a grid $n=p^2$, involving $(\eta_{ij})_{i,j=1,\dots,p}$, Eq.(2) is rewritten in discrete way, Eq.(3)

$$\eta\otimes\eta = acf(k,l) = \frac{1}{\sigma^2}\sum_{i=1}^{p}\sum_{j=1}^{p}\eta(i,j)\eta(i+k,j+l) \qquad (3)$$

Fast Fourier Transforms, FFT, will be used to compute the *acf*s, so the input signal is supposed to be periodic. Hence the above cross-correlations are in fact cyclic cross-correlations: i+k=i+k [p] and j+l=j+l [p]. As will be shown in the results, this assumption proves to be valid for the great majority of the treated cases.

The *acf* reveals the relative importance of randomness regarding periodicity of a profile. It is widely admitted that the *acf* of a pure random profile is well fitted by a declining exponential. By extension it is considered that most of engineering surfaces have an exponential *acf*, as expressed in Eq.(4), where β_X and β_Y are decay factors relative to α direction. Then, on the α-rotated ellipse, which major and minor axes are β_X, β_Y *resp.*, $acf(x,y)=acf(0,0)/10$

$$acf = \exp\left[-2.3\sqrt{\left(\frac{\cos\alpha.x+\sin\alpha.y}{\beta_X}\right)^2+\left(\frac{-\sin\alpha.x+\cos\alpha.y}{\beta_Y}\right)^2}\right] \qquad (4)$$

However, if a periodic component is introduced by the machining process, the resulting *acf* will be a combination of two functions, as illustrated by Petropoulos *et al.* [15], for shaping, turning, grinding and supergrinding: the *acf* shape will quickly differ from an exponential when moving away the origin. Thus the *acf* initial slope and its decrease rate will be kept as surface smoothness indicators: the greater, the sharper.

2.2.b. *acf* curves of the reference surfaces

Along the turning tool displacement axis, vibrations make the surface less smooth than in the orthogonal direction. As it can be seen in Fig.3(1) long wavelengths are predominant along *x* direction, which is corroborated by the *acf*, Fig.5: the surface heights are normalized with respect to σ, then, if wavelengths are more pronounced in a direction, the *acf* will decrease slowly in the orthogonal direction.

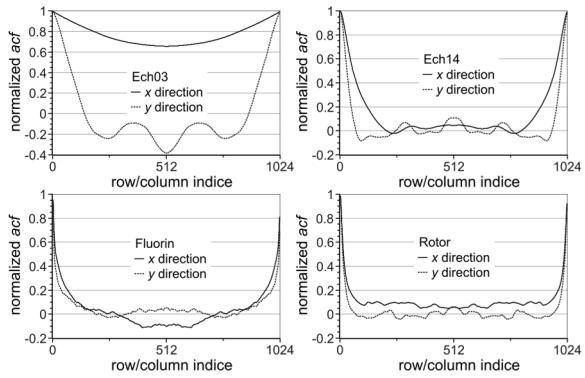


Figure 5: Normalized *acf* of the worn surfaces

Ech03, β_X=1.9, β_Y=0.3, α=-4.0°

Ech14, β_X=0.18, β_Y=0.05, α=7.0°

Fluorin, β_X=0.10, β_Y=0.06, α=-16.0°

Rotor, β_X=0.06, β_Y=0.03, α=-23°

2.2.c. Computational aspects

The computational cost is proportional to n^2, which is prohibitive with grids above n=512x512. However it is reminded that a convolution is faster computed thanks to a FFT, which only needs $n\log(n)$ operations.

$$z \otimes z = \mathrm{FFT}^{-1}[Z^2] = \mathrm{FFT}^{-1}[\mathrm{FFT}(z) * \mathrm{FFT}(z)] \quad (5)$$

The FFT efficiency is due to the assumption that the surface is periodic in both direction. As it will be seen in the part dedicated to the results, this computational trick can induce side effects – artifacts – in the generation of numerical surfaces.

3. Conventional method for the surface numerical generation

The well-known Hu and Tonder [8] procedure is reminded hereafter, with a slight improvement proposed by Bakolas [3].

Step 1

[Determination of the desired spatial characteristics (*acf*), and the four statistical moments μ_i (mean μ, variance σ^2, skewness *Sk* and kurtosis *Ku*)]

Unlike Sk_z and Ku_z which are chosen by the user in order to be representative of the desired worn surface, the prescribed mean μ is set to 0 and the standard deviation σ to 1. However, at the end of the surface generation process, these two moments can be changed with no effect on the third and fourth moments, by scaling and shifting the final surface heights.

Step 2

[Determination of the digital filter *H*.]

The use of a digital filter writes:

$$z = h \otimes \eta \quad (6)$$

where z is the final surface height, η a random white noise and h the digital filter.

In the frequency space it becomes $Z=H.A$, then, $Z\bar{Z} = H\bar{H}A\bar{A}$ which is simply written as $Z^2 = H^2 A^2$. As $Z^2 = \mathrm{FFT}(acf_z)$, and because h is symmetric regarding x and y, $H = \sqrt{(|\mathrm{FFT}(acf_z)|)}/|A|$.

If η is white noise, $|A|$ becomes a constant and can be ignored, Eq.(7)

$$h = \mathrm{FFT}^{-1}\left[\sqrt{(|\mathrm{FFT}(acf_z)|)}\right] \quad (7)$$

Step 3

[Determination of the starting values Sk_η and Ku_η, derived from the desired values Sk_z and Ku_z]

A starting random set η is generated. η is considered as nearly white noise, centered and scaled: $\mu_\eta=0$, $\sigma_\eta=1$. If a digital filter h is applied to η, the z resulting statistical moments are modified, Eq.(8), except for the mean.

$$\begin{aligned}
&\mu_z = 0 \\
&\sigma_z^2 \approx \sum_{k=1}^{n} h_k^2 \\
&Sk_z \approx \frac{\sum_{k=1}^{n} h_k^3}{\left[\sum_{k=1}^{n} h_k^2\right]^{\frac{3}{2}}} Sk_\eta \\
&Ku_z - 3 \approx \frac{\sum_{k=1}^{n} h_k^4}{\left[\sum_{k=1}^{n} h_k^2\right]^{2}} (Ku_\eta - 3)
\end{aligned} \quad (8)$$

In this work, the authors propose the use of an alternative set of equations Eq.(9), rather than the classical one Eq.(8). It is simpler and uses the zero-mean property of η. Indeed h can be written $h = \bar{h} + \mu_h$ where $\bar{h}$ is zero-mean, then $z = (\bar{h} + \mu_h) \otimes \eta = \bar{h} \otimes \eta$.

$$\begin{aligned} \mu_z &= 0 \\ \sigma_z^2 &\approx n\,\sigma_h^2 \\ Sk_z &\approx \frac{1}{\sqrt{n}} Sk_h Sk_\eta \\ Ku_z - 3 &\approx \frac{1}{n}(Ku_h - 3)(Ku_\eta - 3) \end{aligned} \quad (9)$$

The detailed transformations are provided in Appendix 2.

Step 4

[Johnson's transformation of η to match Sk_η and Ku_η]

The Johnson's translation system is usually utilized to transform non-Gaussian data sets into Gaussian data sets. In the present case, the system is reversed in order to obtain data sets η with prescribed moments, starting from Gaussian noise φ. The translator system is composed of three functions reminded in Eq.(10)

$$\begin{aligned} S_U : \eta_{i,i=1,\dots,n} &= \xi + \lambda \sinh\left\{\frac{\varphi_i - \gamma}{\delta}\right\} \\ S_L : \eta_{i,i=1,\dots,n} &= \xi + \lambda \exp\left\{\frac{\varphi_i - \gamma}{\delta}\right\} \\ S_B : \eta_{i,i=1,\dots,n} &= \xi + \lambda \Big/ \left\{1 + \exp\left(-\frac{\varphi_i - \gamma}{\delta}\right)\right\} \end{aligned} \quad (10)$$

The kind of transformation used (Unbounded, Log-normal or Bounded) and the ξ, λ, γ, δ parameters depend on the prescribed moments. Hill *et al.* [16] have provided an algorithm which automatically choose the right transformation with the associated parameters.

Step 5

[z is obtained by digital filtering of η]

The digital filter H is used to average η, and finally $z=\mathrm{FFT}^{-1}(Z=HA)$.

At the end of the process, z is supposed to exhibit the right autocorrelation function *acf* and the right four statistics, μ, σ, *Sk* and *Ku*.

4. Analysis of the conventional method and first improvements

4.1. About Hill's determination of the transformation parameters – limitation 1.

The aforementioned process is limited by Hill's algorithm. Fig.6 shows the kind of transformation used regarding (*Sk, Ku*) value: "U" stands for "Unbounded", "B" stands for "Bounded", "failure" indicates that Hill's algorithm has not converged, and "unreachable" is used below Pearson's limit $Ku<Sk^2+1$. Fig.6(2) highlights the importance of the area for which no solution is provided. As a consequence numerous (Sk_η, Ku_η) required by Step3 can not be reached, and the process stops.

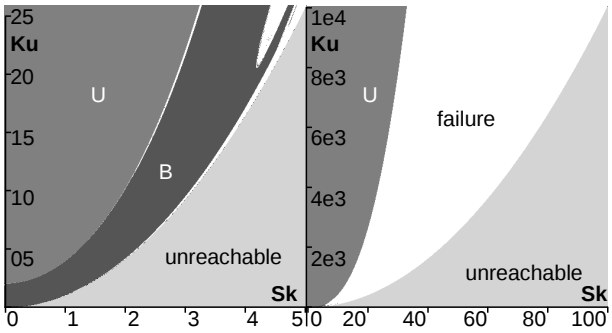


Figure 6: transformation type for *Sk* = 0...5 (left) and *Sk*=0...100 (right)

4.2. About Johnson's translation system – limitation 2.

Hill's algorithm only provides transformation parameters. Several Gaussian series have to be generated until the transformed data set matches the required statistical parameters. Then, if a high precision is required on the results, a lot of series will be necessary. It is explained by the fact that the series are supposed to be uncorrelated infinite series. Besides, as

the process lays on numerically generated random series, there is no "formula" to reproduce a data set; if a particular random serie, among all generated series, gives satisfactory results, there is no savable parameter in order to reproduce it.

The following table, Tab.1, shows the mean and standard deviation of 20 Johnson's generated series, for $(Sk,Ku) = (-2, 20)$ and $n=512^2$: despite the quality of the Gaussian starting random series, the generation of η leads to very variable kurtosis results.

	φ		η	
	Sk	***Ku***	***Sk***	***Ku***
μ	7.10^{-4}	$3.+7.10^{-4}$	$-2.+4.10^{-3}$	19.8
σ	6.10^{-3}	1.10^{-2}	$1.4\ 10^{-1}$	5.6

Table 1. Variability in the process of generating η with Johnson's translator system

4.3. About the relationships Eqs.(6,8) or Eqs. (6,9) – limitation 3.

On the quality of the resulting acf(z)

When η is generated with prescribed statistical moments, it is no more white noise. Therefore $|A|$ is not constant and z does not exhibit the prescribed autocorrelation function *acf(z)*. Furthermore when the decay factors (β_X, β_Y) increase (long wavelengths), the resulting *acf(z)* worsens. An immediate correction consists in applying the formula in Eq.(11).

$$z = \mathrm{FFT}^{-1}\left(H\frac{A}{|A|}\right) \qquad (11)$$

Doing so, *acf(z)* is exactly the same as the one prescribed but η is altered: it is white noise (FFT(η)=$A/|A|$) but its statistical moments are no longer the same. As the statistical moments need anyway to be corrected, as explained below, Eq.(11) is kept so that *acf(z)* will perfectly match the prescribed *acf*.

