Statistical Approaches to Description of Rough Engineering Surfaces at Nano and Microscales

Feodor M. Borodich\textsuperscript{a}, Andrey Pepelyshev\textsuperscript{b} and Ovidiu Savencu\textsuperscript{a,c}

\textsuperscript{a} School of Engineering, Cardiff University, The Parade, Cardiff CF24 0AA, UK
\textsuperscript{b} School of Mathematics, Cardiff University, Senghennydd Road, Cardiff CF24 4AG, UK
\textsuperscript{c} Austrian Center of Competence for Tribology, Viktor Kaplan-Straße 2, 2700 Wiener Neustadt, Austria

Abstract

Statistical models of rough surfaces are widely used in tribology. These models include models based on assumption of normality of the asperity heights or similar assumptions that involve Gaussian distributions, models based solely on properties of the power spectral density of the surface heights along with models based on assumption of fractal character of roughness. It is argued that models describing surface roughness solely by its fractal dimension or its autocorrelation function (or its power spectral density) do not reflect tribological properties of surfaces. Then typical experimental data obtained for rough engineering surfaces prepared by grinding have been studied at nano and microscales. The heights of the micro-asperities were determined by a profilometre (stylus), while the data for nano/atomic scale was obtained by AFM (Atomic Force Microscopy). The assumption of the normal distribution for the roughness heights has been studied by application of various modern tests of normality. It was found that the height distribution of the surfaces under investigation were not Gaussian at both nano and microscales. Hence, the statistical models of rough surfaces under consideration cannot be used for description of the surfaces and there is a need in critical re-examination of the current statistical approaches.

Keywords: Roughness, normality tests, power spectral density, fractals, nanoscale, microscale

1. Introduction

It is well established that topography of solid surfaces involves finite scale roughness regardless of preparation method of the surfaces (Whitehouse and Archard, 1970, Khusu et al. 1975, Goryacheva 1997). In precision engineering practice surface roughness is a key factor in determining successful component performance, particularly in tribological applications (e.g. gears, bearings). Friction, wear and energy dissipation during sliding are strongly influenced by asperity deformation which is, in turn, controlled by the surface profile. Modern nanotechnology considers surfaces whose roughness is below the micro-metre scale and traditional statistical approaches to surface roughness have to be improved to achieve further progress in studying interactions at the nano/micro scales.

Often an analysis of rough surfaces starts from Fourier decomposition of the surface shape using sine and cosine functions. These functions create one of complete orthogonal sets of
functions. If a surface profile is decomposed using non-trigonometric orthogonal sets of functions then this decomposition is called wavelet transform. A surface created by the Fourier synthesis may be described as

\[ z(x) = \sum_{n=1}^{n_{\text{max}}} a_n \sin \left( \frac{2\pi n x}{2L} + \psi_n \right) \]  

where \( \psi_n \) are random phases, and \( n_{\text{max}} \) is the number of used waves. Typically \( a_n = a_1/n \) for engineering surfaces (see, e.g. Greenwood 1992, Morales-Espejel et al. 2000). The nominal shape of the surface is described normally by the longest spatial wavelength. Then the short wavelength shapes are referred to as ”roughness”, and the long wavelength shapes represent ”waviness” of the surface. One of drawbacks of the Fourier synthesis is that the region of definition of sine and cosine functions is infinite. Hence, the obtained surface is periodic, while the real surfaces are bounded domain where they are defined. On the other hand, wavelets are not equal to zero only within bounded regions. Hence, the boundary will affect only the narrow subregions near the boundaries of surfaces synthesized using the wavelet transform (Borodich and Bianchi 2013).

Modern contact solvers allow the researchers to simulate tribological problems with prescribed surface topography. Currently the topography may be described up to atomic scale resolution. However, the results of the simulations would be of little use because one needs to understand which parameters are governing for the tribological process under consideration and how the process behaviour will change when the parameters are varied (Lubrecht and Venner 1999, Borodich and Bianchi 2013). Nowadays many parameters are used for characterization of the roughness (Nowicki 1985). As Whitehouse (1982) noted some of these parameters are useful, but most are not. This confirms that there is no clear understanding what characteristics of rough surfaces are governing ones for tribological processes. Thus, one needs to understand the governed statistical characteristics of surfaces that determine the tribological behaviour at different scales of the real roughness. The present paper is one step in this direction. Various approaches for statistical analysis and representations of rough surfaces are discussed. The studies applied modern procedures for statistical analysis of roughness at nano and microscales. The tests of normality have been applied to experimental data obtained by the state-of-the-art experimental methods: atomic force microscopy (AFM) measurements of subnano/nano roughness and stylus measurements of microscale roughness of metallic surfaces.

The list of surface characteristics used includes various parameters and functions related to the asperity shapes and distributions of the asperities (Nowicki 1985), e.g. height parameters, such as the maximum height of the profile; the root mean square (rms) parameters, such as the rms slope or the rms curvature; parameters associated with horizontal distributions, e.g. the number of intersections of the surface with the average line; parameters describing spatial extend of asperities, such as the high spot count, and so on. The maximum height of the profile \( R_{\text{max}} \), the arithmetical mean deviation of the surface \( R_a \), and the rms height \( R_q \) are the main statistical parameters for a function \( z(x) \) describing the rough profile within an interval \([-L, L] \):

\[
R_a = \frac{1}{2L} \int_{-L}^{L} |z(x)| \, dx, \quad R_{\text{max}} = \max_{x \in [-L, L]} z(x), \quad R_q = \left[ \frac{1}{2L} \int_{-L}^{L} [z(x)]^2 \, dx \right]^{1/2}.
\]
Abbott and Firestone (1933) suggested to calculate the integrals of the horizontal line at a specific level, which lies within the roughness profile, i.e. they introduced in tribology another important parameter of roughness (in fact, it is related to the cumulative distribution function of the surface heights). The parameter is called the bearing area curve or the Abbott-Firestone curve. Its value at a level \( z \) is equal to the length (the area in 2D problem) of the slice of the profile at the level \( z \) (Abbott and Firestone 1933, Greenwood 1992).

