PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Discrete Properties of Random Surfaces

D. J. Whitehouse and M. J. Phillips

Phil. Trans. R. Soc. Lond. A 1978 **290**, 267-298
doi: 10.1098/rsta.1978.0084

DISCRETE PROPERTIES OF RANDOM SURFACES

By D. J. Whitehouse†§ and M. J. Phillips‡

† *Research Department, Rank Taylor Hobson Ltd, Leicester,*

‡ *Department of Mathematics, University of Leicester*

(*Communicated by D. Tabor, F.R.S. – Received* 19 *January* 1978 – *Revised* 26 *April* 1978)

CONTENTS

The problem of measuring surface parameters of tribological importance such as peak curvature has been considerably simplified. Using discrete random process analysis tribological parameters of a wide range of surfaces can now be expressed and measured in terms of just two points on the measured autocorrelation function and the r.m.s. value of the surface. In addition, the contribution each 'scale of size' of asperity makes to an overall parameter can be assessed quantitatively. Many new expressions relating tribological parameters to the autocorrelation function have been derived using a limiting procedure which produces results entirely consistent with equivalent continuous theory. Using this theory it is now possible to predict 'gap' parameters between two surfaces in contact in terms of simple additive parameters of each surface.

Finally a new statistical model of the surface has been developed which encompasses many types of engineering surface.

§ Present address: Department of Engineering, University of Warwick.

1. Introduction

It is now a well established fact that surfaces are important in tribology and in other functional situations (Bowden & Tabor 1954). The realization of this has meant that more and more researchers are measuring surfaces and, furthermore, they are looking at many new parameters of the surface geometry. This is an inevitable development brought about by the growing need to understand more fully the complex interaction between surfaces during experiments.

One of the most important consequences of this interest has been the dramatic increase in the number of investigators using digital methods to investigate surfaces (Greenwood & Williamson 1977). This in itself is good for the subject because it enables the full power and versatility of digital techniques to be used in surface assessment. It has, however, a drawback, because many of the complexities of signal analysis are now brought into the open. When using conventional analogue equipment such as filters, recorders and similar devices, problems associated with frequency responses, noise levels etc. are hidden from the operator. They are still present but plausible answers will continue to emerge because the designer of the equipment has usually set out to match up the various units making up the system. Unfortunately there is no intrinsic matching up procedure in digital methods unless a great deal of digital expertise is available and the instrument understood fully. The researcher usually meets the digital problems of sampling, quantization and numerical analysis head-on and all too often he is not even fully aware of them. With the ever decreasing cost of computing the use, and misuse, of computers in this field will obviously grow. Today it is unsafe for one researcher to compare results of surface parameters obtained by digital methods with another even for the same surface!

There are a number of reasons why it is so difficult for a researcher to make good use of digital data from a surface. One of these is concerned with the analysis. When building up a theory from which to predict the functional behaviour of his experiment he often finds difficulty in finding the relevant surface parameters in the literature and of knowing how what is in the literature relates to his measurements. This is mainly due to the fact that the complex analysis of waveforms has been carried out for the most part by communication engineers (Rice 1944). Naturally they have placed most emphasis on the temporal properties of waveforms such as time series analysis and not on some of the more exotic properties of maxima and minima such as curvatures which are so important in tribology. Neither has it been possible to identify the properties of surfaces which can be ascribed to the different scales of size of asperity making up the surface – another issue of vital importance in tribology and contact phenomena (Archard 1957).

Recently there have been some attempts to rectify this situation. In the first instance some useful surface parameters have been evaluated using statistical methods. Greenwood & Williamson (1966) for example placed a lot of emphasis on the distribution of the peak heights in their models for plastic and elastic contact. Even more recently a major breakthrough has been the introduction of random process analysis. Peklenik (1967–68) used the autocorrelation function as a basis for classifying surfaces whereas Whitehouse & Archard (1970) and Nayak (1971, 1973) expressed tribological parameters in terms of the autocorrelation function, the Nayak work being based on the pioneering two-dimensional work of Longuet-Higgins (1957). Unfortunately use of random process analysis by itself only gives part of the story; it only gives the theoretical relationships between surface parameters and the autocorrelation function it does not show the researcher how the *measured* parameters relate to the *measured* autocorrelation function. Whitehouse & Archard (1970) were the first to tackle the overall problem by using discrete random process analysis in which the profile was considered to be made up of digital or discrete ordinates. They then developed expressions for some of the important parameters such as peak curvature in terms of these discrete ordinates and the correlation function. For instance in their analysis a peak was defined only when the central ordinate of a set of three was the highest. Intrinsically this method of approach is attractive because it gives the foundation for understanding digital results obtained from surfaces; it also provides a very simple way of exploring how the scale of size of asperity changes its properties (this is done by merely changing the interval between samples). Neither of these problems had been investigated previously.

Although this early work by Whitehouse & Archard (1970) was a useful starting point in bridging the gap between theory and practice it was somewhat restricted. Only surfaces having a normal (gaussian) distribution of ordinate heights and an exponential autocorrelation were considered. Although the former constraint is not serious the latter can be; it confines the analysis to only one type of surface and can cause some theoretical difficulties because of the properties of exponential functions. However, ease of manipulation makes this model attractive to use and later Whitehouse (1978) investigated more of the problems of the digital analysis of surfaces. He was able to show that of the three digital problem areas, the distance between data points (here called ordinates), the quantization in the measurement of the height of the ordinate, and the numerical model, it was the distance or interval between ordinates which posed most of the problems. He was also able to show that the three-point analysis method of determining peaks and peak properties was sufficiently accurate for most practical purposes. What is badly needed it to extend this early work and to generalize it so that many different types of surface can be included in order that the nature and extent of the problem of describing and measuring surfaces in a discrete form can be carried out properly. One of the principal aims of this paper is to help satisfy this need and if possible simplify the measurement of such difficult parameters as the variance in peak curvature and other similar parameters.

In this paper a nominally gaussian (normal) distribution of ordinate heights taken from a profile of the surface will be assumed: all practical examples used in the paper adhere to this condition. Because of the comments above only the effects of sampling interval will be considered and the numerical model will be restricted to the three-point model. In §2 general expressions are developed relating tribological parameters of a surface profile to two points in the autocorrelation function. Many significant new parameters are evaluated in this way using discrete random process analysis.