On the quality of the resulting statistical moments (Sk_{zc}, Ku_{zc}), theoretical background

The skewness and kurtosis belong to higher-order statistics. These scalar values are particular cases of higher-order statistical functions that are sensitive to phase relationships between multiple frequencies. To explain the origin of Eq.(8) a short piece of Higher-Order Statistics (HOS) follows, without the proofs, [17,18].

The p*th*-order moment function of a real stationary random process η is defined as

$$m_p^{\eta}(\tau_1,\tau_2,\ldots,\tau_{p-1}) \stackrel{\text{def}}{=} \mathrm{E}\left[\eta_k,\eta_{k+\tau_1},\ldots,\eta_{k+\tau_{p-1}}\right] = \frac{1}{n}\sum_{k=1}^{n}\eta_k\,\eta_{k+\tau_1}\ldots\eta_{k+\tau_{p-1}}$$

which only depends on the lags $(\tau_1,\tau_2,\ldots,\tau_{p-1})=\boldsymbol{\tau}_{p-1}$. Therefore the 2*nd*-order moment function $m_2^{\eta}(\tau_1)$ is the η autocorrelation function and then the zero*th*-lag moment $m_2^{\eta}(0)$, the mean square value.

The p*th*-order cumulant is defined as the p*th*-moment, with correction terms from lower moments, so chosen as to make the result additive under convolution of the **p**robability **d**ensity **f**unctions, *pdf* ([19] Appendix C, « Convolutions and Cumulants »). The following relationships between moment and cumulant sequences of zero-mean are – for orders p=1, 2, 3, 4:

- $c_1^{\eta}=m_1^{\eta}=\mathrm{E}[\eta_k]=0$, mean value
- $c_2^{\eta}(\tau_1)=m_2^{\eta}(\tau_1)-(m_1^{\eta})^2=m_2^{\eta}(\tau_1)$, autocorrelation sequence

- $c_3^{\eta}(\tau_1,\tau_2)=m_3^{\eta}(\tau_1,\tau_2)-m_1^{\eta}(\ldots)=m_3^{\eta}(\tau_1,\tau_2)$, skewness function
- $c_4^{\eta}(\tau_1,\tau_2,\tau_3)=m_4^{\eta}(\tau_1,\tau_2,\tau_3)-m_2^{\eta}(\tau_1)m_2^{\eta}(\tau_3-\tau_2)$ $-m_2^{\eta}(\tau_2)m_2^{\eta}(\tau_3-\tau_1)-m_2^{\eta}(\tau_3)m_2^{\eta}(\tau_2-\tau_1)$, kurtosis function

By putting $\tau_1=\tau_2=\tau_3=0$ in the previous relationships, the following constants appear:

- $\gamma_2^{\eta}=\mathrm{E}[\eta_k^2]=c_2^{\eta}(0)$ (variance),
- $\gamma_3^{\eta}=\mathrm{E}[\eta_k^3]=c_3^{\eta}(0,0)$ (skewness $\gamma_3^{\eta}/[\gamma_2^{\eta}]^{3/2}$),
- $\gamma_4^{\eta}=\mathrm{E}[\eta_k^4]-3[\gamma_2^{\eta}]^2=c_4^{\eta}(0,0,0)$ (normalized excess of kurtosis $\gamma_4^{\eta}/[\gamma_2^{\eta}]^2$)

A first result to be noticed is that, **1/** if η is assumed to be *iid*, e.g. white noise, Eq.(12) is verified:

$$\begin{cases} c_p^{\eta}(\boldsymbol{\tau}_{p-1})=\gamma_p^{\eta}, \text{ if } \boldsymbol{\tau}_{p-1}=0 \\ c_p^{\eta}(\boldsymbol{\tau}_{p-1})=0 \quad , \text{otherwise} \end{cases} \tag{12}$$

A second point is that **2/** the p*th*-order cumulant function of a non-Gaussian stationary random signal η can be written as (for *p*=3, 4 only)

$c_p^{\eta}(\boldsymbol{\tau}_{p-1})=m_p^{\eta}(\boldsymbol{\tau}_{p-1})-m_p^{G}(\boldsymbol{\tau}_{p-1})$ where $m_p^{G}(\boldsymbol{\tau}_{p-1})$ is the p*th*-order moment function of an equivalent Gaussian signal that has the same mean value and autocorrelation sequence as η. Hence for a Gaussian process $c_3^{\eta}(\tau_1,\tau_2)=c_4^{\eta}(\tau_1,\tau_2,\tau_3)=0$. Anyway, if η is Gaussian $c_p^{\eta}(\boldsymbol{\tau}_{p-1})=0$ for all *p*, be it colored or white process. So it can be said that cumulants not only display the amount of higher-order correlation, but also provide a measure of the distance of the random process from gaussianity.

When applied to a linear filter, $z_k=\sum_{i=1}^{n} h_i\eta_{i+k}$ the p*th*-order cumulant writes:

$$c_p^{z}(\boldsymbol{\tau}_{p-1})=\sum_{i_0}..\sum_{i_{p-1}} h_{l-i_0}h_{l-i_1+\tau_1}\ldots h_{l-i_{p-1}+\tau_{p-1}} \times c_p^{\eta}(i_1-i_0,\ldots i_{p-1}-i_0)$$

and its zero*th*-lag value, Eq.(13):

$$c_p^{z}(0)=\sum_{i_0}\sum_{i_1}\ldots\sum_{i_{p-1}} h_{l-i_0}h_{l-i_1}\ldots h_{l-i_{p-1}} \times c_p^{\eta}(i_1-i_0,i_2-i_0,\ldots i_{p-1}-i_0) \tag{13}$$

The third point is that **3/** if η is assumed to be *iid,* Eq. (13) becomes:

$$c_p^{z}(0)=\gamma_p^{\eta}\sum_{k=1}^{n} h_k^p \tag{14}$$

Eq.(14) finally yields the Eq.(8) relationships for *p*=2, 3, 4.

To sum up the important points:

1/-3/ white noise cumulants are identically zero, except zero lag cumulants, then there exist simple relationships between the noise statistics and the filtered signal, Eq. (14) and more particularly Eq.(8).

2/ Gaussian noise cumulants are identically zero: filtered Gaussian noise remains Gaussian.

To conclude on this piece of theory, if η can not be considered as white noise, Eq.(8) relationships are no longer satisfied because of cross-correlations that not vanish in Eq.(13). For instance, for extremely skewed distributions regarding the number of points *n*, the cross-products of η autocorrelation do not reduce to zero. As illustrated on Fig.7, the problem can be

overcome by increasing the number of points; the cross-product mean obviously decreases but the standard deviation also.

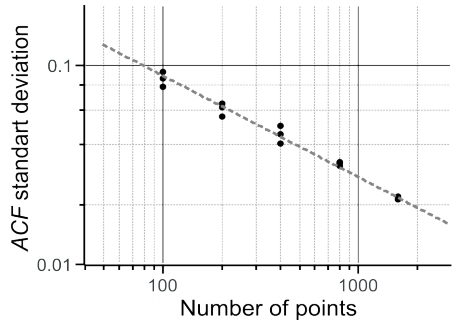


Figure 7: standard deviation of a data set normalized autocorrelation (Sk=-2, Ku=8)

If it is not possible to increase the number of points – what usually happens, because n is a prescribed parameter – Eq.(8) induces errors on (Sk_η , Ku_η) as proved by the following case:

- prescribed parameters : n=512x512, (Sk_z, Ku_z)=(-3, 30) ; $\beta_X/512=\beta_Y/512=\tau$=0.3 then ($Sk_\eta$, Ku_η)=(-99.4, 9940.3)
- generated η: ($Sk_{\eta c}$, $Ku_{\eta c}$)=(-98.3, 9940.4)
- After linear filtering: (Sk_{zc}, Ku_{zc})=(-3.5, 39.1), and *acf(z)* is presented on Fig.8.

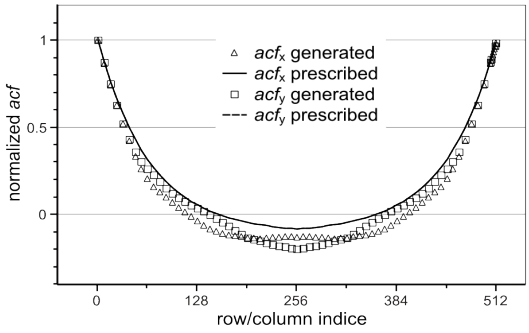


Figure 8: comparison of the generated *acf* and the prescribed one

To conclude, as the final moments need anyway to be corrected, the use of Eqs.(9,11) is fully justified because it at least ensures the right *acf(z)*.

On the quality of the resulting statistical moments (Sk_{zc}, Ku_{zc}), parametric study

In order to assess the error introduced with Eqs.(9,11), a small parametric study is carried out:

- $\beta_X/512=\beta_Y/512=\tau=\{0.01, 0.02, \dots , 0.64\}$
- (Sk_z, Ku_z)={(0,3) , (0,6) , (-2, 12)} *ie* a Gaussian signal, a non -Gaussian symmetric *pdf* signal and a skewed signal *resp*.
- n=512x512

For each case 5 calculations are performed because of the stochastic nature of the process. *errKu* is the relative difference between the prescribed kurtosis Ku_z and the calculated kurtosis Ku_{zc} values, *errSk* is the absolute difference between the prescribed skewness Sk_z and the calculated skewness Sk_{zc} values (because of the zero value, relative difference is not possible). For the (Sk_h, Ku_h) cases that violate Pearson's inequality, a star (*) replaces the result in Tab. 2.

	(0,3)		(0,6)		(-2,12)	
τ	**errSk**	**errKu (%)**	**errSk**	**errKu (%)**	**errSk**	**errKu (%)**
0.01	0.01	0.23	0.02	18.93	0.63	29.00
0.02	0.00	0.72	0.00	18.73	0.59	28.63
0.04	0.02	1.92	0.01	19.50	0.57	27.78
0.08	0.01	1.50	0.04	19.75	0.64	28.80
0.16	0.08	3.18	0.03	21.80	*	*
0.32	0.09	2.06	0.06	21.17	*	*
0.64	0.01	10.04	0.17	25.54	*	*

Table 2. Errors induced by the use of (*Sk, Ku*) moment relationships

It can be concluded that the errors become large for high values of τ (very long wavelengths) and increasing

values of (Sk_z, Ku_z). So the question is, is there any relationship between (Sk_z, Ku_z) and (Sk_{zc}, Ku_{zc}) ?