In addition to the above parameters, it was suggested to consider the surface roughness as a random process or signal (see, e.g. Linnik and Khusu 1954, Longuet-Higgins (1957a,b), Whitehouse and Archard (1970), Nayak (1971, 1973), Khusu at al. 1975, Greenwood 1992). However, these approaches considered mainly microscale roughness and did not split the nano and microworlds, while it is important to realize that at angstrom, nano and micrometer scales different interaction mechanisms are involved in tribological processes.

Linnik and Khusu (1954) and Whitehouse and Archard (1970) suggested to study rough surfaces employing the profile auto-correlation function \( R(\delta) \). Whitehouse and Archard (1970) argued that a profile of a random rough surface can be represented by the waveform of a random signal that is completely defined by two parameters: a height distribution and an auto-correlation function

\[
R(\delta) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} [z(x + \delta) - \bar{z}] [z(x) - \bar{z}] \, dx = \langle [z(x + \delta) - \bar{z}] [z(x) - \bar{z}] \rangle,
\]

(2)

The latter parameter can be substituted by its Fourier transform – the power spectral density \( G(\omega) \):

\[
G(\omega) = \frac{2}{\pi} \int_{0}^{\infty} R(\delta) \cos \omega \delta \, d\delta \quad \text{and} \quad \bar{z} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} z(x) \, dx
\]

(3)

where \( \bar{z} \) is the average value (the mean line) of the profile function \( z(x) \), \( \omega \) is the signal frequency and \( \delta \) is the lag. The moments \( m_n \) of the spectral density \( G(\omega) \) provides additional parameters of the surface roughness

\[
m_n = \int_{\omega_0}^{\infty} \omega^n G(\omega) \, d\omega.
\]

Here the profile length \( \lambda_0 \) correspond to the wavenumber \( \omega_0 = 2\pi/\lambda_0 \). For studies of rough surfaces, the structure function \( S(\tau) \) given by

\[
S(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} [z(x + \tau) - z(x)]^2 \, dx
\]

is also employed in addition to the above described parameters and functions (Whitehouse 2001a).

Employing the random signal approach, Sayles and Thomas (1978) calculated the spectral density functions of many natural surface profiles and presented experimental relations between normalized \( G \) and wavelength. They showed that can be approximately presented as \( G(\omega) = 2\pi\Lambda/\omega^2 \) or

\[
G(\omega) \sim 1/\omega^\psi
\]

(4)

where the exponent \( \psi = 2 \) and \( \Lambda \) is called the topothesy of the surface, It follows from discussions by Whitehouse (2001b) and Borodich (2013b) that the latter term is not well defined.
These results presented in logarithmic coordinates span nearly eight decades in wavelength. The fractal approach to surface roughness was trigged by Berry and Hannay (1978) who argued that the surfaces are fractal and their geometric properties were discussed in detail by Mandelbrot (1977). For combining measurements of roughness for 23 types of various surfaces, the \( \lg(G) \sim \lg(\omega) \) relation was presented as a united line spaned nearly eight decades in wavelength. This graph was the cause of many claims that surface topography has fractal behaviour through many orders of magnitude. Then the fractal approach was promoted by Berry and Lewis (1980) by detailed discussion of the Weierstrass-Mandelbrot function and by providing a continuous approximation of its power spectrum.

The paper is organized as follows:

In §2 we give some preliminary information concerning various approaches to statistical studies of rough surfaces. In 2.1 we discuss the early approaches based on assumption of normality of the height distribution.

In 2.2 the fractal approaches and other models based on the use of power spectral density functions are critically examined. Although Borodich (1993) and Borodich and Onishchenko (1993) (see also Borodich and Onishchenko 1999, Borodich 2013c) gave several counterexamples showing that the fractal dimension alone cannot reflect the contact properties of the surfaces (in addition, Whitehouse 2001b) gave a very negative appraisal to the use of fractals in tribology, papers devoted to the fractal approach to roughness prefer to avoid any discussion of these arguments. Hence, the fractal approach is discussed in this paper and it is argued many approaches that pretend to be the fractal ones, are based not on scientifically justified information but on myths about fractals and the use of ill-defined terminology, i.e. instead of the state-of-the-art approaches they use the state-of-the-street ones.

In 2.3 it will be shown that the approaches that uses solely the surface power spectrum \( G(\omega) \) for representation of the surface roughness, have in fact the same drawbacks as the above fractal ones. It is argued that the surface power spectrum alone cannot reflect the contact properties of the surfaces.

One of the purposes of the paper is testing of normality of the height distribution of the roughness. This is because the classic statistical approaches (Linnik and Khusu 1954, Longuet-Higgins (1957a,b), Whitehouse and Archard (1970), Nayak (1971, 1973), Khusu at al. 1975) assume normality of the height distribution of the rough surfaces. It is known that normality (or lack thereof) of an underlying data distribution can have an effect to a greater or lesser degree on the properties of estimation or inferential procedures used in the analysis of the data (Thode 2002). Thus, in 2.4 seven established tests of normality of data are described.

In §3 we took for our studies typical experimental data obtained for rough engineering surfaces prepared by grinding. The heights of the micro-asperities were determined by a profilometre (stylus), while the data for nano/atomic scale was obtained by AFM (Atomic Force Microscopy). The assumption of normal (Gaussian) distribution of the roughness heights of engineering surfaces at micro- and nano-scales is re-examined by application of various modern tests of normality of data including the Kolmogorov-Smirnov, Lilliefors, Shapiro-Wilk, Pearson and other tests, i.e. the above mentioned normality tests are used for determining whether a sample of observations (surface heights) could have come from a normal (Gaussian) distribution. It was found that the surfaces under investigation were not Gaussian at both micro- and nano-scales. Hence, the above mentioned classic approaches that used the assumption of normality of roughness, cannot be applied to the considered surfaces.
2. Preliminaries

Here the classic approaches to random rough surfaces are considered along with recent approaches to analysis of rough surfaces that concentrated mainly on fractal approaches and the use of power spectral density functions.