The theoretical validity of these expressions is proved in the first part of §3 by taking the sample interval to zero and showing that the formulae derived converge to those of continuous random process theory. Armed with this knowledge continuous expressions are derived, using this limiting procedure, for a number of potentially important parameters not hitherto available in the literature. An example of this is the correlation coefficient between peak height and curvature. Others are the standard deviation of the peak height and curvature distributions. The remainder of §3 is devoted to finding a statistical model of a surface which is suitably versatile to be able to be used for many types of surface, and yet at the same time remain theoretically sound. This model was needed to explore the relationship between sampling interval and the tribological parameters reported later in the paper.

The first part of §4 is devoted to showing the results of some tests on real surfaces in order to demonstrate the practical validity of the relationships developed in §2. This was carried out for a number of different types of surface. During this subsection the practical procedure for using and interpreting the formulae derived in the early part of the paper is given. Section 4 concludes with the results obtained by using the statistical model developed in §3 to explore the variation in the surface parameters as the sampling interval is changed. Again the interpretation of results is explained.

Section 5 consists of a discussion of the paper and its implications.

2. Analysis of discrete parameters of a generalized random surface

In order to investigate the three-point analysis of a random surface, it is necessary to apply suitable constraints to the joint distribution of three equally spaced ordinates y_{-1}, y_0 and y_1, taken from a profile of the surface. An ordinate is a discrete measurement of the profile. It is assumed, without loss of generality, that the ordinates have a probability distribution with zero mean and unit variance. The joint probability distribution of the three ordinates will be assumed to be a trivariate normal (gaussian) distribution. The variance–covariance matrix V will be given by

$$V = \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{pmatrix}, \tag{2.1}$$

where ρ_1 is the correlation coefficient between adjacent ordinates spaced a distance h apart and ρ_2 is the correlation coefficient between ordinates spaced a distance $2h$ apart. It is necessary for this matrix V to be positive definite and this gives the inequality

$$2\rho_1^2 - 1 < \rho_2 < 1.$$

A multivariate normal distribution of a vector $\boldsymbol{Y}$ will be denoted by $\boldsymbol{Y} \sim N[\boldsymbol{\mu}, \boldsymbol{V}]$ where $\boldsymbol{\mu}$ is the vector of means and V is the variance–covariance matrix.

2.1. *Distributions derived from linear combinations of ordinates*

Some useful characteristics of a surface profile can be expressed simply in linear combinations of the ordinates. Because of the linearity normal distributions will be obtained. Two such characteristics are the slope and curvature of the profile, which will be denoted by m and c respectively. The curvature is taken as the second differential of the surface profile, as the surface slopes are generally small. If

$$s_{-1} = y_0 - y_{-1} \tag{2.2}$$

and

$$s_1 = y_0 - y_1 \tag{2.3}$$

then the slope is given by

$$m = (s_{-1} - s_1)/2h \tag{2.4}$$

and the curvature is given by

$$c = (s_{-1} + s_1)/h^2. \tag{2.5}$$

The joint distribution of S_{-1}, S_1 and Y_0 is

$$(S_{-1}, S_1, Y_0)' \sim N\left[\begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix} 2-2\rho_1 & 1-2\rho_1+\rho_2 & 1-\rho_1 \\ 1-2\rho_1+\rho_2 & 2-2\rho_1 & 1-\rho_1 \\ 1-\rho_1 & 1-\rho_1 & 1 \end{pmatrix}\right]. \tag{2.6}$$

Hence the joint distribution of the slope M and the central ordinate Y_0 is

$$(M, Y_0)' \sim N\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix} (1-\rho_2)/2h^2 & 0 \\ 0 & 1 \end{pmatrix}\right], \tag{2.7}$$

and of the curvature C and the central ordinate Y_0 is

$$(C, Y_0)' \sim N\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}(6-8\rho_1+2\rho_2)/h^4 & 2(1-\rho_1)/h^2\\ 2(1-\rho_1)/h^2 & 1\end{pmatrix}\right]. \tag{2.8}$$

The marginal distributions are normal. So

$$Y_0 \sim N[0, 1], \tag{2.9}$$

$$M \sim N[0, (1-\rho_2)/2h^2] \tag{2.10}$$

and

$$C \sim N[0, (6-8\rho_1+2\rho_2)/h^4]. \tag{2.11}$$

The conditional distributions are also normal, though M and Y_0 are independent random variables. The distribution of the profile height Y_0 for a given curvature C $(=c)$ is

$$(Y_0 | C=c) \sim N\left[\frac{(1-\rho_1)\,h^2c}{(3-4\rho_1+\rho_2)}, \frac{1-2\rho_1^2+\rho_2}{3-4\rho_1+\rho_2}\right]. \tag{2.12}$$

A feature of tribological importance to emerge is that the higher parts of the profile have bigger curvatures c, as defined by (2.5). This can be seen from the correlation coefficient between C and Y_0, given by

$$\operatorname{corr}(C, Y_0) = \frac{2(1-\rho_1)}{(6-8\rho_1+2\rho_2)^{\frac{1}{2}}}, \tag{2.13}$$

which is always positive. This correlation coefficient is $(\frac{2}{3})^{\frac{1}{2}}$ if $\rho_1 = \rho_2 = 0$, when the ordinates are independently distributed. The distribution of the modulus of M is a folded normal distribution (see Johnson & Kotz 1970), with a mean given by

$$E(|M|) = \frac{1}{h}\left(\frac{1-\rho_2}{\pi}\right)^{\frac{1}{2}}. \tag{2.14}$$

When $\rho_2 = \rho_1^2$, the case when the correlations are exponential, (2.10) and (2.11) reduce to the equations derived by Whitehouse & Archard (1970).

2.2. *Distributions of peaks* (*using truncation*)

A peak is defined, with the use of three ordinates, when the central ordinate y_0 is higher than the other two. The condition is therefore that

$$y_{-1} < y_0 > y_1$$

or

$$s_{-1} > 0 \quad \text{and} \quad s_1 > 0.$$

This event $\{S_{-1} > 0, S_1 > 0\}$, which implies that a peak occurs, will be denoted by T. Hence the distribution for peaks will be obtained by constraining the joint probability density function $f(s_{-1}, s_1, y_0)$ subject to these conditions (i.e. truncating the random variables S_{-1} and S_1 below, at zero).