- Because τ has evidently an influence on the results, it is set to a moderate value (τ=0.08).
- (Sk_h , Ku_h) value is taken in the range $[0,-50]x[10,8000]$, then (Sk_z, Ku_z) is calculated according Eq.(9) and (Sk_{zc}, Ku_{zc}) is the computed value after digital filtering Eq.(11)
- (Sk_h , Ku_h) = (49.2, 5453.3) is a constant because it only depends on τ, according to Eq.(4).

On Fig.9 it can be seen that for a given Sk_z , different Sk_{zc} are obtained: one can conclude that the kurtosis has an effect on the skewness prediction, which is due to the η whitening process. On the contrary, a rather good correlation exists between Ku_z and Ku_{zc} whatever the skewness.

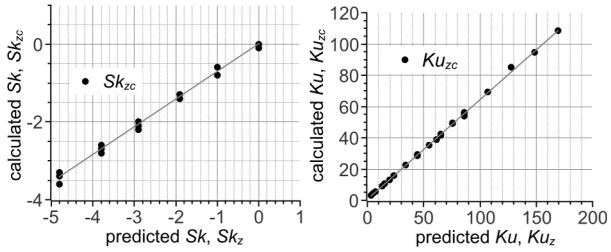


Figure 9: Skewness obtained after filtering (Sk_{zc}) vs predicted skewness (Sk_z) with Eq.(9), (left) and kurtosis obtained after filtering (Ku_{zc}) vs predicted kurtosis (Ku_z) with Eq.(9), (right)

Hence, for a given τ, the relationships can be corrected in order to gain accuracy; in the present case (τ=0.08) that leads to Eq.(15).

$$\begin{aligned} Sk_z &= \frac{1}{1.40\sqrt{n}} Sk_h Sk_\eta \\ Ku_z - 3 &= \frac{1}{1.57n}(Ku_h - 3)(Ku_\eta - 3) \end{aligned} \quad (15)$$

It has to be recalled that Gaussian noise must be transformed into a Gaussian signal, therefore the fitting functions have been corrected accordingly.

Tabs. 3,4 detail the error on skewness and kurtosis after correction.

Sk_η	Skzc	Sk_z	errSk
-50	-3.4	-3.4	-0.01
-40	-2.7	-2.7	-0.04
-30	-2.1	-2.1	0.02
-10	-0.7	-0.6	-0.04
0	0.0	0.0	0.01

Table 3. Errors induced on the skewness

Ku_η	Ku_{zc}	Ku_z	errKu %	Ku_η	Ku_{zc}	Ku_z	errKu %
10	3.1	3.0	1.4	2500	36.1	35.1	3.1
100	4.3	4.2	1.4	2800	40.1	38.8	3.4
200	5.6	5.4	3.7	3000	42.8	41.9	2.0
500	9.6	9.3	3.6	3500	49.4	49.4	0.6
600	10.9	10.6	3.3	4000	56.0	54.9	2.3
800	13.6	13.0	4.3	5000	69.3	69.2	0.2
1000	16.2	15.9	2.3	6000	82.6	85.1	2.9
1500	22.9	22.7	0.8	7000	95.8	94.7	1.2
2000	29.5	28.9	2.1	8000	109.1	108.4	0.6

Table 4. Errors induced on the kurtosis

To summarize the preceding analysis, the corrected (Sk,Ku) formulas Eq.(15) improve the accuracy of the results, but they strongly depend on τ. So the last analysis to be performed quantifies the effect of τ on Eq. (9).

τ values are taken in the range [0.01, 0.64], h is generated according to Eq.(7) and (Sk_h, Ku_h) is calculated; Fig.10 shows a strong correlation between (Sk_h, Ku_h) and τ.

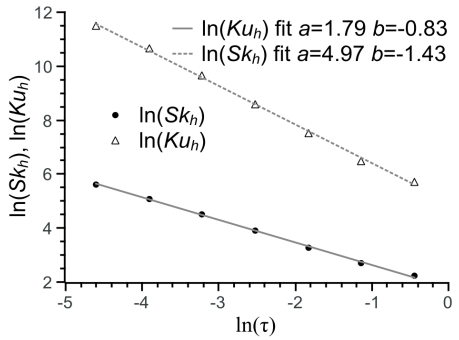


Figure 10: (Sk_h, Ku_h) as a function of τ, in other words the digital filter statistics regarding the decay factor of the *acf*

Then if (Sk_h, Ku_h) are calculated thanks to the fitting function found, Eq.(16), the error on Sk_h remains below 6% and the error on Ku_h below 12%, which is accurate enough to pursue the analysis.

$$Sk_h = \frac{6.02}{\tau^{0.83}} \text{ and } Ku_h = \frac{144}{\tau^{1.43}} \qquad (16)$$

It must be said that the (Sk_h,Ku_h) fit can not be extrapolated to lower values of τ because as proved in Appendix 3, $\lim_{\tau \to 0} Sk_h \approx \sqrt{n-3}$ and $\lim_{\tau \to 0} Ku_h \approx n-2$, which is quite different from Eq.(16). However such cases should not happen: if τ <<0.01 then $h \approx \delta_1$ (h=1 if i=1, h=0 otherwise), hence $z \approx \eta$ which presents no interest.

With prescribed (Sk_z, Ku_z), (Sk, Ku) formulas of Eq.(9), are used to deduce (Sk_η, Ku_η) with respect to τ, thanks to (Sk_h, Ku_h) fit . Pearson's limit is reached whenever Ku-Sk^2-1<0, so the non-dimensional parameter $P_M = (Ku_\eta - Sk_\eta^2 - 1)/Ku_\eta$ is plotted against τ and when P_M becomes negative it means that (Sk_η, Ku_η) can not be generated. Fig.11 represents the evolution of the aforementioned criterion with (Sk_z, Ku_z)=(-2, 12) as an example. As it clearly appears if τ is greater than 0.08, (Sk_η, Ku_η) violates Pearson's limit and can not be generated: the classical Hu and Tonder's procedure stops.

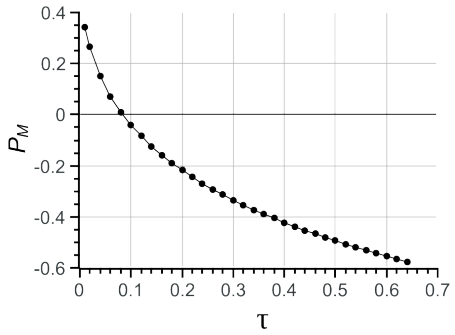


Figure 11: "Pearson's criterion" as a function of the decay factor τ

Fig.12 represents the decay factor τ maximum usable value as a function of Ku_z for the most common values of Sk_z (-1, -2, -3, -4, -5).

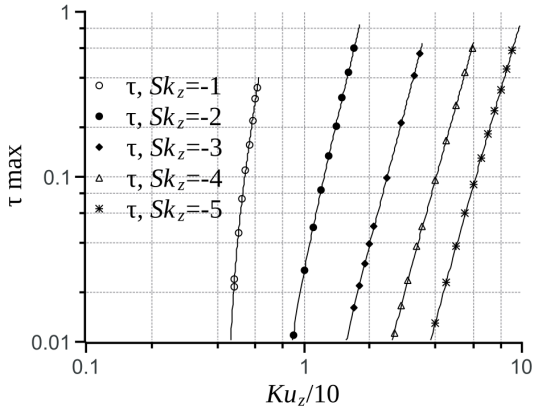


Figure 12: maximum of the decay factor τ as a function of Ku_z for common values of Sk_z

It illustrates the fact that, despite quite common values of (Sk_z, Ku_z) and τ, some cases are nonsolvable because of impossible values of (Sk_η, Ku_η).

A similar case is encountered when the prescribed kurtosis is below 3 associated to high τ values of the *acf*: the kurtosis Ku_η will even turn negative, which is impossible.

Conclusion on the conventional method

Limitation 1: Numerous (*Sk,Ku*) values are not reachable with Hill's algorithm.

Limitation 2: Among the reachable (*Sk,Ku*) values, some need a lot of tries to generate an accurate (*Sk,Ku*) serie.

Limitation 3: Assuming that the final *acf* matches the prescribed one, using the proposed correction Eq.(11), *z* statistics do not match the prescribed ones.

5. New hybrid method

To get rid of the aforementioned limitations, three major improvements should be proposed:

a) Any (Sk_η, Ku_η) value must be reachable, provided the fact that Pearson's inequality remains satisfied – *limitation 1.*

b) A more robust method must be used to generate η; ideally speaking, a deterministic algorithm should be defined – *limitation 2.*

c) The final surface heights must be corrected if necessary, to better match the prescribed moments – *limitation 3.*

5.1. Use of an analytical function instead of a data set transformation, improvements (a) and (b)

Instead of generating random numbers, an analytical function can be used to generate the heights $\eta_i=\eta(x_i)$ along a well-chosen axis. The advantage is that the height distribution will be directly provided with the desired statistical moments in a deterministic way. For reasons that will be explained later, three cases are considered:

$$\begin{cases} 1.10\,Sk_z^2+1>Ku_z\geqslant 1.00\,Sk_z^2+1 & \text{(a)} \\ 1.34\,Sk_z^2+1>Ku_z\geqslant 1.10\,Sk_z^2+1 & \text{(b)} \\ Ku_z\geqslant 1.34\,Sk_z^2+1 & \text{(c)} \end{cases} \quad (17)$$

Most of encountered engineering surfaces belong to the latter situation, Eq.(17.c). Some cases Eq.(17.b) are called binary-like height distributions and are rarely observed. Almost no real surface is highly binary, Eq. (17.a). However, *step 3* of Hu and Tonder's algorithm can require Ku_η in the three ranges.

5.1.a. Usefulness and relevance of an analytical height function

Most of unimodal surfaces are fitted with a tangent function

As presented above, Figs.3,1-4 exhibit very different morphologies. The very first operation performed on the surfaces is the normalization, with respect to their standard deviation. Then, for each surface a vector filled with 2048 representative heights is built and ordered.