2.1. Statistical approaches to rough surfaces

2.1.1. Early models of rough surfaces

One of the first models of contact between rough elastic solids was presented by Zhuravlev (1940) employing statistical arguments. In particular, he assumed that all asperities are elastic spheres of the same radius $R$ located at various heights, i.e. it was a statistical multi-level model of roughness. He assumed also that the number of asperities at a specific height may be characterized by a distribution function $\phi(\xi)$, i.e. the number increases as one considers the deeper levels $\xi$. Applying the Hertz contact theory, he derived general expressions for the true contact area $A$ and the contact force $P$

$$A = \frac{\pi R}{2N} \int_{0}^{x/2} [x - (\xi_1 + \xi_2)]\phi(\xi_1)\phi(\xi_2)d\xi_1d\xi_2, \quad (5)$$

$$P = \frac{\sqrt{2R}}{3\pi kN} \int_{0}^{x/2} [x - (\xi_1 + \xi_2)]^{3/2}\phi(\xi_1)\phi(\xi_2)d\xi_1d\xi_2. \quad (6)$$

Here $k = (1 - \nu^2)/\pi E$ is the effective contact modulus of surfaces having the elastic modulus $E$ and the Poisson ratio $\nu$, $x$ is the relative approach due to compression of the surfaces, and $N$ is the total number of summits of the protuberances located at depth $x$. As an example, he considered that $\phi(\xi)$ is linear. In this case it follows from (6) that $A \sim P^{10/11}$.

Kragelsky (1948) modified the Zhuravlev approach by modeling a rough surface as a collection of elastic rods of various heights based on a flat rigid surface. He noted that one would need to employ the numerical integration of general expressions, if the Gaussian distribution of rod heights were employed. Greenwood and Williamson (1966) used an exponential distribution of the sphere summits to show that the true contact area $A$ is proportional to the applied compressive load $P$.

In the Archard model (1957) of rough surfaces, a sphere of radius $R_0$ is covered by spherical protuberances of radius $R_1$ that are evenly distributed over the surface of the sphere. These latter types of spheres, in turn, are covered by smaller protuberances of radius $R_2$ ($R_2 \ll R_1 \ll R_0$). Although the model is not based on surface roughness statistics, it reflects a very important observation of multilevel structure of the roughness.

2.1.2. Modelling of surfaces as random processes

The idea to consider an irregular curve as a realization of a random process was applied initially to statistical analysis of signals (see, e.g. Pollock 1999). To the best of our knowledge the first attempt to use the random process approach in tribology is due to Limnik and Khusu (1954). They suggested to represent a profile of polished surface as a sample from some normal stationary process whose auto-correlation function can be approximated as

$$R(\delta) = R(0)\exp(-\alpha|\delta|).$$
Thus the roughness is described by the parameters $R(0)$ and $\alpha$. The asperity heights were calculated from the middle line of the profile, that, in turn, was calculated by the least square method. This was leading to introduction of additional parameters. Using this description, Linnik and Khusu (1954) (see also Khusu et al. 1975) measured (i) the fraction of the horizontal line at a specific level, which lies within the roughness profile, i.e. they calculated the Abbott–Firestone curve, and (ii) the total area of the profile cove above this horizontal line. They reported pretty good agreement of the obtained results with their particular experimental observations. They suggested also to use the Abbe test for checking that the expectations of all values are the same, i.e. there is no trends in data.

Longuet-Higgins (1957a,b) presented a careful statistical analysis of a random surface of moving fluid. Assuming that the surface is Gaussian, he derived many statistical properties of the surfaces. In particular he derived expressions for (i) the probability distribution of the surface elevation; (ii) the average number of zero-crossing per unit distance along a line; (iii) the average length of the contours per unit area; and (iv) the average density of maxima and minima per unit area of the surface. The approach can be applied directly only to liquids only because the surface effects of solids and their nano-scale features differ from properties of liquids.

Nayak (1971, 1973) presented a careful statistical analysis of random surfaces of solids. He compared statistics extracted from a surface and those of a profile. Assuming that the roughness is random, isotropic and Gaussian, he found serious differences in the distributions of heights of maxima and in the mean gradient. He presented functional relations connected the power spectral density for a single profile and the surface. Finally, he presented some statistical results for anisotropic Gaussian surfaces.

As it has been mentioned above, Whitehouse and Archard (1970) argued that a profile of a random rough Gaussian surface can be represented by the waveform of a random signal that is completely defined by two parameters: a height distribution and an auto-correlation function $g(2)$.

Thus, the above classic statistical models that treat rough surfaces as random processes, assume normality of the height distribution. Of course, other distributions were also used, e.g. Zhuravlev (1940) used a linear distribution, while Greenwood and Williamson (1966) employed an exponential distribution of the asperity summits. However, these distributions were introduced mainly for solving some illustrative examples.