So the probability density function of a peak at height y_0 is given by

$$\begin{aligned} f(y_0|T) &= \frac{\operatorname{prob}(S_{-1} > 0,\ S_1 > 0 | Y_0 = y_0) f(y_0)}{\operatorname{prob}(S_{-1} > 0,\ S_1 > 0)} \\ &= \frac{\operatorname{prob}(T|Y_0 = y_0) f(y_0)}{\operatorname{prob}(T)} \\ &= \frac{\Phi\left(y_0\left(\frac{1-\rho_1}{1+\rho_2}\right)^{\frac{1}{2}}, y_0\left(\frac{1-\rho_1}{1+\rho_1}\right)^{\frac{1}{2}}; \frac{\rho_2-\rho_1^2}{1-\rho_1^2}\right)\phi(y_0)}{\Phi\left(0,\ 0; \frac{1-2\rho_1+\rho_2}{2-2\rho_1}\right)}, \end{aligned} \tag{2.15}$$

where
$$\phi(y_0) = \frac{\exp\left(-\tfrac{1}{2}y_0^2\right)}{(2\pi)^{\frac{1}{2}}} \tag{2.16}$$

and
$$\Phi(z, z; \rho) = \int_{-\infty}^{z}\int_{-\infty}^{z} \frac{\exp\left(-\tfrac{1}{2}(s_{-1}^2 + s_1^2 - 2\rho s_{-1}s_1)/(1-\rho^2)\right) \mathrm{d}s_{-1}\,\mathrm{d}s_1}{2\pi(1-\rho^2)^{\frac{1}{2}}}. \tag{2.17}$$

This result follows as $f(y_0|T)$ is the probability density function of the non-truncated variable of a trivariate normal distribution, when the other two variables are truncated below, at zero.

Of particular interest in tribology is the way in which the density of peaks changes from surface to surface, the way in which the mean peak height depends on the correlations between ordinates, and the relationship between peak height and the curvature of the peak. The mean peak height is obtained from the probability density function given by (2.15), and is given by

$$E(Y_0|T) = \frac{[(1-\rho_1)/\pi]^{\frac{1}{2}}.}{2\Phi\left(0,\ 0;\ \dfrac{1-2\rho_1+\rho_2}{2-2\rho_2}\right)}. \tag{2.18}$$

The variance of peak height is given by

$$\sigma^{*2} = \operatorname{var}(Y_0|T)$$
$$= 1 + \frac{\dfrac{(1-\rho_1)}{2\pi}\left(\dfrac{1-\rho_2}{3-4\rho_1+\rho_2}\right)^{\frac{1}{2}}}{\Phi\left(0,\ 0;\ \dfrac{1-2\rho_1+\rho_2}{2-2\rho_1}\right)} - \frac{\dfrac{(1-\rho_1)}{4\pi}}{\left[\Phi\left(0,\ 0;\ \dfrac{1-2\rho_1+\rho_2}{2-2\rho_1}\right)\right]^2}. \tag{2.19}$$

These results can be obtained from results for the moments of the truncated trivariate normal distribution given by Tallis (1961) and Finney (1962). Both of these moments have been used extensively in tribological situations. Mitchell & Rowe (1968) used them to predict the sealing performance of surfaces.

The density of peaks counted on a profile when digital techniques are used is better expressed in terms of the probability that an ordinate is a peak. In terms of importance the peak density can hardly be over emphasized. In the steel industry the peak count is one of the central parameters of interest. If the probability that an ordinate is a peak is N then

$$N = \operatorname{prob}(T)$$
$$= \Phi\left(0,\ 0;\ \frac{1-2\rho_1+\rho_2}{2-2\rho_1}\right)$$
$$= \frac{1}{\pi}\arctan\left[\left(\frac{3-4\rho_1+\rho_2}{1-\rho_2}\right)^{\frac{1}{2}}\right]. \tag{2.20}$$

When $\rho_2 = \rho_1^2$ then the results of (2.15), (2.18), (2.19) and (2.20) reduce to those obtained by Whitehouse & Archard (1970).

For the case of independent ordinates, when $\rho_2 = \rho_1 = 0$, then (2.18) reduces to

$$E(Y_0|T) = \tfrac{3}{2}\pi^{-\frac{1}{2}}, \tag{2.21}$$

and (2.19) reduces to
$$\sigma^2 = \operatorname{var}(Y_0|T) = 1 + \frac{3^{\frac{1}{2}}}{2\pi} - \frac{9}{4\pi}. \tag{2.22}$$

The importance of (2.21) and (2.22) is in the fact that they enable the limiting values to be obtained as the sampling interval h increases.

Another important property of peaks is their curvature, especially when contact properties are being considered. Whether or not surface peaks deform elastically or plastically is determined by the plasticity index, according to Greenwood & Williamson (1966). The plasticity index ψ is given by

$$\psi = \frac{E}{H}\left(\frac{\overset{*}{\sigma}}{R}\right)^{\frac{1}{2}}, \tag{2.23}$$

where E is the elastic modulus of the surface, H is the hardness, $\overset{*}{\sigma}$ is the r.m.s. (standard deviation) of the peak height distribution and R is the radius of curvature of the peaks. Obviously in all wear situations R is a critical factor. The distribution of the curvature of a peak has a probability density function given by

$$f(c|T) = \frac{\text{prob}\,(0 < S_1 < ch^2|C = c)\,f(c)}{\text{prob}\,(T)}, \tag{2.24}$$

where $f(c)$ is the probability density function of the distribution of the profile curvature given by (2.11). Hence

$$f(c|T) = \frac{\left[\Phi\left(\frac{ch^2}{[2(1-\rho_2)]^{\frac{1}{2}}}\right)-\Phi\left(\frac{-ch^2}{[2(1-\rho_2)]^{\frac{1}{2}}}\right)\right]h^2\,\phi\left(\frac{ch^2}{(6-8\rho_1+2\rho_2)^{\frac{1}{2}}}\right)}{N(6-8\rho_1+2\rho_2)^{\frac{1}{2}}}, \text{ for } c>0,$$

$$= 0, \quad \text{for} \quad c \leqslant 0, \tag{2.25}$$

where

$$\Phi(z) = \int_{-\infty}^{z} \phi(y)\,\mathrm{d}y. \tag{2.26}$$

So the probability density function of this distribution is obtained from the convolution of the joint probability density function of a bivariate normal distribution when both variables are truncated below at zero. The mean and variance of the peak curvature can be obtained from the results for the moments obtained for the truncated bivariate normal distribution by Weiler (1959) and Rosenbaum (1961). So

$$E(C|T) = \frac{3-4\rho_1+\rho_2}{2Nh^2[\pi(1-\rho_1)]^{\frac{1}{2}}} \tag{2.27}$$

and

$$\text{var}\,(C|T) = \frac{(3-4\rho_1+\rho_2)}{4h^4\pi(1-\rho_1)}\left[8\pi(1-\rho_1)+\frac{2[(3-4\rho_1+\rho_2)\,(1-\rho_2)]^{\frac{1}{2}}}{N}-\frac{(3-4\rho_1+\rho_2)}{N^2}\right]. \tag{2.28}$$