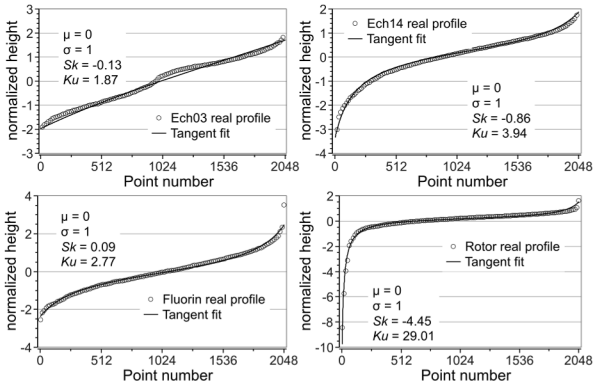


Figure 13: tangent fit on real profiles

Ech03, x_a=0.96, x_b=1.37, x_c=4.08, x_d=0.86

Ech14, x_a=0.20, x_b=0.43, x_c=0.74, x_d=0.25

Fluorin, x_a=0.38, x_b=0.35, x_c=0.92, x_d=-0.07

Rotor, x_a=0.03, x_b=0.20, x_c=0.27, x_d=0.20

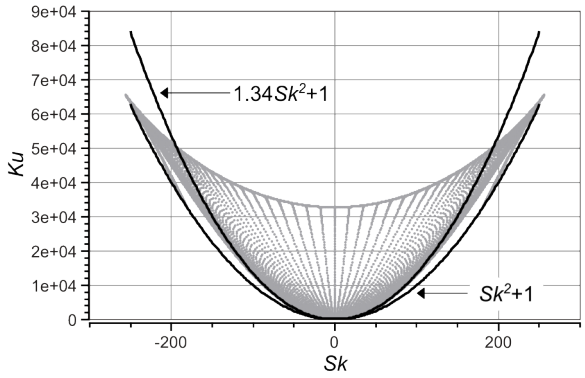
Fig.13 illustrates the surface profiles (circle curve). The curves exhibit a straight central part (except Fig.13 (1))

along with starting/ending smaller portions of more pronounced slopes. Despite the similarity, the well-known Abott-Firestone's curve is not used because of its inherent averaging method of construction. The trick consists then in finding an analytical function able to well fit the aforementioned vector of data. Two candidates have been selected, the inverse hyperbolic tangent function (*atanh*) and the tangent function (*tan*). When regularly discretized, both functions offer the essential ability to reach high values (when *tan* is close to ±π/2 and *atanh* close to ±1) along with lot of moderate values, besides which, only two parameters are needed to control the function shapes.

It should be nonetheless noticed that the *atanh* function can not easily reach high kurtosis values, as proved by the following parametric study. Let us consider the *tan* function regularly discretized over the interval I_1 (Eq.18). A set of 256^2 values is generated, and the third and fourth statistical moments calculated.

$$I_1=\left[-\frac{\pi}{2}(1-a),\frac{\pi}{2}(1-b)\right]\ (a,b)\in[10^{-10},1] \qquad (18)$$

The map *Ku vs Sk* is plotted on Fig.14. It can be *a priori* deduced that a large range of (*Sk, Ku*) is covered.



Let us now consider the *atanh* function and its varying interval I_2. In the same manner, the (*Sk,Ku*) map is plotted (Fig.15) and it appears that the range of (*Sk,Ku*) values is much more limited.

$$I_2=\left[-(1-a),+(1-b)\right]\ (a,b)\in[10^{-14},1] \qquad (19)$$

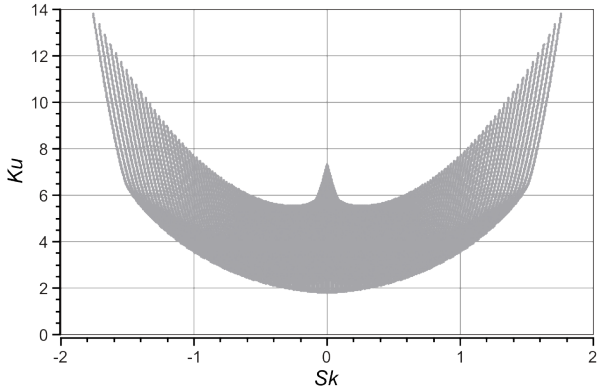


Figure 15: (*Sk,Ku*) values reached by the *atanh* function over the interval I_2

Fig.16 gives a rather simple explanation of that matter of fact. The *atanh* function converges very quickly towards its asymptote unlike the *tan* function, as also proved by the equalities of Eq.(20).

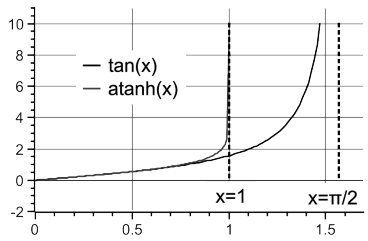


Figure 16: comparison between the *tan* function and the *atanh* function behaviors

$$\int_0^1 \operatorname{atanh}(x)\,dx = \ln(2);\ \int_0^{\pi/2} \tan(x)\,dx = \infty \qquad (20)$$

Lot of other analytical functions are suitable to fit the kind of studied profiles but, to the authors' knowledge, none of them is as close to the profiles as the tangent function, with only two parameters.

Thus, the height-ordered surface distribution should be close to a tangent function with well-chosen limits, as expressed in Eq.(21).

$$\tan(x_i) = \tan\left[-\frac{\pi}{2}(1-a) + \frac{i-1}{n-1}\frac{\pi}{2}[2-(a+b)]\right] \quad i \in \{i,\dots n\}$$
$$\text{and } x_i \in [-\frac{\pi}{2}(1-a), \frac{\pi}{2}(1-b)] \qquad (21)$$

Prior any analytical development, the tangent function fitting capacity – $\tan(x).x_c+x_d$ – has been checked on a large variety of surfaces, including Figs.3,1-4, on a classical spreadsheet. As a result, it is found that, despite high skewness and kurtosis levels to be fitted, the tangent function remains a good candidate for height generation. Fig.13 illustrates the tangent fitting ability with x_a, x_b, x_c and x_d ($x_a = \frac{\pi}{2}a$ and $x_b = \frac{\pi}{2}b$) which are determined with the spreadsheet solver.

Comments

- The lack of fit that appears on Fig.13(1) was foreseeable: Ech03 is a “strong” bimodal surface (Fig.4), its profile exhibits a central curvature change that the tangent function can not fit. That is the reason why the present study is limited to unimodal or weakly bimodal surfaces. However bimodal profiles can be handled by mixing two unimodal profiles.
- Some (*Sk*, *Ku*) values are hardly reached with the *tan* function, as appears in Fig.14 between the black curves.
- As expected, for large kurtosis values and negative skewness, Fig.13(4), x_a becomes small: the -π/2 asymptote of the tangent function appears.

5.1.b. *Transformation of a data set sum into an integral*

The use of an analytical representation of the heights is of limited interest if the sums, as expressed in Eq.(1), can not be avoided because it becomes time consuming for large *n*. So, how will be the statistical moments calculated in a discrete problem? Considering the mean and recalling that Simpson's method involves such a sum and that it links it to the function integral, the relationship Eq.(22) is deduced, with $n=2p$ the number of points, and $h=(\beta-\alpha)/(n-1)$ the interval length.

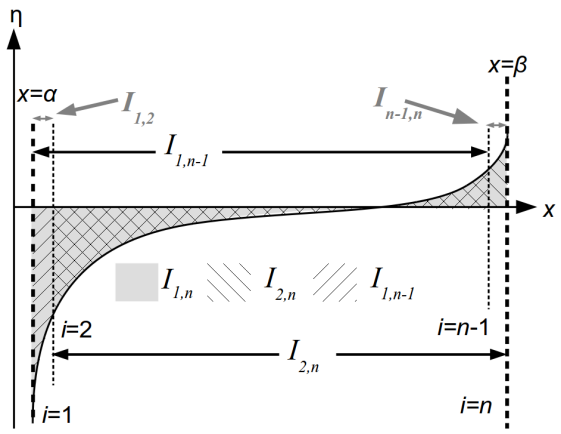


Figure 17: Relationship between the integral $I_{1,n}$, and the integrals starting from odd and even points. $I_{1,n}=0.5*(I_{2,n}+I_{1,2}+I_{1,n-1}+I_{n-1,n})$

The goal being to establish a relationship between $I_{1,n}$ (analytical) and $S_\eta=\frac{1}{n}\sum_{i=1}^{n}\eta_i$, $I_{1,n}$ is decomposed, as shown in Fig.17, where $I_{1,2}$ and $I_{n-1,n}$ are approximated in Eq.(23), yielding Eq.(24).

$$\begin{aligned} I_{1,n} &= \int_\alpha^\beta \eta(x)dx \\ &\approx \frac{h}{3}\left[\eta(x_1)+4\sum_{i=1}^{p}\eta(x_{2i})+2\sum_{i=1}^{p-1}\eta(x_{2i+1})+\eta(x_{2p+1})\right] \\ &\approx \frac{h}{3}\left[\eta_1+4\sum_{i=1}^{p}\eta_{2i}+2\sum_{i=1}^{p-1}\eta_{2i+1}+\eta_{2p+1}\right] \\ &= \frac{1}{2}[I_{1,2}+I_{2,n}+I_{1,n-1}+I_{n-1,n}] \end{aligned} \tag{22}$$

$$\begin{aligned} I_{1,2} &\approx \frac{h}{6}\left(\eta_1+4\eta_{1+0.5}+\eta_2\right) \\ I_{n-1,n} &\approx \frac{h}{6}\left(\eta_{n-1}+4\eta_{n-0.5}+\eta_n\right) \end{aligned} \tag{23}$$

$$\begin{aligned} S_\eta &= \frac{1}{n}\sum_{i=1}^{n}\eta_i \approx \frac{n-1}{n(\beta-\alpha)}I_{1,n} \\ &+\frac{1}{12n}\left[9(\eta_1+\eta_n)+(\eta_2+\eta_{n-1})-4(\eta_{1+0.5}+\eta_{n-0.5})\right] \end{aligned} \tag{24}$$

Based on this principle, the computation of the four moments becomes therefore instantaneous and does not depend on the data set length *n*. The following integrals, Eq.(25), can therefore be analytically determined by hand (but can turn to be tedious), or calculated on a symbolic calculation software.

$$\begin{aligned} \mathrm{Int}_1 &= \int_\alpha^\beta \eta(x)dx \\ \mathrm{Int}_2 &= \int_\alpha^\beta (\eta(x)-\mu)^2 dx \\ \mathrm{Int}_3 &= \int_\alpha^\beta \left(\frac{\eta(x)-\mu}{\sigma}\right)^3 dx \\ \mathrm{Int}_4 &= \int_\alpha^\beta \left(\frac{\eta(x)-\mu}{\sigma}\right)^4 dx \end{aligned} \tag{25}$$

As a result, μ is determined as a function of Int_1, then, σ as a function of Int_2 then, *Sk(a,b)* and *Ku(a,b)* as a function of Int_3 and Int_4 *resp.*, with $\alpha=-\pi/2(1-a)$ and $\beta=+\pi/2(1-b)$.

5.1.c. Data set with prescribed moments

The remaining task is to control the shape of the tangent function, so that the discrete form will match the desired statistical moments.

Modifying a function shape can be achieved by parameterizing its limits. As an example, for the tangent function, the closer to $\pm\pi/2$ the larger the tangent values and the higher the kurtosis. Eq.(26) defines the domain limits, using two parameters *a* and *b*.

$$x\in\left[-\frac{\pi}{2}(1-a),\frac{\pi}{2}(1-b)\right]\ \{a,b\}\in]0,1[\tag{26}$$

An optimization algorithm is finally used to find the right values of *a* and *b*, such that the functional *f*, Eq. (27), is minimized, with imposed *Sk* and *Ku*.

$$f(a,b)=\left[Sk^2-Sk^2(a,b)\right]+|Ku-Ku(a,b)| \tag{27}$$

Once *a* and *b* give satisfactory results regarding *Sk* and *Ku*, the set of points $(\eta_i)_{i=1,n}$ is shifted and scaled to match prescribed values of μ and σ with no subsequent effect on *Sk* and *Ku*.

Therefore to this point, a data set η has been generated

and it matches the prescribed four first statistical moments.