Note that to employ models introduced by Zhuravlev (1940) and later by Greenwood and Williamson (1966), one would need to estimate the average radius of surface asperities. If the surface profile $z(x)$ is measured with a constant step $\delta x$, i.e. there is an array $z_k$, then the radius of curvature ($R$) of an asperity $z_k$ may be calculated using $z_{k-1} < z_k < z_{k+1}$ (Greenwood 1992)

$$1/R = -(z_{k-1} + z_{k+1} - 2z_k)/\delta x^2.$$ 

It was found later (see, e.g. Whitehouse and Archard 1970) that different steps gave different values of $R$, i.e. the mean radius of curvature depends on the scale of consideration. One of the reasons that the fractal approach to surface roughness description was introduced, is that there was a hope that this approach provides a scale independent characterization of roughness.
2.2. Fractal Approaches to Surface Roughness Description

Fractal is a general term that is used to describe both some mathematical ideas related to set theory and a collection of semi-empirical approaches based on ideas of similarity (Borodich 2013a). was Mandelbrot (1977) introduced the term fractals by without a proper mathematical definition. One of the authors (FB) suggested to consider as fractals sets having non-integer fractal dimension. He has also suggested to distinguish mathematical and empirical fractals (see, e.g. Borodich 2013a,b). The latter can be also referred to as physical fractals. The use of fractals assumes usually that an object under consideration is covered by spheres (or cubes) of size $\delta$. Mathematical fractals are characterized by their fractal dimensions (FDs), these include the box-counting dimension, the Hausdorff dimension and others. The mathematical FDs can be calculated by calculating limits when $\delta \to 0$. It is assumed often that physical fractals have the so-called number-radius relation that can be written as a power-law

$$N(\delta) \sim \delta^{-D}, \quad (7)$$

when the size of the cover is between the upper $\Delta_*$ and lower $\delta_*$ cut-offs of the fractal behaviour, i.e. $\delta_* \leq \delta \leq \Delta_*$. The value $D$ is referred to as FD, $N(\delta)$ is the number of elements of the cover, and $\delta$ is the current size in the elements.

One can see that mathematical fractals have practically nothing in common with physical fractals. The power law of physical fractals is observed only for scales from a bounded domain with a lower cut-off $\delta_*$, while the mathematical FD can be calculated only for $\delta \to 0$. Therefore, mathematical fractals can be used as models of real life objects, in particular surface topography. However, such models have a number of drawbacks.

To handle fractal surface models in a rigorous manner, one of the authors (FB) introduced two different types of fractal surfaces: the Cantor-Borodich ($C_B$) profile and the parametric-homogeneous (PH) functions (Borodich and Mosolov 1991, 1992, Borodich 1993, 1998a,b). The PH-functions $b_d$ of degree $d$ is defined as

$$b_d(p^k x; p) = p^{kd} b_d(x; p)$$

for any integer $k$. Originally the $C_B$ profile had two scaling parameters $f_x$ and $f_z$ to describe the structure of its asperities. In each generation the width $L_i$ is given by $L_i = L_0 f_x^i$ and the height $h_i$ is $h_i = h_0 f_z^i$, respectively.

Studying solutions to problems of contact for $C_B$ profile and for PH-punches, one can see that a number of statements about fractals are often non-critically adopted in the literature. It is often claimed (see, e.g. Majumdar and Bhushan, 1991) that the Weierstrass-Mandelbrot (WM) function that was introduced by Mandelbrot (1977) and that was discussed in detail by Berry and Lewis (1980)

$$W(x; p) = \sum_{n=-\infty}^{\infty} p^{(D-2)n} (1 - \cos p^nx)$$

reflects the general fractal properties of rough surfaces (here $p$ is a fixed scaling parameter). It is also believed that the scaling properties $W(x; p)$

$$W(px; p) = p^{2-D} W(x; p) \quad (8)$$
are in close connection with the FD of its graph. In fact, the above scaling (8) of \( W(x;p) \) is because the WM function is a PH-function with \( d = (2 - D) \) and parameter \( p \). In addition, one has to realize that the graph of \( W(x;p) \) is not statistically flat but it has a specific trend that can be described as \( x^{2-D} \). Hence, \( W(x;p) \) is a particular example of a fractal PH-function and it cannot be considered as the general fractal model of rough surfaces (Borodich 2013c, Borodich and Bianchi 2013).

Further, it is often assumed that scaling properties of rough surfaces are defined by the exponent \( H \) (the Hurst exponent), and the fractal dimension \( D \) of the surface or a profile can be calculated as \( H = E - D \). Here \( E \) is the Euclidean dimension of the space: \( E = 3 \) for a surface and \( E = 2 \) for a curve. This is usually explained by an example saying that under quasi-homogeneous scaling of coordinates \( x \rightarrow \lambda_1 x \) and \( y \rightarrow \lambda_2 y \) the fractal surfaces appear as approximately the same or

\[ x \rightarrow \lambda x; \quad z(x) \rightarrow \lambda^H z(x). \] (9)

Here \( z(x) \) is the function of heights, while \( \lambda_1 = \lambda > 0 \) and \( \lambda_2 = \lambda^H \) are scaling factors. The exponent \( H \) is referred to as the self-affine exponent or the Hurst exponent. Using the above rather vague definition, it was suggested to call fractals as self-affine or self-similar depending on their statistical invariantness under respectively quasi-homogeneous or homogeneous \((\lambda_1 = \lambda_2)\) coordinate dilation.

In fact, the term 'self-affine fractal' does not have a rigorous mathematical definition. It is known (see, e.g. Davies and Hall, 1999) a random curve, that is described as a stochastic function of \( x \), may be considered as statistically self-similar "if the scaling of its argument can be undone by an appropriate scaling of the curve itself". However, the claims about self-affinity of self-similarity of rough surfaces are not usually supported by studies of their statistical properties. Further, it is known (see for details Borodich 1998a,b) that FDs of mathematical fractal curves will not change under action of any quasi-homogeneous dilation of coordinates, i.e. the graphs of the following one-dimensional fractal PH-curves \( b_1 \) and \( b_2 \)

\[
\begin{align*}
  b_1(x;p) &= xb_0(x;p), \\
  b_2(x;p) &= x^2b_0(x;p), \\
  b_0(x;p) &= x^{-\beta}W(x;p)
\end{align*}
\] (10)

have the same FDs as the original WM function \( W(x;p) \). However, they have respectively degrees \( d = 1 \) and \( d = 2 \). Hence, if one would use the above definitions of self-affine or self-similar fractal curves, then he would need to say that \( b_1(x;p) \) is a self-similar fractal (its graph has a linear trend) and \( b_2(x;p) \) is a self-affine fractal (its trend is described by a quadratic parabola), while actually their FDs are the same. Hence, the term 'Hurst exponent' is ill-defined and the common statement, that the fractal dimension is always closely related to the exponent \( H \) of self-affine fractals, in general is wrong.