From a knowledge of the mean and variance of the peak curvature, the proportion of peaks on the surface which will elastically or plastically deform could be estimated. This is an important tribological feature which has so far been neglected. When $\rho_2 = \rho_1^2$ then the mean peak curvature given by (2.27) reduces to the result obtained by Whitehouse & Archard (1970).

It is possible to obtain similar results for the height and curvature of valleys, which can also be important in frictional studies involving the passage of fibres over rollers, or similar applications. An interesting result is that the conditional distribution of peak height for a given curvature is the same as for the conditional distribution of the profile height for the same given curvature, as given by (2.12). This follows from a result given by Lawley (1943) and is because Y_0 is not truncated. Hence

$$(Y_0|C = c, T) \sim N\left[\frac{(1-\rho_1)\,h^2c}{(3-4\rho_1+\rho_2)}, \frac{1-2\rho_1^2+\rho_2}{3-4\rho_1+\rho_2}\right]. \tag{2.29}$$

Combining (2.24) and (2.25) with (2.29) we obtain the joint probability density function of peak height and curvature as

$$f(c, y_0 | T) = \frac{\text{prob}\,(0 < S_1 < ch^2 | C = c) f(c, y_0)}{\text{prob}\,(T)}$$

$$= \frac{1}{N}\left[\Phi\left(\frac{ch^2}{[2(1-\rho_2)]^{\frac{1}{2}}}\right) - \Phi\left(\frac{-ch^2}{[2(1-\rho_2)]^{\frac{1}{2}}}\right)\right] f(c, y_0), \quad \text{for} \quad c > 0,$$

$$= 0, \quad \text{for} \quad c \leqslant 0, \tag{2.30}$$

where $f(c, y_0)$ is the joint probability density function of the profile height and curvature. So as before with (2.25) the peak probability density function is obtained by multiplying the profile probability density function by prob $(0 < S_1 < \text{ch}^2 | C = c)$ and then normalizing. This connection between peaks and profile is a consequence of the definition of peaks in terms of a triplet of ordinates. Whitehouse (1971, 1978) has shown that a more refined definition of a peak is rarely if ever needed in practice. If $\rho_2 = \rho_1^2$ then (2.30) reduces to a result which corresponds to that given by Whitehouse & Archard (1970).

The correlation coefficient of peak height and curvature can be obtained using (2.30), in terms of the variance of peak height. Thus

$$\text{corr}\,(C, Y_0 | T) = \left[1 - \frac{(1 - 2\rho_1^2 + \rho_2)}{(3 - 4\rho_1 + \rho_2)\,\text{var}\,(Y_0 | T)}\right]^{\frac{1}{2}}, \tag{2.31}$$

using (2.22). This correlation coefficient is always positive so that the higher the peaks the larger is their curvature (which makes physical sense). From a result observed by Regier & Hamdan (1971), it can be shown that

$$0 \leqslant \text{corr}\,(C, Y_0 | T) \leqslant \text{corr}\,(C, Y_0) \leqslant 1.$$

Formulae in §2 can in some instances be made simpler by use of the structure function $\bar{S}(\tau)$ rather than the auto-correlation $\rho(\tau)$ where

$$\bar{S}(\tau) = E(y(x) - y(x+\tau))^2 = \text{var}\,(y(x) - y(x+\tau)) \tag{2.32}$$

thus

$$\bar{S}(\tau) = 2\sigma^2(1 - \rho(\tau)). \tag{2.33}$$

It can be seen in §2 that many expressions involve the differences $1 - \rho_1$, $1 - \rho_2$, etc. so that the simplifications made possible by equation (2.33) are obvious. In effect peak and slope parameters are generally better specified in terms of structure functions than of correlation functions because peaks and slopes, in discrete analysis, are defined in terms of differences. Thus

$$\text{corr}\,(C, Y_0) = \frac{\bar{S}_1}{(4\bar{S}_1 - \bar{S}_2)^{\frac{1}{2}}} \tag{2.34}$$

$$N = \frac{1}{\pi} \arctan\left(\frac{4\bar{S}_1 - \bar{S}_2}{\bar{S}_2}\right)^{\frac{1}{2}} \tag{2.35}$$

$$E(|M|) = \frac{1}{h}\left(\frac{\bar{S}_2}{2\pi}\right)^{\frac{1}{2}} \tag{2.36}$$

for example where $\bar{S}_1 = 2(1 - \rho_1)$ and $\bar{S}_2 = 2(1 - \rho_2)$, σ^2 being taken as unity.

3. Statistical models of random surfaces – analogue and digital

3.1. *Limiting behaviour of the correlation coefficients between ordinates*

So far distributions have been obtained for those surfaces characteristics deemed to be of most interest in tribological situations. These expressions have been derived in terms of ordinates of the profile; little has been said about the correlation coefficients between ordinates on which the formulae depend. These are ρ_1, the correlation between ordinates a distance h apart, and ρ_2, the correlation between ordinates $2h$ apart. The only constraints on ρ_1 and ρ_2 are those obtained by making the matrix in (2.1) positive definite. So ρ_1 and ρ_2 will vary as h varies, depending on the shape of the autocorrelation function of the surface. As h approaches zero

$$\lim_{h\to 0} \rho_1(h) = \lim_{h\to 0} \rho_2(h) = 1, \tag{3.1}$$

and as h approaches infinity
$$\lim_{h\to \infty} \rho_1(h) = \lim_{h\to \infty} \rho_2(h) = 0. \tag{3.2}$$