5.1.d. Limitations and solutions

Near Pearson's limit ($Ku \approx Sk^2+1$) the height distribution becomes binary and, the numerical generation is no longer satisfactory, as illustrated in Fig.18.

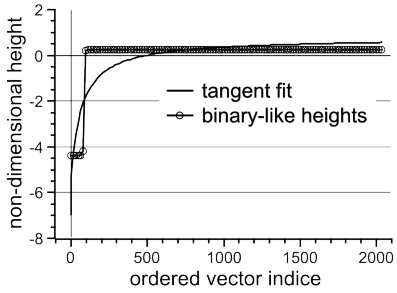


Figure 18: height ordered binary distribution with its tangent fit

Binary-like distributions are often needed during the conventional Step 3, for long correlation lengths. Because the relations in Eq.(8) are very approximate, it can lead to non reachable values of (Sk_η, Ku_η) such that $Ku_\eta \leqslant Sk_\eta^2+1$. Instead of giving up, or arbitrarily modifying the input (Sk_z, Ku_z), it is better to threshold Ku_η using a binary distribution and complete the process with the approximate value of Sk_η and its corresponding Ku_η, $Ku_\eta \approx Sk_\eta^2+1$. The resulting generated surface will be anyway corrected thanks to improvement (c), as shall be seen later.

To solve this problem of weak fitting, another function, Eq.(28), is introduced. It is built so that it exhibits a starting and an ending plateau; therefore, if the domain extends from low values to large values of x, the discrete representation will be binary-like, Fig.19.

$$\begin{aligned} e(x_i) &= sign(x_i)[1-\exp(-|x_i|)] \\ x_i &= -\frac{1}{a}(1-a)+\frac{i-1}{n-1}\left(\frac{1}{a}+\frac{1}{b}-2\right) \\ x_i &\in \left[-\frac{1}{a}(1-a),\frac{1}{b}(1-b)\right] \quad \{a,b\} \in]0,1[\end{aligned} \tag{28}$$

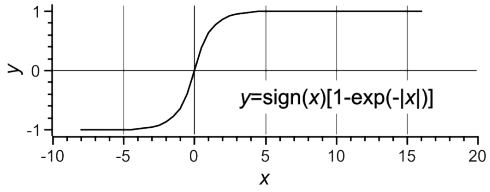


Figure 19: exponential function over the domain [-7,15] (*Sk*,*Ku*)=(-0.7, 1.6)

To assess the exponential function ability to fit binary distributions, the same parametric study as for the tangent function is carried out, and the (*Sk*, *Ku)* map is plotted over the interval I_3, Eq.(29), Fig.20

$$I_3 = [-\frac{1}{a}(1-a),\frac{1}{b}(1-b)] \quad (a,b) \in [10^{-14},1] \tag{29}$$

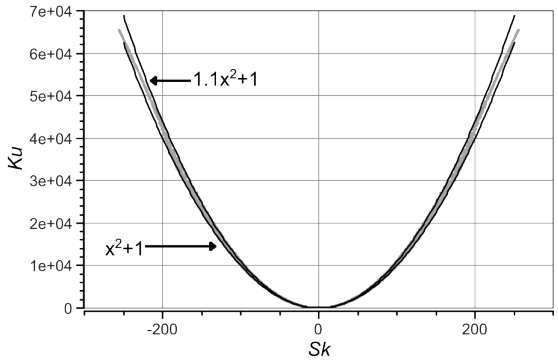


Figure 20: (*Sk*,*Ku*) values reached by the *exp* function over the interval I_3

Large values of *Ku* are reached but they are bounded by two limits: the Pearson's limit and $Ku=1.1Sk^2+1$. In addition, the generation of a tangent distribution revealed to be more accurate when $Ku \geqslant 1.34\,sk^2+1$, Fig.14. Therefore, there exists a transitional region where the *tan* and *exp* functions give moderate quality results: it has been approximatively identified for *Ku* between $1.1Sk^2+1$ and $1.34Sk^2+1$, Fig.21.

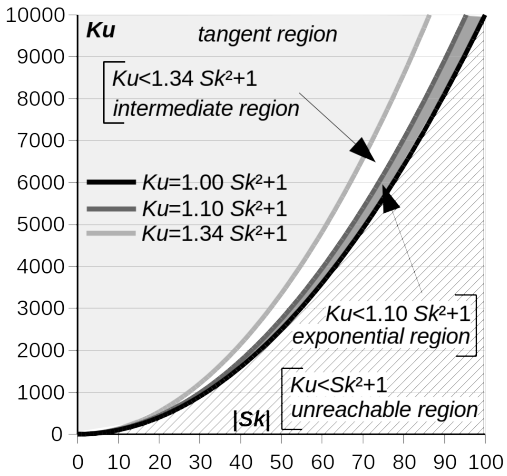


Figure 21: functions used regarding the targeted (*Sk,Ku*)

In this intermediate region, a linear combination of tangent and exponential distributions with the same kurtosis *Ku* is used. Let $(t_i)_{i=1,...,n}$ be a *tan* distribution and $(e_i)_{i=1,...,n}$ an *exp* distribution. A bimodal distribution composed of both distributions with the same probability can be expressed as:

$$(h_i)_{i=1...2n}=\begin{cases}\sigma_t(t_i)+\mu_t & i=1...n\\ \sigma_e(e_i)+\mu_e & i=n+1...2n\end{cases} \tag{30}$$

Then, using the equations in Appendix 4,

$$\begin{aligned}&\mu_h=0\rightarrow\mu_e=-\mu_t\ ,\ \mu_t\in[0,1[\\ &\sigma_h^2=1\rightarrow\sigma_t^2+\sigma_e^2=2(1-\mu_t^2)\ \text{then}\\ &\begin{cases}\sigma_t=\sqrt{2(1-\mu_t^2)}\cos\gamma\\ \sigma_e=\sqrt{2(1-\mu_t^2)}\sin\gamma\ ,\ \gamma\in\left[0,\frac{\pi}{2}\right]\end{cases}\end{aligned} \tag{31}$$

Hence,

$$\begin{aligned}2\,Sk_h&=Sk_t\sigma_t^3+Sk_e\sigma_e^3+3\mu_t(\sigma_t^2-\sigma_e^2)\\ 2\,Ku_h&=Ku(\sigma_t^4+\sigma_e^4)+4\mu_t(Sk_t\sigma_t^3-Sk_e\sigma_e^3)+\\ &\quad 6\mu_t^2(\sigma_t^2+\sigma_e^2)+2\mu_t^4\end{aligned} \tag{32}$$

An optimization routine is used to determinate the two unknowns μ_t and γ, which is quite instantaneous. If one wishes a final set of only *n* terms instead of 2*n* terms, Sk_h and Ku_h must be doubled, and half terms of the bimodal distribution dropped.

To conclude on this part, whatever the profile being fitted, the tangent function can be used for most of cases, and a complementary exponential function is introduced for limit cases.

5.2. Efficiency, accuracy

The new approach proposed by the authors can thus be modeled as follows:

$$(Sk,\ Ku)\rightarrow(Sk_{a,b},\ Ku_{a,b})\rightarrow(Sk_\eta,\ Ku_\eta)$$

It means that once prescribed (*Sk*, *Ku*), (*a*,*b*) are determined leading to ($Sk_{a,b}$, $Ku_{a,b}$). Eq.(33) is obtained thanks to Eq.(24) with $\alpha=-\frac{\pi}{2}(1-a)$ and $\beta=\frac{\pi}{2}(1-b)$ for the tangent function. The integral is determined formally as a function of *a* and *b*.

$$\begin{aligned}\mu_{a,b}&=\frac{1}{n}\sum_{i=1}^{n}\tan(x_i)\approx\frac{n-1}{n\frac{\pi}{2}(2-a-b)}\left[\int_{-\frac{\pi}{2}(1-a)}^{\frac{\pi}{2}(1-b)}\tan(x)\,dx\right]\\ &+\frac{1}{12n}\left[9(\tan(x_1)+\tan(x_n))+(\tan(x_2)+\tan(x_{n-1}))\right.\\ &\left.-4(\tan(x_{1+0.5})+\tan(x_{n-0.5}))\right]\end{aligned} \tag{33}$$

Knowing $\mu_{a,b}$ the standard deviation is also determined as a function of *a* and *b*, Eq.(34)

$$\begin{aligned}\sigma_{a,b}^2&=\frac{1}{n}\sum_{i=1}^{n}(\tan(x_i)-\mu_{a,b})^2\\ &\approx\frac{n-1}{n\frac{\pi}{2}(2-a-b)}\left[\int_{-\frac{\pi}{2}(1-a)}^{\frac{\pi}{2}(1-b)}(\tan(x)-\mu_{a,b})^2\,dx\right]+...\end{aligned} \tag{34}$$

Then $Sk_{a,b}$, $Ku_{a,b}$ follows, Eq.(35).

$$
\begin{aligned}
Sk_{a,b} &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\tan(x_i)-\mu_{a,b}}{\sigma_{a,b}}\right)^3 \\
&\approx \frac{n-1}{n\frac{\pi}{2}(2-a-b)}\left[\int_{-\frac{\pi}{2}(1-a)}^{\frac{\pi}{2}(1-b)}\left(\frac{\tan(x)-\mu_{a,b}}{\sigma_{a,b}}\right)^3 dx\right]+\ldots \\
Ku_{a,b} &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\tan(x_i)-\mu_{a,b}}{\sigma_{a,b}}\right)^4 \\
&\approx \frac{n-1}{n\frac{\pi}{2}(2-a-b)}\left[\int_{-\frac{\pi}{2}(1-a)}^{\frac{\pi}{2}(1-b)}\left(\frac{\tan(x)-\mu_{a,b}}{\sigma_{a,b}}\right)^4 dx\right]+\ldots
\end{aligned}
\qquad (35)
$$

Finally, the minimization problem Eq.(27) is solved and (a, b) determined. Then, a data set is generated with (a,b) resulting in (Sk_η, Ku_η).

So, the first validation to be carried out is the ability of the method to analytically reach any (Sk,Ku) value and being close to its value.

(Sk, Ku) → ($Sk_{a,b}$, $Ku_{a,b}$)

Fig.22 shows the validation domain (Sk,Ku) ∈ [0,100]x[1,10001]; for a given (Sk,Ku), an analytical function is automatically chosen (tangent, exponential, or intermediate), and (a,b) is determined. The domain is covered with 512x512 validation points.