As it has been mentioned above, it is claimed often that the empirical fractals are very often observed. In some sense it is true, however, for very short range. Some time ago an extended analysis of data published by Physical reviews journals was performed by a group of researchers from Jerusalem (Malcai et al. 1997, see also Avnir et al. 1998). It follows from their analysis that the overwhelming majority of reported physical fractals have a span of about 1.5 orders of magnitude, i.e. \( \Delta_*/\delta_* = 10^{1.5} \approx 31.6 \). Mandelbrot (1998) argued that these limited-range examples of power law correlations are "unfortunate side effects of enthusiasm, imperfectly controlled by refereeing". However, the Jerusalem group disagreed and argued that
the dominant fractals observed in Nature are the limited-range empirical fractals (Biham et al. 1998). The span of about 1.5 orders of magnitude is also observed in analysis of fractal surfaces (Borodich 2013b). In addition, the values of the surface FD obtained by different methods may differ considerably (Wendt et al. 2002). Whitehouse (2001b) criticized the fractal approaches. He argued that “it is fundamentally wrong to use scale-invariant models to describe or interpret scale-variant mechanisms”. As it has been mentioned above, rough surfaces demonstrate the power-law number-radius relation usually within a 1.5 or 2 orders of magnitude. Thus, empirical fractals do not give scale-invariant parameters for description of rough surfaces (Borodich 2013b, Borodich and Evans 2013).

Mandelbrot underlined that his books on fractals (see, e.g. Mandelbrot 1977) are not written in mathematically rigorous manner. However, it is difficult to deal rigorously even with some statements from mathematical books and papers on fractal. For example, it has been stated (Davies and Hall, 1999) that all the common definitions of fractal dimensions may be used for statistical models of rough surfaces based on the Gaussian-process, and all definitions gave the same numerical value for the FD. However, this statement was not supported by mathematical evidence. Further, speaking about autocorrelation of fractal functions \( R(\tau) \), Falconer (1990) wrote that if \( R(\tau) \) of fractal function \( z \) satisfies \( R(0) - R(\tau) \sim \tau^{2(2-s)} \) for small \( \tau \) then it is reasonable to expect that the box dimension of graph \( z \) equal \( s \) and it is reasonable to expect a graph with the spectral density \( 4 \) to have a graph of fractal dimension \( s = (5 - \psi)/2 \). We believe that such reasons have to be supported by statistical analysis of surfaces.

To conclude the discussion of fractal approaches, let us discuss some contact problems for fractal punches. Although the \( C_B \) structural function has infinite length for \( 0 < f_x < 0.5 \) and \( 0.5 < f_z < 1 \), and therefore, it can be considered as a fractal curve, it does not describe real rough surfaces because this model does not have enough flexibility. It was suggested by Warren and Krajcincovic (1996) to modify the \( C_B \) model and additionally to the scaling parameters \( f_x \) and \( f_z \) of the original Cantor-Borodich profile, to introduce an integer number \( s \) as an additional fitting parameter. This parameter \( s, s \geq 2 \) is equal to the number of segments in each next generation of the structure. They suggested to use the modified \( C_B \) type profiles for synthesizing fractal surfaces. The FD of the modified \( C_B \) model can be calculated as

\[
D = 1 - \frac{\ln f_z}{\ln s f_x} + \frac{\ln s}{\ln s f_x}.
\]

The analysis of solutions to contact problems for both \( C_B \)-type profiles (Borodich and Onishchenko 1993, 1999) and the PH-functions (Borodich 1993, 1998a,b, Borodich and Galanov 2002, Kindrachuk and Galanov 2014) showed that the \( P - h \) relations (here \( h \) is the approach between the punch and an elastic surface; and \( P \) is the external load) do not depend on the fractal dimension of the surfaces. Hence, the fractal dimension of a surface does not determine its contact properties. It will be shown below that the autocorrelation function of a surface does not determine its contact properties neither.

2.3. Approaches to roughness based solely on the surface power spectrum

Let us study the power spectral density for the surface roughness with the profile \( z(x) \). Let \( \bar{z} \) be the mean line of the surface profile function \( z(x) \). Without loss of generality, we let \( \bar{z} = 0 \) because we can always shift the origin of the coordinate axis. The definitions of the
auto-correlation function $R_z(\delta)$ and its Fourier transform, the power spectral density $G_z(\omega)$ has been given above by (2) and (3). Because $\bar{z} = 0$, it follows from (2) that

$$R_z(\delta) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} z(x + \delta)z(x) \, dx.$$ 

Now we consider another profile $\xi(x)$ defined by $\xi(x) = -z(x)$, i.e. we consider a reflected body. Note that $\xi(x)$ is the inverted profile (a complementary profile) compared to the profile $z(x)$. It is easy to see that

$$\bar{\xi} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} y(x) \, dx = -\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} z(x) \, dx = 0.$$ 

Then the auto-correlation function $R_\xi(\delta)$ for the profile $\xi(x)$ has the form

$$R_\xi(\delta) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} [y(x + \delta) - \bar{\xi}][y(x) - \bar{\xi}] \, dx = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \xi(x + \delta)\xi(x) \, dx$$

$$= \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} z(x + \delta)z(x) \, dx = R_z(\delta).$$

Therefore, we have obtained that the profiles $z(x)$ and $\xi(x)$ have the same auto-correlation function, in turn, it follows from (3) that both profiles have the same power spectrum. Thus, the profile has the same auto-correlation function and the power spectrum in both $(x, z)$ and $(x, \xi)$ coordinate systems. However, they may have the different bearing area (Abbott) curves. Hence, the power spectral density as a single parameter does not characterize tribological properties of rough surfaces.