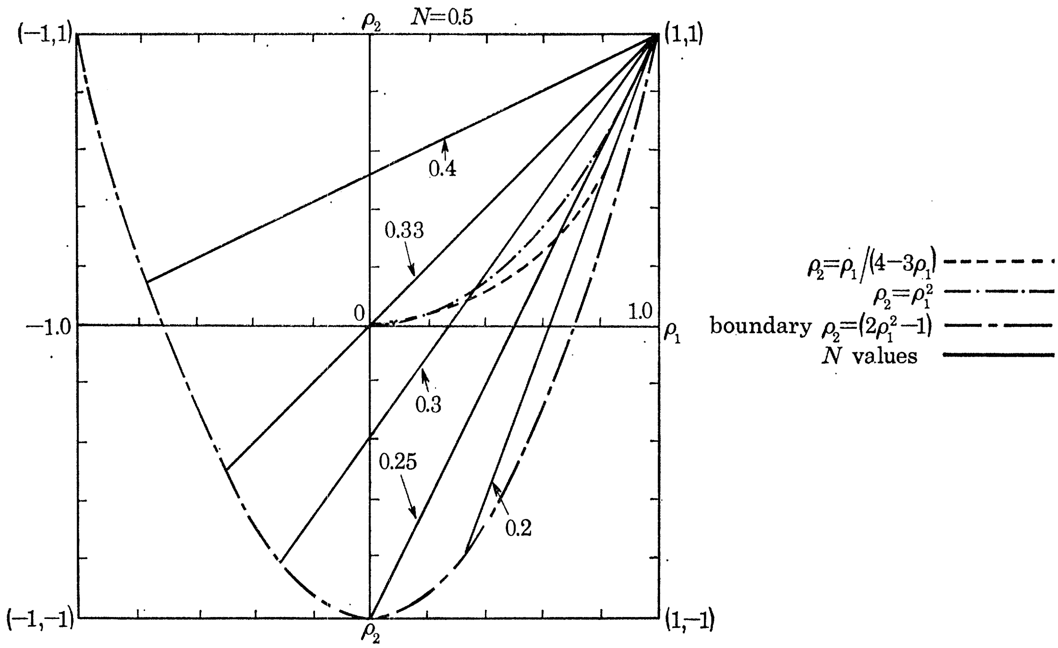

Figure 1. The contours of peak density (the probability that an ordinate is a peak) in the correlation domain showing two correlation functions.

If ρ_2 is plotted against ρ_1 as h varies in the ρ_1, ρ_2 domain (see figure 1), then the curve will start at (1, 1) for $h = 0$ and end at (0, 0) for $h = \infty$. The shape of the curve will depend on the autocorrelation function. The boundary imposed by the positive definite matrix are given by a straight line between the points (−1, 1) and (1, 1), and by a parabola through the points (−1, 1), (0, −1) and (1, 1). The expressions derived in §2 are not valid if the autocorrelation function chosen gives values outside this boundary. However, these need not be considered for other violations also occur. For instance, in a second order random surface the only solution which lies on the boundary is the one in which the autocorrelation function is cosinusoidal and not random. In the next sub-section some parameters of the distributions will be investigated as h changes for different autocorrelation functions. This will illustrate how much variation could be expected when measurements are taken with different spacings between the ordinates.

3.2. *Behaviour of discrete parameters in the correlation domain*

The probability that an ordinate is a peak is given by N in (2.20). Rearrangement of this shows that the contours for N are a family of straight lines all starting from the point (1, 1) and having slopes between 0 and $\frac{1}{2}$. The particular autocorrelation function chosen for the statistical model of the surface determines which values of N would be obtained. For example, for the exponential autocorrelation function, which gives the curve $\rho_2 = \rho_1^2$, the values of N will vary between $\frac{1}{4}$ and $\frac{1}{3}$ as h varies. This was discussed in some detail by Whitehouse & Archard (1970). Other autocorrelation functions give different results. For instance if the correlations satisfy

$$\rho_2 = \rho_1/(4-3\rho_1) \tag{3.3}$$

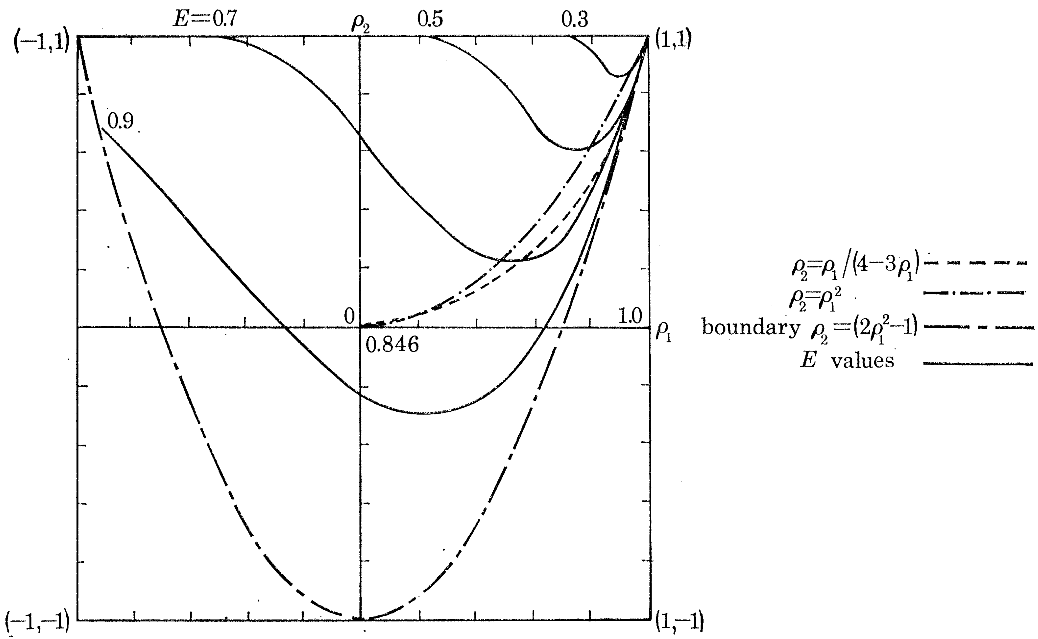

FIGURE 2. The contours of mean peak height, showing two correlation functions and the limiting value at zero correlation.

then the values of N vary between 0 and $\frac{1}{3}$ as h varies (see figure 1). For the mean peak height, which is given by (2.18), the contours are not straight lines (figure 2), though they still pass through the point (1, 1). The values of the mean peak height vary between 0 and $(\frac{1}{2}\pi)^{\frac{1}{2}}$. For the exponential autocorrelation function the curve $\rho_2 = \rho_1^2$ will cross the contours from 0 to $3/(2\pi^{\frac{1}{2}})$, while for the autocorrelation function with the curve given by (3.3) the contours between $\frac{1}{2}(\frac{1}{3}\pi)^{\frac{1}{2}}$ and $3/(2\pi^{\frac{1}{2}})$ will be crossed.