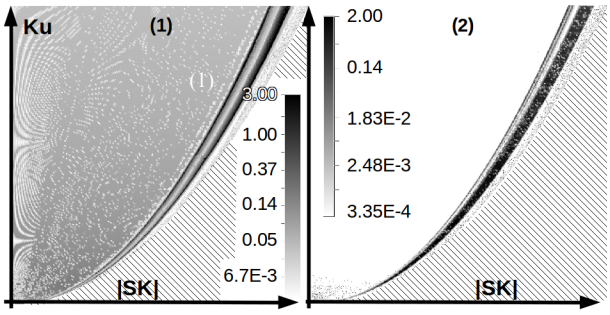


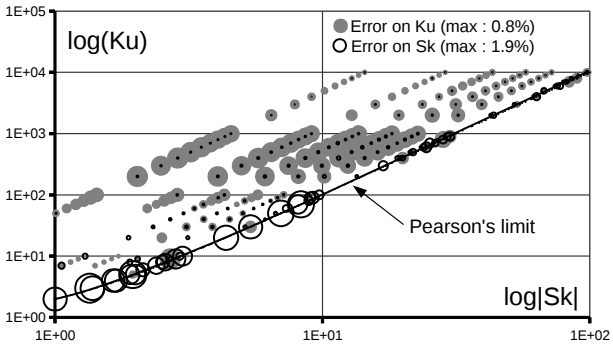
Figure 22: error map (%) of the hybrid method. (*Sk,Ku)* in [0,100]x[1,10001]. Error on *Sk* (1), and *Ku* (2)

The errors (%) on Sk (1) and Ku (2) remain at small values. The error maxima are located in the transitional region because it accumulates the errors of both tangent and exponential distributions, as previously explained. However, this intermediate region is not representative of real surfaces: it is only needed as calculus intermediate at the end of Step 3.

Because (a,b) ensures an analytical ($Sk_{a,b}$, $Ku_{a,b}$) close to the desired value, the quality of the transformation into a data η set must now be assessed.

($Sk_{a,b}$, $Ku_{a,b}$) → (Sk_η, Ku_η)

This validation part is more time consuming, so the domain is not covered with so many points: the points are scattered across the domain as shown on Fig.23. The differences ($Sk_{a,b}$- Sk_η), and ($Ku_{a,b}$- Ku_η) are negligible: the errors don't exceed 1.9% on $Sk_{a,b}$ and 0.8% on $Ku_{a,b}$.



One can conclude that the hybrid analytical/numerical method is very accurate. It is furthermore efficient because the results are obtained in no time, which makes it possible to carry out large parametric studies.

It can be finally concluded that the hybrid approach is able to generate any (Sk_η, Ku_η) data set, in a range of use much larger than Hill's.

After that the data set is generated, with the desired statistical moments, it is randomly shuffled and ready for digital filtering of Step 5.

5.3. Use of an ordering operator, improvements (c)

At the end of Step 5, the spatial characteristic of the generated surface are rigorously the same as the prescribed one. However, the statistical properties can have turned to be far from those wanted because of the approximations of Step 3.

Let O_z be the height ordering operator of the above surface. By applying O_z^{-1} on the ordered data, the starting surface would be obviously retrieved. So, an ordered data set with the right statistical moments is generated, thanks to the hybrid method and, O_z^{-1} is applied. As a result, the final statistical moments are rigorously the same as those prescribed, and the autocorrelation function remains very close to the prescribed one.

To illustrate the aforementioned operation, let us start with an initial profile, continuous line, Fig.24.

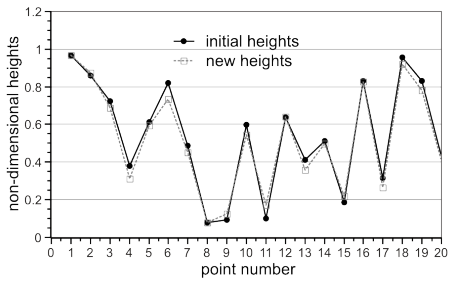


Figure 24: initial profile and corrected profile

The initial profile is ordered and the associated reverse operator is applied to a smoother ordered profile, dashed line, Fig.25.

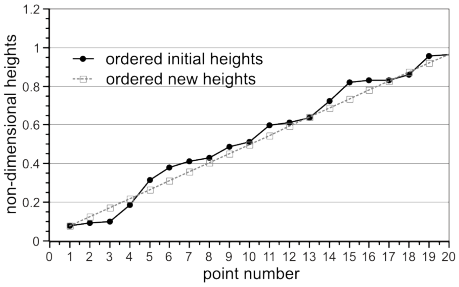


Figure 25: ordered profiles

As a result, a slightly different, corrected profile is obtained, dashed line Fig.24.

Fig.26 is a recap of the overall generating process; it enlightens the principal differences with the classical algorithm along the gray boxes. The operator “transform” is used to stack matrix data on vectors and vice versa.

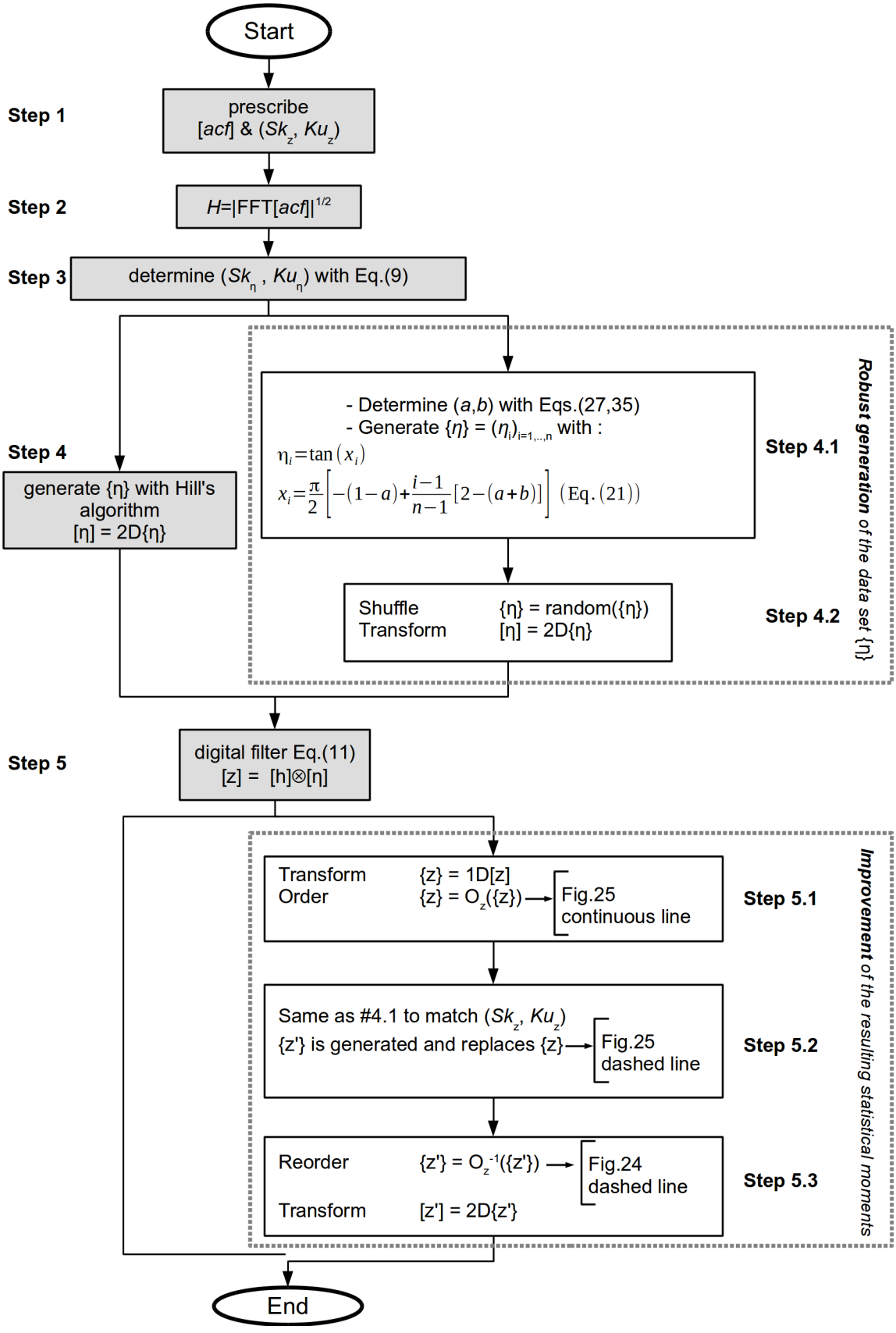


Figure 26: overall generating process

As the data set η is randomly shuffled, after the filtering, the resulting *Sk* can be more or less close to the final desired value. The authors propose to carry out a few tries, three or four are enough, and keep the one which yields the best results. This way, *z*' will be closer to *z* and its autocorrelation function closer to the prescribed one.

5.4. Results

The hybrid method is used to generate surfaces that share exactly the same statistic and spatial parameters as the reference surfaces, Figs.27,28. The final heights {*z*'}, step 5.2 Fig.26, are not numerically generated but recover from the original surfaces.

It can be noticed that:

- the height scales are not the same as those of Fig.3: it is just because the heights are normalized with respect to their standard deviation.
- The generated Ech03, Fig.27(1) exhibits vertical lines and Ech14, Fig.27(2) shows artifacts (like pin holes). Because of the use of FFTs, it is assumed that the surfaces are periodic. This assumption remains valid for short (β_X, β_Y) even if the roughness has a clear orientation, Fig.27(3). However if a long wavelength pattern is present with an orientation different from 0 [π/2], the non-periodicity gap makes artifacts appear.

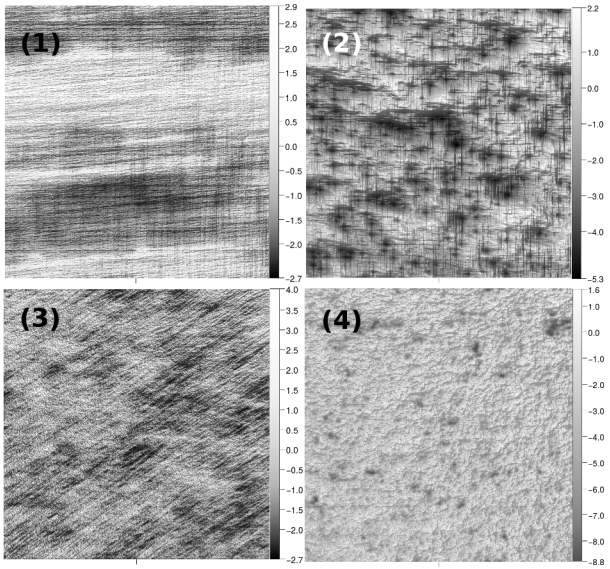


Figure 27: Generated surfaces matching the reference surfaces of Fig.3 – (1) Ech03, (2) Ech14, (3) Fluorin, (4) Rotor

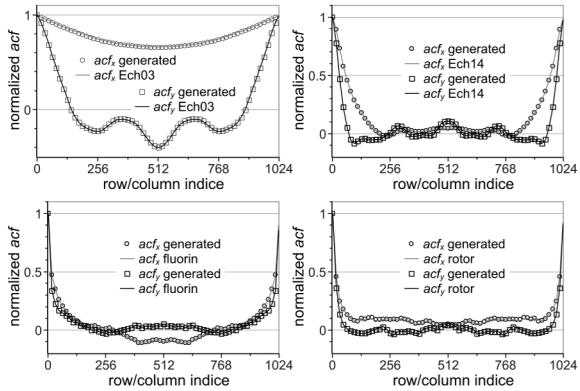


Figure 28: Generated and reference surface *acf* comparison

An important limitation seems to be brought out from the use of FFTs but:

- it only deals with existing surfaces which have long wavelength patterns,
- in addition, one only needs to rotate the original surface so that the pattern orientation is 0 [π/2], as done below.