Let us mention another approach to contact problems between randomly rough surfaces introduced by Persson (see, e.g. Persson, 2001). Borodich (2002) argued that this approach is wrong saying that two punches having the same fractal surface, e.g. $C_B$ profile, but situated either above or below the surface show usually different asymptotics in load-displacement relations. Nevertheless, the approach is non-critically adopted for numerical simulations by many researchers. On the other hand, the authors believe the approach does not have any mathematical justification. Persson assumes that the surface is a self-affine fractal whose power spectrum has the power-law form. Hence, the approach has the same drawbacks as the above mentioned fractal ones. However, even if he would not use this assumption, it follows from (11), the approach is still wrong because it is based solely on the use of the surface power spectrum $G(\omega)$ for representation of the surface roughness. Thus, the approach would not describe contact properties of the surfaces.

2.4. Normality test

One of the popular models of roughness is based on the assumption that the roughness profile follows a Gaussian process with some correlation function. In particular, it is assumed that the statistical sample of height measurements is normally distributed. The assumption of normality is especially critical when characteristics of roughness profiles are derived.

The main tests for the assessment of normality are the Kolmogorov-Smirnov (KS) test, the Lilliefors (LF) test, the Shapiro-Wilk (SW) test, the Anderson-Darling (AD) test, the Cramer-von Mises (CVM) test, the Pearson test, the Shapiro-Francia (SF) test (see, e.g. Thode 2002).
Each test is based on a particular test statistic which is a quantitative estimator of proximity between the theoretical normal distribution and an observed sample of measurements. Based on the test statistic, the test determines the \( p \)-value. The \( p \)-value can be interpreted as the significance at the scale \([0,1]\) that the hypothesis of the normality is true for the observed measurements. If the \( p \)-value is less than the acceptable significance level, say 5\%, then the hypothesis of normality is rejected, that is, we conclude that the height distribution is not normal. Note that the acceptable significance level is a probability to reject the hypothesis of normality even if it is true.

To describe the tests, we introduce notations: \( z_1, \ldots, z_n \) is a permutation of the sample \( z_1, \ldots, z_n \) such that \( z_1 \leq \ldots \leq z_n \), \( p(i) = \Phi((z(i) - \bar{z})/s) \), where \( \Phi \) is the cumulative distribution function of the standard normal distribution, and \( \bar{z} \) and \( s \) are mean and standard deviation of the sample.

**The KS test**

The Kolmogorov-Smirnov (KS) test for a sample is based on comparison of an empirical distribution function (edf) \( F_N(t) \) and the theoretical cumulative distribution function (cdf) \( F(t) \) of the test distribution; specifically, the test statistic is

\[
T = \max_z \left| F(z) - F_N(z) \right|
\]

where \( F(z) \) is the cdf of Gaussian (normal) distribution and

\[
F_N(z) = \begin{cases} 
0, & z < z(1), \\
i/n, & z(i) \leq z < z(i+1), \ i = 1, \ldots, n - 1, \\
1, & z \geq z(n),
\end{cases}
\]

is the edf of the sample \( (z_1, z_2, \ldots, z_N) \).

A limitation of the KS test is its high sensitivity to extreme values.

**The LF test**

The Lilliefors (LF) test is a correction of the KS test in a way to be more conservative. The test statistic is the maximal absolute difference between empirical and theoretical cdf as

\[
D = \max \{D^+, D^-\}, \quad D^+ = \max_{i=1,\ldots,n} (i/n - p(i)), \quad D^- = \max_{i=1,\ldots,n} (p(i) - (i - 1)/n).
\]

The LF test is slightly better than the KS test because the LF test is more conservative.

**The AD test**

The Anderson-Darling (AD) test is based on a squared difference between the edf and theoretical cdf, the test statistic is

\[
Z = (1 + 0.75/n + 2.25/n^2) \left( -n - \frac{1}{n} \sum_{i=1}^{n} (2i - 1)(\ln(p(i)) + \ln(1 - p(n-i+1))) \right).
\]

The Anderson-Darling test is sensitive to discrepancies in the tails of the distribution.

**The CVM test**
The Cramer-von Mises (CVM) test is also based on a squared difference between the edf and theoretical cdf. The test statistic is

$$W = (1 + 0.5/n) \left( \frac{1}{12n} + \sum_{i=1}^{n} \left( p(i) - \frac{2i - 1}{2n} \right)^2 \right).$$

The CVM test is uniform to discrepancies in different parts of the distribution.

**The SW test**

The Shapiro-Wilk (SW, or Shapiro) test is based on the correlation between the data and the corresponding normal scores,

$$W = \frac{\left( \sum_{i=1}^{n} c_i z(i) \right)^2}{\sum_{i=1}^{n} (z(i) - \bar{z})^2}, \quad c_i = \frac{m_i}{\sum_{j=1}^{n} m_j^2}$$

where $m_i = \Phi^{-1}((i - 3/8)/(n + 1/4))$. The Shapiro-Wilk test has better power than the KS test and the LF test. The Shapiro-Wilk test is known as the best choice for testing the normality of data.

**The Pearson test**

The Pearson chi-square test is built on creating the classes are build in such a way that they are equiprobable under the hypothesis of normality, where the number of classes is chosen as $m = 2n^{2/5}$. The Pearson statistic is given by

$$P = \sum_{j=1}^{m} \frac{(C_j - E_j)^2}{E_j},$$

where $C_j$ is the number of counted observations and $E_j$ is the number of expected observations under the hypothesis of normality in the $j$-th class.