These two examples of possible autocorrelation functions suffice to illustrate the fact that the possible variations in measured parameters on one type of surface may be larger than the differences between different types of surface. The variation will depend on how the parameters vary as h, the sampling interval, varies. Also it is possible to get very different behaviour between surfaces with different autocorrelation functions as h varies from 0 to ∞. This point will be emphasized when practical results are considered in §4.1. It now remains to consider how the discrete approach ties in with the continuous analysis found in communication theory and what

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

surface models need to be considered in order to meet the range of real surfaces met with in practice. A main objective of this paper is to use this model to quantify the real variations in the surface parameters likely to be met with in practice.

3.3. *Limiting behaviour of distributions*

In this paper peaks have been defined by considering a triple set of ordinates, equally spaced at a distance h. This is a discrete approximation to a peak on the real surface which forms part of a continuous boundary. The continuous definition of a peak is given by that point on the boundary (or profile) where the slope changes from a positive to a negative value. It is necessary for the discrete results, obtained using three ordinates, to converge to the results obtained for the continuous profile as h approaches zero.

Rice (1944), Longuet-Higgins (1957) and Bendat (1958) have evaluated the peak distribution on a continuous waveform. They used the distribution of the amplitude of the waveform and its first and second derivatives. To determine the distribution of peak height for a normal (gaussian) surface they used a statistical model of the waveform whose autocorrelation function $\rho(\tau)$ had the first and third derivatives at the origin zero. So the autocorrelation function can be expressed by using Taylor's theorem in terms of only the second and fourth derivatives D_2 and D_4, where

$$D_r = \left.\frac{\mathrm{d}^r\rho(h)}{\mathrm{d}h^r}\right|_{h=0}, \tag{3.4}$$

for r a positive integer. Rice (1944) obtained the probability density function of the distribution of peak height Y, in terms of D_2 and D_4, as

$$f(y) = \left(\frac{D_4 - D_2^2}{D_4}\right)^{\frac{1}{2}} (2\pi)^{\frac{1}{2}}[\phi(w) + w\Phi(w)]\,\phi(y), \tag{3.5}$$

where
$$w = \frac{(-D_2)\,y}{(D_4 - D_2^2)^{\frac{1}{2}}}. \tag{3.6}$$

The mean and variance of the peak height was not evaluated because they are not very significant in communication theory, although they are in tribology. However, from (3.5)

$$E(Y) = (\tfrac{1}{2}\pi)^{\frac{1}{2}}(-D_2)/(D_4)^{\frac{1}{2}} \tag{3.7}$$

and
$$\operatorname{var}(Y) = 1 + (1 - \tfrac{1}{2}\pi)\,(-D_2)^2/D_4. \tag{3.8}$$

Rice (1944) and Bendat (1958), in addition to the probability density function of the peak distribution given by (3.5), obtained results for m_0, the mean number of peaks in unit distance, and n_0, the mean number of crossings of the mean line in unit distance, in terms of D_2 and D_4. These are

$$m_0 = \frac{1}{2\pi}\left(\frac{D_4}{-D_2}\right)^{\frac{1}{2}} \tag{3.9}$$

and
$$n_0 = (-D_2)^{\frac{1}{2}}/\pi. \tag{3.10}$$

Now the results obtained in §2 can be compared with those given above by first expressing the autocorrelation function in the Taylor's expansion and then investigating the behaviour as the sampling interval h tends to zero, when $\rho_1 = \rho(h)$ and $\rho_2 = \rho(2h)$, where

$$\rho(h) = 1 - (-D_2)\frac{h^2}{2!} + D_4\frac{h^4}{4!} + O(h^4). \tag{3.11}$$

We can use this limiting process method for both profile and peak results. Then the distributions of the profile obtained by the discrete analysis in §2.1 becomes for example

$$(M, Y_0)' \sim N\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}(-D_2) & 0\\ 0 & 1\end{pmatrix}\right] \tag{3.12}$$

from (2.7),

$$(C, Y_0)' \sim N\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}D_4 & -D_2\\ -D_2 & 1\end{pmatrix}\right] \tag{3.13}$$

from (2.8); and therefore

$$M \sim N[0, -D_2] \tag{3.14}$$

from (2.10), and

$$C \sim N[0, D_4] \tag{3.15}$$

from (2.11). The conditional distribution of Y_0 given $C\,(=c)$ is

$$(Y_0|C=c) \sim N\left[\left(\frac{-D_2}{D_4}\right)c,\ 1-\frac{(-D_2)^2}{D_4}\right] \tag{3.16}$$

from (2.12). This result has important tribological significance. The correlation coefficient between C and Y_0 is given by

$$\operatorname{corr}(C, Y_0) = -D_2/(D_4)^{\frac{1}{2}} = n_0/2m_0 \tag{3.17}$$

from (2.13), (3.9) and (3.10). The results obtained in §2.2 for the peak distributions can be compared as h tends to zero with the continuous results of Rice (1944) and Bendat (1958). It can be shown that the results of the probability density function, mean and variance of the peak height given by (2.15), (2.18) and (2.19) become the results given by (3.5), (3.7) and (3.8) obtained by Rice (1944). This clearly vindicates the use of sampled data methods to analyse surfaces.