Ech03 is rotated to ensure that its principal axes are parallel to *x* and *y* (α=0). As a consequence the

statistical moments change, but it remains the very same surface Fig.29(1). It is a very challenging case because Ku_z<3, along with long wavelengths, Fig.30; Eq.(9) yields a Ku_η value much below 1, which is mathematically impossible. So, the hybrid method keeps the Sk_η value given by Eq.(9), with a modified Ku_η, $Ku_\eta = Sk_\eta^2+1$. At the end of the process, z exhibits the right spatial characteristics but its statistical parameters do not exactly match the prescribed ones. However, as z is replaced by z', having the right (Sk,Ku), the final surface is very close to the original surface from both statistical and spatial points of view, as shown on Figs.29,30.

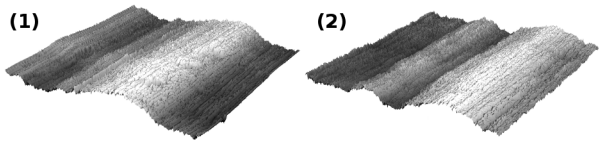


Figure 29: rotated Ech03 (1), generated surface (2); *Sk*=0.37, *Ku*=2.03

Some marked peaks can appear on the generated surface, then to be utilizable in a realistic context, the surface has to be manually cleared of the higher peaks. It will affect a little its characteristics but the result will be a good compromise between the original surface look and the original surface characteristics.

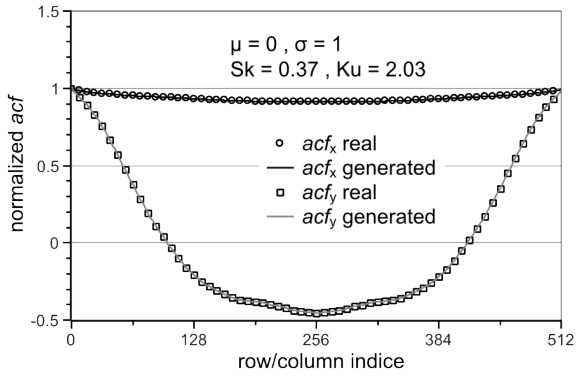


Figure 30:normalized autocorrelation function along x and y axes for the case of Fig.29

Anyway, this kind of surface should not be generated as is; the sinus-like wave should be subtracted before any numerical treatment, to work on shorter wavelengths. However it proves that the hybrid method can also handle very tricky cases.

As concerns theoretical surfaces, Fig.31 is an illustration of the hybrid method accuracy. Four surfaces are generated, with n=512x512, μ=0, σ=1, Sk=-3 and Ku=15. The spatial characteristics are as follows: (1) $\beta_X/512=\beta_Y/512=0.02$; (2) $\beta_X/512=0.04$, $\beta_Y/512=0.02$; (3) $\beta_X/512=0.12$, $\beta_Y/512=0.08$; (4) $\beta_X/512=0.48$, $\beta_Y/512=0.32$

The statistical moments of the generated surface are obtained with an error lower than 0.5% The computing time is higher compared to Hill's algorithm but it remains of the order of a second, on a common computer. The case (4) is the most severe one but it can be seen on Fig.32 that the autocorrelation functions along x and y axes are perfectly matched.

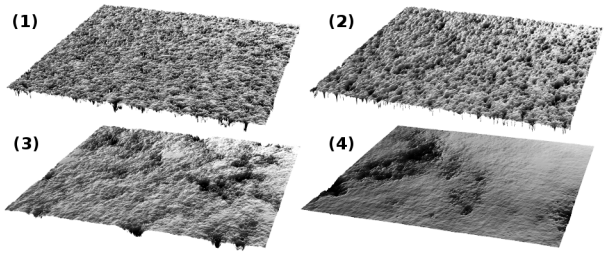


Figure 31: from top left to bottom right, Sk=-3, Ku=15 (1) $\beta_X/512=\beta_Y/512=0.02$ (2) $\beta_X/512=0.04,\beta_Y/512=0.02$ (3) $\beta_X/512=0.12,\beta_Y/512=0.08$ (4) $\beta_X/512=0.48,\beta_Y/512=0.32$

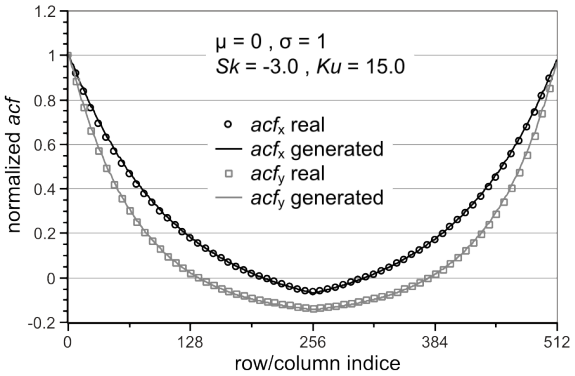


Figure 32: normalized autocorrelation function along *x* and *y* axes for the most severe case (long *acf* lengths).

6. Conclusion

From a general point of view, the gap in accuracy and efficiency of the conventional method, in the task of generating rough surfaces, is filled.

First, the traditional technique fails for numerous prescribed (*Sk, Ku*) cases. Many more become unsolvable if the surface to be modeled has long wavelengths. Second, the conventional method is based on the assumption of infinite series, which worsens the results of small series: a lot of tries are required to match the desired results.

An original and solid approach is then proposed. It is able :

- to handle any (*Sk, Ku*) case, even for small series (provided that $Ku<n$, as explained in Appendix 3)
- to generate rough surfaces with high correlation lengths without any accuracy loss.

The accuracy is proved to be highly satisfactory as well as the efficiency, because of very short computing times.

7. References

[1] Bhushan B. Surface Roughness Analysis and Measurement Techniques. In: Bushan B (ed) *Modern Tribology Handbook, vol.1. 1st ed. CRC Press;2000, chap 2.*

[2] Minet C, Brunetière N, Tournerie B and Fribourg D. Analysis and Modeling of the Topography of Mechanical Seal Faces. *Trib. Trans.* 2010; 53(6): 799-815.

[3] Bakolas V. Numerical Generation of Arbitrarily Oriented non-Gaussian Three-dimensional Rough Surfaces. *Wear* 2003; 254(5): 546-554.

[4] Whitehouse DJ. *Handbook of Surface and Nanometrology.* 2nd ed. CRC Press, 2010, p.73.

[5] Patir N. A Numerical Procedure for Random Generation of Rough Surfaces. *Wear* 1978; 47(2): 263-277 .

[6] TW Kim, Bhushan B and Cho YJ. The Contact Behavior of Elastic/Plastic non-Gaussian Rough Surfaces. *Tribology Letters* 2006; 22(1): 1-13.

[7] Chilamakuri SK and Bhushan B. Contact Analysis

of non-Gaussian Random Surfaces. *Proc. IMechE, Part J: J. Eng. Trib.* 2010; 212(1): 19-32.

[8] Hu YZ and Tonder K. Simulation of 3-D Random Rough Surface by 2-D Digital Filter and Fourier Analysis. *Int. J. Mach. Tools Manufact.* 1992; 32(1-2): 83-90.

[9] Wu JJ. Simulation of non-Gaussian Surfaces with FFT. *Trib. Int.* 2004; 37(4): 339-346.

[10] Reizer R. Simulation of 3D Gaussian surface topography. *Wear* 2011; 271(3-4): 539-543.

[11] Luo T, Liu X, Chen Y, Huang W and Liu Z. Design and Laser Fabrication of Controllable non-Gaussian Roughness Surfaces at Microscale. *App. Surf. Sci.* 2013; 276(1): 95-100.

[12] Sedlačeka M, Podgornik B, Vižintin J. Correlation Between Standard Roughness Parameters Skewness and Kurtosis and Tribological Behavior of Contact Surfaces. *Trib. Int.* 2012; 48: 102–112.

[13] Pearson K. Mathematical Contributions to the Theory of Evolution, XIX; Second Supplement to a Memoir on Skew Variation. *Phil. Trans. Roy. Soc. (A)* 1916; 216: 432.

[14] Wikipedia. Moment (mathematics), http://en.wikipedia.org/wiki/Moment_(mathematics)

[15] Petropoulos GP, Pandazaras CN and Davim JP. Surface Texture Characterization and Evaluation Related to Machining. In: Davim JP (ed) *Surface Integrity in Machining.* Springer Science & Business Media, 2010, pp 37-66.

[16] Hill ID, Hill R and Holder RL. Fitting Johnson Curves by Moments. *Appl. Stat.* 1976; 25(2) :180-189.

[17] Nikias CL, Mendel JM. Signal Processing with Higher-Order Spectra. *Signal Processing Magazine*, IEEE, 1993;10(3): 10-37.

[18] Mendel JM. Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications. *IEEE Proc.*, 1991;19: 278-305.

[19] Jaynes ET. *Probability Theory: The Logic of Science.* Cambridge University Press, Cambridge, 2003.

[20] Kim TH, White H. On More Robust Estimation of Skewness and Kurtosis: Simulation and Application to the S&P 500 Index. *Finance Research Letters* 2004; 1(1): 56-73.

8. Appendix 1

Notation

a, b integration parameters

acf_z $acf(z)$, z autocorrelation function

c_p^{η} η pth-order cumulant

E mathematical expectancy

h bimodal distribution of the intermediate zone ; interval length of the discretized interval $I_{i\ i=1,2,3}$

h, H digital filter, H=FFT(h)

$I_{1,2}$, $I_{1,n-1}$, $I_{1,n}$, $I_{2,n}$, $I_{n-1,n}$ decomposition of the interval $I_{i\ i=1,2,3}$

I_1, I_2, I_3 *tan*, *atanh* and *exp* function *resp.* intervals

$\mathrm{Int}_{i,\ i=1,\ 2,\ 3,\ 4}$ analytically calculated integrals

m_p^{η} η *p*th-order moment function

m_p^{G} *p*th-order moment function of an η equivalent Gaussian signal

n number of points of the distributions

O_z height-ordering operator

p number of rows/columns of a surface

Sk, Ku skewness and kurtosis

($Sk_{a,b}$, $Ku_{a,b}$) theoretical skewness and kurtosis

(Sk_{zc}, Ku_{zc}) calculated *z* skewness and kurtosis

($Sk_{\eta c}$, $Ku_{\eta c}$) calculated η skewness and kurtosis

t, e tangent and exponential distributions *resp.*

Va variance, $Va = \sigma^2$

x, y image directions, horizontal and vertical *resp.*

x_a, x_b, x_c, x_d tangent fitting parameters.

z, Z final random distribution, Z=FFT(*z*)

z' corrected final random distribution

α principal direction of the heights

α, β integration bounds

β_X , β_Y decay factors along *x* and *y resp.*

γ parameter that links $(e_i)_{i=1,\ldots,n}$ and $(t_i)_{i=1,\ldots,n}$

γ_p^{η} η *p*th-order centred statistical moment

η, *A* starting random distribution, *A*=FFT(η)