**The SF test**

The Shapiro-Francia (SF) test is simply the squared correlation between the ordered sample values and the (approximated) expected ordered quantiles from the standard normal distribution, the test statistic is

$$F = \frac{\left( \sum_{i=1}^{n} \mu_i z(i) \right)^2}{\sum_{i=1}^{n} (z(i) - \bar{z})^2},$$

where $\mu_i$ are standard normal ordered statistics.

### 3. Analysis of roughness data

For simulations of friction between the surfaces, Borodich and co-workers have proposed several deterministic models of rough surfaces (Borodich and Mosolov 1991, 1992, Borodich and Onishchenko 1993, 1999). In particular, they have developed a model where a nominally flat slider is represented as a multiscale, hierarchical system (MHS) of connected deformable components where the nano-scale asperities are responsible for adhesive interactions between surfaces, while the micro-scale asperities are the main cause for the interlocking component of friction (Savencu and Borodich 2014, Savencu 2016). This model takes into account not
only the hierarchical multilevel structure of roughness but also the adhesive interactions, the deformation of asperities, transfer the deformations between levels of the hierarchical structure and the vertical degree of freedom of the asperities. In addition, the model takes into account the Polonsky-Keer effect (Polonsky and Keer 1996a,b): ... plastic deformation at an asperity micro-contact becomes difficult and then impossible when the asperity size decreases below a certain threshold value on the order of the microstructural length, i.e. the nano-scale asperities do not have plastic deformations even under very high pressure. Thus, to reflect different mechanisms of elastic/plastic deformations of asperities and adhesive/cohesive interactions between surfaces that in turn depend on the environmental conditions of the surfaces, one needs to study the surface statistics at the appropriate scales. The results of such studies are presented below.

The heights of the micro-asperities were determined in 3D by measuring the surface of metallic samples by a Taylor Hobson Form Talysurf 2 profilometre with the sample mounted on a y-stage to obtain a series of offset profiles and thus a 3D map of surface heights. Similar profiles can be found in the PhD thesis of Weeks (2015). The heights of the nano-asperities were determined by analysing the measurement of the sample surface by the Park XE-100 AFM from Park Systems. This AFM was described in detail by Brousseau et al. (2015).

3.1. Description of original data

Let us consider several datasets obtained by the above described measurements of roughness. In Figure 1, one can see the roughness of a typical copper surface obtained by AFM, where measurements $z_k$ are taken at points $x_k = 0.15625k \, \mu m$, $k = 0, 1, 2, \ldots, K$, $K = 256$. One can observe that the asperities have a zig-zag shape.

Figure 1: The roughness data of a copper sample obtained by AFM, scale units are $\mu m$.

Figure 2 presents the roughness obtained by profilometre on a steel gear surface, where measurements $z_k$ are taken at points $x_k = 1.5k \, \mu m$, $k = 0, 1, 2, \ldots, K_2$, $K_2 = 3334$. One can see that the pattern for the interval $[0, 1000]$ consists of 5 sub-patterns that are very similar to each other; this is because the profilometre measures roughness of 5 parallel segments that are close to each other. Note that the roughness for the 5-th segment moderately differs from other segments. Therefore, we will further study roughness for the 1-st segment, which is shown in Figure 2. One can see that asperities typically have a parabolic shape in contrast to a zig-zag shape of asperities at nano-scale obtained by the AFM measurements.

Let us depict roughness with natural aspect ratio. In Figure 3 roughness measurements obtained by AFM for some segments is shown. One can observe that the distance between
Figure 2: The roughness data of steel obtained by a profilometre for 5 segments, scale units are µm.

Figure 3: The roughness data of steel obtained by a profilometre for the 1-st segment, scale units are µm.

points of measurements looks large and the linear interpolation looks smooth.

Figure 4: The roughness data of copper obtained by AFM for 3 small segments with natural aspect ratio, scale units are µm.

In Figure 5 roughness obtained by a profilometre is shown with natural aspect ratio. One can see that roughness at the micro-metre scale is less smooth than at the nano-scale. The local minimum near \( x = 70 \) that looks very sharp in Figure 3 does not look sharp in Figure 5. This is an effect caused by quasi-homogeneous representation of data in Figure 3, i.e. because the \( x \) and \( z \) coordinates have different scales.

Let us now consider the combination of two roughness obtained at different scales. Specif-
3.2. Descriptive statistics and normality tests

Let us now study the height distribution for roughness data. We consider the series $z_1, z_2, \ldots, z_K$ as a sample and in Figure 7 we show the histogram of heights for roughness data obtained by AFM. One can see that the distribution of heights is not very well fitted by the normal distribution.

In Figure 8 we show the histogram of heights for roughness data obtained by a profilometre. We can see again that the distribution of heights is not very well fitted by the normal distribution. Thus, to get a conclusion if the distribution is normal or not, modern statistics
has developed many tests of normality (Thode 2002). Some of these tests have been described in the previous section.

Let us apply normality tests to roughness measurements at nano and microscales which are drawn in Figures 9–13. These figures present both the roughness profiles and the height distributions. One can observe waviness of original measurements and therefore, we consider also data with removed waviness.

Figure 7: The histogram of heights for roughness data of copper obtained by AFM for the segment [0, 40] with the fitted normal distribution, \( x \)-scale unit is \( \mu m \), vertical scale is counts.

Figure 8: The histogram of heights for roughness data of steel obtained by a profilometre for the segment [0, 1000] with the fitted normal distribution, \( x \)-scale unit is \( \mu m \), vertical scale is counts.