Discrete formulae, when taken to the limit, can be usefully employed to give the analogue formulae for some useful tribological parameters not hitherto evaluated. For instance (3.7) and (3.8) can be expressed in terms of the limiting form of the correlation coefficient between C and Y_0, given by (3.17). Thus

$$E(Y_0|T) = (\tfrac{1}{2}\pi)^{\frac{1}{2}}\operatorname{corr}(C, Y_0) \tag{3.18}$$

and

$$\operatorname{var}(Y_0|T) = 1+(1-\tfrac{1}{2}\pi)\,[\operatorname{corr}(C, Y_0)]^2 \tag{3.19}$$

Also a new analogue formula for the probability density function of peak curvature can be found by taking the limit of (2.25), which gives

$$\begin{aligned} f(c|T) &= \frac{c}{D_4}\exp\left(\frac{-c^2}{2D_4}\right), \quad \text{for} \quad c>0,\\ &= 0, \quad \text{for} \quad c \leqslant 0. \end{aligned} \tag{3.20}$$

This is a Rayleigh distribution, not a gamma distribution as was originally suggested by Greenwood & Williamson (1966). The distribution of peak curvature depends only on D_4 and not on D_2. This is *not* an unexpected result as the curvature can be defined in terms of the second differential of the profile only, in the region of the peak, as the first differential is zero. Consequently D_4, derived from the autocorrelation of the second differential would be expected to play a major rôle. The distribution of the height of the envelope of a narrow band random surface also has a Rayleigh distribution.

From (3.20) it is possible to obtain

$$E(C|T) = (\tfrac{1}{2}\pi D_4)^{\frac{1}{2}} \tag{3.21}$$

and

$$\operatorname{var}(C|T) = D_4(2-\tfrac{1}{2}\pi). \tag{3.22}$$

These are also the limiting values of (2.27) and (2.28). Also the probability that a peak has curvature greater than c is given by

$$\operatorname{prob}(C > c|T) = \exp\left(\frac{-c^2}{2D_4}\right). \tag{3.23}$$

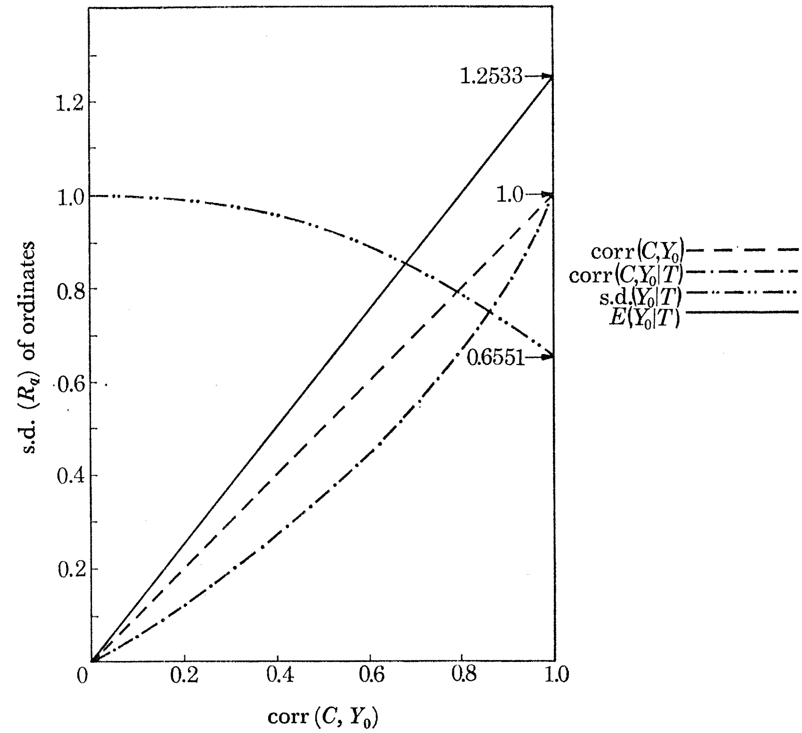

FIGURE 3. Limiting distribution of peak height, showing four peak parameters.

The conditional distribution of the *peak* height for a given curvature is the same as that for the conditional distribution of the *profile* height for a given curvature, and hence is a normal (gaussian) distribution as given by (3.16). Hence combining (3.16) and (3.20) gives the joint probability density function of the peak height and curvature as

$$f(c, y_0|T) = \frac{c}{[2\pi D_4(D_4-D_2^2)]^{\frac{1}{2}}} \exp\left[\frac{-(D_4 y_0^2 + 2D_2 y_0 c + c^2)}{2(D_4-D_2^2)}\right], \quad \text{for } c > 0,$$
$$= 0, \quad \text{for} \quad c \leqslant 0. \tag{3.24}$$

This is also the limiting value of (2.30). From (3.24), or by taking the limiting value of (2.31), the correlation coefficient between peak height and curvature is given by

$$\operatorname{corr}(C, Y_0|T) = \frac{(-D_2)}{(D_4)^{\frac{1}{2}}}\left[\frac{2-\frac{1}{2}\pi}{1+(1-\frac{1}{2}\pi)\,D_2^2/D_4}\right]^{\frac{1}{2}}. \tag{3.25}$$

Hence the mean and standard deviation of the peak height, given by (3.18) and (3.19), and the correlation coefficient between peak height and curvature, given by (3.25), are simple functions of the correlation coefficient between profile height and curvature, given by (3.17). The functions are given in figure 3.

So far only boundary conditions have been placed on the correlation coefficient values, in order to allow the discrete analysis to be valid. Now it is necessary to know what range of surface models need to be examined in order to cater for practical surfaces and, of equal importance, what additional constraints have to be imposed on such models in order to satisfy theoretical considerations. Both considerations have to be satisfied if analytical predictions of tribological situations are to be meaningful.

3.4. *Suitability of statistical models*

It is evident from the above discussion that ideally a statistical surface model should be such that the autocorrelation function satisfies the condition that D_2 and D_4 exist. All of the results obtained by either analogue or discrete methods (when taken to the limit) only truly allow peak height and curvature properties if these two derivatives exist. Such comments have been made previously by many authors, for instance Bendat (1958), Davenport & Root (1958) and Nayak (1971), the behaviour of the autocorrelation function at the origin being taken as the criterion of the suitability of the model. This is not, however, the only way of presenting this critical information. Perhaps a more practical way is to consider the restrictions on the power spectral density function $g(\omega)$ of the surface model. The power spectral density function gives the autocorrelation function $\rho(\tau)$ by means of the relationship

$$\rho(\tau) = \int_0^\infty g(\omega) \cos(\omega\tau)\, \mathrm{d}\omega. \tag{3.26}$$

The reciprocal Fourier transform of (3.26) gives $g(\omega)$ in terms of $\rho(\tau)$ as

$$g(\omega) = \frac{2}{\pi}\int_0^\infty \rho(\tau) \cos(\omega\tau)\, \mathrm{d}\omega, \tag{3.27}$$

which is a statement of the well known Weiner–Khintchine relationship (see, for example, Cox & Miller 1965).