ξ, λ, γ, δ Johnson's transformation parameters

φ Gaussian noise

μ mean

μ_i i^{th} statistical moment

σ standard deviation

τ decay factor ($\tau=\beta_X=\beta_Y$)

$\boldsymbol{\tau}_{p-1}=\tau_1,\tau_2,\ldots,\tau_{p-1}$ HOS set of lags

Δx, Δy shifting quantities along *x* and *y resp.*

Acronyms

acf autocorrelation function

FFT Fast Fourier Transform

HOS Higher-Order Statistics

iid independent identically distributed random variables

pdf probability density function

9. Appendix 2

Relationships between cross-products and non-centred moments

Let us first define the four non-centred moments of $(x_i)_{i=1,\ldots n}$:

$$\mu_x=\mathrm{E}(x);\ V_x=\mathrm{E}(x^2);\ S_x=\mathrm{E}(x^3);\ K_x=\mathrm{E}(x^4) \quad (36)$$

Developing $(\mu_x)^2$, the following relationship holds

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} x_i x_j = n\mu_x^2 - V_x \quad (37)$$

Developing $V_x\mu_x$, the following relationship holds

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} x_i^2 x_j = n\mu_x V_x - S_x \qquad (38)$$

Developing $(\mu_x)^3$, the following relationship holds

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\sum_{k=1,k\neq i\neq j}^{n} x_i x_j x_k = n^2\mu_x^3 - 3n\mu_x V_x + 2S_x \qquad (39)$$

Developing $(V_x)^2$, the following relationship holds

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} x_i^2 x_j^2 = n V_x^2 - K_x \qquad (40)$$

Developing $S_x\mu_x$, the following relationship holds

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} x_i^3 x_j = n\mu_x S_x - K_x \qquad (41)$$

Developing $V_x\mu_x^2$, the following relationship holds

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\sum_{k=1,k\neq i\neq j}^{n} x_i^2 x_j x_k = n^2\mu_x^2 V_x - n(2\mu_x S_x + V_x^2) + 2K_x \qquad (42)$$

Developing μ_x^4, the following relationship holds

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\sum_{k=1,k\neq i\neq j}^{n}\sum_{l=1,l\neq k\neq i\neq j}^{n} x_i x_j x_k x_l = n^3\mu_x^4 - 6n^2\mu_x^2 V_x + n(8\mu_x S_x + 3V_x^2) - 6K_x \qquad (43)$$

Relationships between non-centred moments and central statistical moments

$$\begin{aligned} Va_x &= V_x - \mu_x^2 \\ Sk_x &= \frac{S_x - 3\mu_x V_x + 2\mu_x^3}{(V_x - \mu_x^2)^{3/2}} \\ Ku_x &= \frac{K_x - 4\mu_x S_x + 6\mu_x^2 V_x - 3\mu^4}{(V_x - \mu_x^2)^2} \end{aligned} \qquad (44)$$

Relationships between z and η statistical moments

Mean

Linking μ_z and μ_η is an easy task, because of the summation over i of η_{k+i} in Eq.(45).

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} z_i = \sum_{k=1}^{n} h_k \frac{1}{n}\sum_{i=1}^{n}\eta_{k+i} = \sum_{k=1}^{n} h_k \times \frac{1}{n}\sum_{i=1}^{n}\eta_i \\ &\Rightarrow \mu_z = \sum_{k=1}^{n} h_k \times \mu_\eta = 0 \end{aligned} \qquad (45)$$

Standard deviation

Remarking that

$$\frac{1}{n}\sum_{i=1}^{n}\eta_{j+i}\eta_{k+i} = \frac{1}{n}\sum_{i=1}^{n}\eta_i\eta_{p+i}, \text{ with } p = k-j \neq 0$$

and with the assumption that,

$$\frac{1}{n}\sum_{i=1}^{n}\eta_i\eta_{p+i} \approx \left[\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\eta_i\eta_j\right] / (n-1) \text{ with } j = p+i$$

and using Eq.(37), the relationship between the variances of the input signal η, the filter h and the output z is as follows:

$$\begin{aligned} z_i &= \sum_{j=1}^{n} h_j \eta_{j+i} \\ z_i^2 &= \sum_{j=1}^{n} h_j \eta_{j+i} \cdot \sum_{k=1}^{n} h_k \eta_{k+i} \\ &= \sum_{j=1}^{n}\sum_{k=1}^{n} h_j h_k \eta_{j+i}\eta_{k+i} \\ \frac{1}{n}\sum_{i=1}^{n} z_i^2 &= \sum_{j=1}^{n}\sum_{k=1}^{n} h_j h_k \left(\frac{1}{n}\sum_{i=1}^{n}\eta_{j+i}\eta_{k+i}\right) \\ &= \sum_{j=1}^{n} h_j^2 \left(\frac{1}{n}\sum_{i=1}^{n}\eta_{j+i}^2\right) \\ &+ \sum_{j=1}^{n}\sum_{k=1,k\neq j}^{n} h_j h_k \left(\frac{1}{n}\sum_{i=1}^{n}\eta_{j+i}\eta_{k+i}\right) \\ &\approx n V_h V_\eta + n(n\mu_h^2 - V_h)\left(\frac{-V_\eta}{n-1}\right) \\ &\approx n Va_h V_\eta = n Va_h = Va_z \end{aligned} \qquad (46)$$

remembering that η has zero mean and unit variance.

Skewness

With Eqs.(38,39) relationships, owing to the following assumption:

$$\frac{1}{n}\sum_{i=1}^{n}\eta_i\eta_{p+i}\eta_{q+i}\approx\frac{1}{(n-1)(n-2)}\left[\frac{1}{n}\sum_{i,p,q=1,i\neq p\neq q}^{n}\eta_i\eta_p\eta_q\right]\approx\frac{1}{n^2}\left[\frac{1}{n}\sum_{i,p,q=1,i\neq p\neq q}^{n}\eta_i\eta_p\eta_q\right]$$

the z third non-centred moment

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}z_i^3&=\sum_{j,k,l=1}^{n}h_jh_kh_l\left(\frac{1}{n}\sum_{i=1}^{n}\eta_{j+i}\eta_{k+i}\eta_{l+i}\right)\\&=\sum_{j=1}^{n}h_j^3\left(\frac{1}{n}\sum_{i=1}^{n}\eta_{j+i}^3\right)\\&+3\sum_{j,k=1\,j\neq k}^{n}h_j^2h_k\left(\frac{1}{n}\sum_{i=1}^{n}\eta_{j+i}^2\eta_{k+i}\right)\\&+\sum_{j,k,l=1\,j\neq k\neq l}^{n}h_jh_kh_l\left(\frac{1}{n}\sum_{i=1}^{n}\eta_{j+i}\eta_{k+i}\eta_{l+i}\right)\end{aligned}\tag{47}$$

results in,

$$Sk_z\approx n^{-\frac{1}{2}}Sk_hSk_\eta\tag{48}$$

Kurtosis

With Eqs.(40-43) relationships, owing to the following assumption:

$$\frac{1}{n}\sum_{i=1}^{n}\eta_i\eta_{j+i}\eta_{p+i}\eta_{q+i}\approx\frac{1}{(n-1)(n-2)(n-3)}\left[\frac{1}{n}\sum_{i,j,p,q=1,i\neq j\neq p\neq q}^{n}\eta_i\eta_j\eta_p\eta_q\right]\approx\frac{1}{n^3}\left(1+\frac{6}{n}\right)\left[\frac{1}{n}\sum_{i,p,q=1,i\neq p\neq q}^{n}\eta_i\eta_j\eta_p\eta_q\right]$$

the z fourth non central moment

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}z_i^4&=\sum_{j,k,p,q=1}^{n}h_jh_kh_ph_q\left(\frac{1}{n}\sum_{i=1}^{n}\eta_{j+i}\eta_{k+i}\eta_{p+i}\eta_{q+i}\right)\\&=\sum_{j=1}^{n}h_j^4\left(\frac{1}{n}\sum_{i=1}^{n}\eta_{j+i}^4\right)\\&+4\sum_{j,k=1,k\neq j}^{n}h_j^3h_k\left(\frac{1}{n}\sum_{i=1}^{n}\eta_{j+i}^3\eta_{k+i}\right)\\&+3\sum_{j,k=1,k\neq j}^{n}h_j^2h_k^2\left(\frac{1}{n}\sum_{i=1}^{n}\eta_{j+i}^2\eta_{k+i}^2\right)\\&+6\sum_{j,k,p=1,p\neq k\neq j}^{n}h_j^2h_kh_p\left(\frac{1}{n}\sum_{i=1}^{n}\eta_{j+i}^2\eta_{k+i}\eta_{p+i}\right)\\&+\sum_{j,k,p,q=1,q\neq j\neq k\neq p}^{n}h_jh_kh_ph_q\left(\frac{1}{n}\sum_{i=1}^{n}\eta_{j+i}\eta_{k+i}\eta_{p+i}\eta_{q+i}\right)\end{aligned}\tag{49}$$

results in,

$$Ku_z-3\approx n^{-1}(Ku_h-3)(Ku_\eta-3)\tag{50}$$

As expected, Gaussian noise (Sk=0, Ku=3) is transformed into a Gaussian data set, because its third and fourth moments remain unchanged by digital filtering.

10. Appendix 3

Maximum of kurtosis

As the kurtosis increases when the distribution becomes "taller and skinnier", its maximum is reached when a single point as a different value from all other points:

$$\eta_i=\begin{cases}-k\ ,\ \text{i=1},\ k\in\mathbb{R}_+^*\\0\ ,\ \text{i=2...n}\end{cases}\tag{51}$$

then

$$\begin{aligned}\mu&=\frac{-k}{n}\\\sigma&=\frac{k}{n}\sqrt{n-1}\\Ku&=\frac{1}{n-1}(n^2-3n+3)\approx n-2\\Sk&\approx\sqrt{n-3}\ \text{(binary distribution)}\end{aligned}\tag{52}$$

11. Appendix 4

Bimodal distribution

Let the subscripts t and e be related to a tangent and exponential distribution respectively. Then, mixing the two distributions with the probability p and (1-p) respectively will rise the following relationships between the statistical moments, [20].

$$
\begin{aligned}
\mu_h &= p\,\mu_t + (1-p)\,\mu_e \\
\sigma_h^2 &= p(\sigma_t^2 + \delta_t^2) + (1-p)(\sigma_e^2 + \delta_e^2) \\
Sk_h &= p(Sk_t\sigma_t^3 + 3\delta_t\sigma_t^2 + \delta_t^3) + \\
&\quad (1-p)(Sk_e\sigma_e^3 + 3\delta_e\sigma_e^2 + \delta_e^3) \\
Ku_h &= p(Ku_t\sigma_t^4 + 4\,Sk_t\delta_t\sigma_t^3 + 6\,\delta_t^2\sigma_t^2 + \delta_t^4) + \\
&\quad (1-p)(Ku_e\sigma_e^4 + 4\,Sk_e\delta_e\sigma_e^3 + 6\delta_e^2\sigma_e^2 + \delta_e^4)
\end{aligned}
\qquad (53)
$$

where $\delta_t = \mu_t - \mu_h$ and $\delta_e = \mu_e - \mu_h$.

Contents