Figure 9: Roughness data of a copper sample obtained by AFM: original (black), waviness (red) and profile with removed waviness (blue); \( x \)-scale unit is \( \mu m \). The height distributions are given on the left and the right respectively.
Figure 10: Roughness data of a bronze sample obtained by AFM: original (black), waviness (red) and profile with removed waviness (blue); x-scale unit is µm. The height distributions are given on the left and the right respectively.

Figure 11: Roughness data of a bronze sample obtained by AFM: original (black), waviness (red) and profile with removed waviness (blue); x-scale unit is µm. The height distributions are given on the left and the right respectively.

We show $p$-values for both data (original and with removed waviness) in Table 1. One can observe that many $p$-values are much smaller that 0.05 and therefore we reject the hypothesis of normality for these measurements.

Note that if the $p$-value for a specific test is larger than 0.05 then this test does not reject the hypothesis of normality. However, one can see that the results for all such tests have weak power. Other tests for the same data yield very small $p$-values. Combining results of several tests for the same data, one can conclude that the hypothesis of normality has to be rejected. Thus, the height distribution of the surfaces under investigation were not Gaussian at both micro- and nano-scales.
4. Conclusion

It has been shown that papers devoted to characterization of surface roughness using fractal approach are often based on wrong assumptions or ill-defined terminology. An example of such an ill-defined term is the Hurst exponent. It has been argued that tribological properties of rough surfaces cannot be characterized by models based solely on the surface fractal dimension. Hence, some rather popular models of fractal contact cannot be used. The same conclusion about wrong formulation of contact problems is applicable to models based solely on the power spectral density of the surface roughness. Because the power spectral density of a rough surface and a complementary surface are the same, this function provides the same characteristics for both surfaces, while their contact properties are different (see also a discussion by Borodich 2002). Indeed, valleys of a polished surface become hills of the complementary surface. To feel
Table 1: The p-values of various normality tests for several roughness measurements at nano and microscales for several metallic samples.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Material</th>
<th>KS</th>
<th>LF</th>
<th>SW</th>
<th>AD</th>
<th>CVM</th>
<th>Pearson</th>
<th>SF</th>
</tr>
</thead>
<tbody>
<tr>
<td>original data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nano</td>
<td>copper</td>
<td>.033</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>.012</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Nano</td>
<td>bronze 1</td>
<td>.108</td>
<td>.001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Nano</td>
<td>bronze 2</td>
<td>.046</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Micro</td>
<td>steel</td>
<td>.031</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Micro</td>
<td>bronze</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

data with removed waviness

| Nano      | copper     | .155  | .003   | < .001 | < .001 | < .001 | .003    | < .001 |
| Nano      | bronze 1   | .542  | .122   | .001   | .004   | .003   | .437    | .001  |
| Nano      | bronze 2   | .049  | < .001 | < .001 | < .001 | < .001 | < .001  | < .001 |
| Micro     | steel      | .002  | < .001 | < .001 | < .001 | < .001 | < .001  | < .001 |
| Micro     | bronze     | < .001 | < .001 | < .001 | < .001 | < .001 | < .001  | < .001 |

the difference, one can touch a pin by a soft part of a human body (e.g. by hand). Clearly, the effect will be different depending on the side the pin is approached, while the power spectral density of the pin is the same for both sides.

The main purpose of the present paper was not to present specific experimental data on surface roughness but rather to discuss the current models of statistical approaches to description of rough surfaces. It was noted that the most classic models of random surfaces are based on assumption of normality of the asperity heights or similar assumptions that involve Gaussian distributions. It has been suggested to check this assumption by application of various modern tests of normality of data that include the Kolmogorov-Smirnov, Lilliefors, Shapiro-Wilk, Pearson, Anderson-Darling, Cramer-von Mises and Shapiro-Francia tests. Of course, there are many other tests of normality (see a discussion by Thode 2002). However, the tests used are the most popular ones. The typical experimental data obtained for rough engineering surfaces prepared by grinding have been studied at micro and nano-scales. The heights of the micro-asperities were determined by a profilometre (stylus), while the data for nano/atomic scale was obtained by AFM (Atomic Force Microscopy). Our statistical analysis is directly related to scales of applicability of different physical mechanisms of interactions between surfaces. Because it was found that the height distribution of the surfaces under investigation were not Gaussian at both micro- and nano-scales, the above classic statistical approaches are not applicable. Of course, it is difficult to expect that the linear distribution used by Zhuravlev (1940) or the exponential distribution of the asperity summits used by Greenwood and Williamson (1966) for solving some illustrative examples, will well describe the surface statistics. However, one could try to test other models, e.g. one could expect that the Markov surfaces suggested by Whitehouse (2001b) are more appropriate for rough surface description. We believe that even if the height distribution of a surface is normal or another distribution of heights describes very well the roughness, this is of rather restricted use for solving appropriate contact problems. There is a need in development of new statistical
models of rough surfaces that can be incorporated in existing numerical solvers.

**Acknowledgements**

The results have been presented at the 6th Vienna International Conference NANO-TECHNOLOGY – VIENNANO '15 (Borodich et al. 2015) and the first CARBTRIB International Workshop on ”Nano-phenomena and Functionality of modern Carbon-Based Tribo-Coatings”, Cardiff, 2016. The authors are grateful to the Leverhulme Trust for financial support of their collaboration within the framework of the CARBTRIB International Network.

The studies of O. Savencu at Cardiff university were financially supported by the "Austrian COMET-Program" within the framework of the COMET K2 Excellence Centre of Tribology (project XTribology, no. 824187 and no. 849109) via the Austrian Research Promotion Agency (FFG) and the provinces of Niederösterreich, Vorarlberg and Wien. The support is very much appreciated.

Thanks are due to our colleagues Dr. Emmanuel Brousseau and Dr. Ingram Weeks (Cardiff) for their help in getting the roughness experimental data using an AFM and a profilometre respectively.

**References**