For D_2 and D_4 to exist, D_1 and D_3 must also exist and be zero. So for r a positive integer, if D_{2r-1} exists then

$$D_{2r-1} = \lim_{\tau\to 0}\,(-1)^r \int_0^\infty \omega^{2r-1} g(\omega) \sin(\omega\tau)\, \mathrm{d}\omega = 0, \tag{3.28}$$

and if D_{2r} exists then

$$D_{2r} = (-1)^r \int_0^\infty \omega^{2r} g(\omega)\, \mathrm{d}\omega \tag{3.29}$$

is finite. So the infinite integrals in (3.28) and (3.29) must converge; if D_{2r} is to exist then $\omega^{2r} g(\omega)$ must converge to zero fast enough for the infinite integrals to converge for $r = 1$ and 2. It will often be more convenient to work with $g(\omega)$ as frequency limitations are often imposed by the instruments.

The theoretical constraints having been decided on, it is necessary to specify a model which will provide just enough versatility to cope with most real surfaces. Obviously not every surface likely to be used in engineering could be included, but certain broad types have to be taken into account. Three such types will now be given.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Type 1. Surfaces produced by conventional finishing processes, such as grinding, lapping polishing, electric discharge machining etc., in which there is almost a purely random waveform, which may well be limited in bandwidth.

Type 2. Surfaces which have two elements, one random and one deterministic, in which the random element is modulated (multiplied), by the deterministic. Surfaces like this are often produced in single tool cutting operations like diamond turning, in which the random element is introduced by a variable amplitude or phase of the feed mechanism. This type is often called a narrow band random surface indicating that the power spectral density function is centred around the fixed frequency.

Type 3. Surfaces in which the two elements are not modulated but are simply added together. This often happens in type 1 surfaces, where the machine has introduced a small amount of chatter, or in type 2 where the slideway has a systematic error of movement.

3.5. *Statistical models*

Ideally in order to be able to simulate practical surfaces the statistical model chosen should be able to take on all the possible types discussed above. The effect of digital methods could then be investigated for most surfaces. In the past a number of models have been used to simulate surfaces. These have been models of a random nature, because most conventional finishing processes such as grinding have this character. Because of this it is natural to think in terms of exponentials when building the probability model. These arise because there is evidence from Whitehouse (1971) that the *peak positions* of many surfaces obey a close approximation to a Poisson distribution law. So there is a uniform probability in any interval of a peak being at any position. It can also be shown that the general autocorrelation function can be written in term of conditional probabilities. Thus

$$\begin{aligned}\rho(\tau) &= E[y(x)\cdot y(x+\tau)]\\ &= E_\zeta[E(\tau|\zeta)]\\ &= \sum_{\text{all }\zeta} E(\tau|\zeta)f(\zeta),\end{aligned} \tag{3.30}$$

where ζ_1 is the event that $y(x)$ and $y(x+\tau)$ all lie within an impression left by one grit, ζ_2 is the event that they lie within the impression left by two grits, and so on, and $f(\zeta)$ is the corresponding probability function. If for example the grits are assumed to be square then (3.30) reduces to

$$\rho(\tau) = \exp(-\lambda_s|\tau|), \tag{3.31}$$

where λ_s is the density of grit edges. Various other forms for $\rho(\tau)$ result depending on the distribution of the individual grits, though they all involve the sum of exponential terms with low order polynomial coefficients. In view of this it is plausible to consider the use of a simple exponential function for the autocorrelation function of the surface model. This produces the simple relation between ρ_1 and ρ_2 of $\rho_2 = \rho_1^2$. The power spectral density function of the autocorrelation function

$$\rho(\tau) = \exp(-\alpha|\tau|) \tag{3.32}$$

is given by

$$\begin{aligned}g(\omega) &= \frac{2}{\pi}\int_0^\infty \exp(-\alpha|\tau|)\cos(\omega\tau)\,\mathrm{d}\tau\\ &= \frac{\alpha}{\pi(\alpha^2+\omega^2)}.\end{aligned} \tag{3.33}$$

So obviously, from (3.33), $\omega^{2r}g(\omega)$ does not tend to zero as ω tends to infinity for either $r = 1$ or 2.

Hence from (3.29) D_2 and D_4 do not exist, so that the mean number of peaks or crossings, given by (3.9) and (3.10), do not exist. Also the discrete results for peak behaviour *do not* converge to the analogue results.

The same situation arises when a similar autocorrelation function, the exponential cosine (Bendat 1958), is used. This also has a slowly decaying spectrum which causes the derivatives D_2 and D_4 not to exist. This is because if

$$\rho(\tau) = \exp(-|\tau|)\cos(2\pi\theta\tau) \tag{3.34}$$

then

$$g(\omega) = \frac{2}{\pi}\int_0^\infty \exp(-|\tau|)\cos(2\pi\theta\tau)\cos(\omega\tau)\,d\tau$$

$$= \frac{2}{\pi}\frac{1+(2\pi\theta)^2+\omega^2}{[1+(\omega+2\pi\theta)^2][1+(\omega-2\pi\theta)^2]}. \tag{3.35}$$

So the spectrum given by (3.35) belongs to a first order system, in the sense that at high frequencies it decays at 6 dB per octave, i.e. ω^{-2} in the power spectral density function, as does the pure exponential autocorrelation function. It is obvious from the constraints imposed by (3.28) and (3.29) that the decay of the spectrum must be *faster* than ω^{-4}, in order that the peak distributions may exist. It is the *degree* of the decay of the spectrum which determines the suitability of the model. Evidently to satisfy this condition imposed on the power spectral density function it must decay at a rate of ω^{-6}, or faster. Modification of the simple exponential autocorrelation function to comply with this condition gives the following sum of exponentials. Thus

$$\rho(\tau) = \frac{\gamma\beta(\gamma+\beta)\exp(-|\tau|)}{(\gamma-1)(\beta-1)(1+\gamma+\beta)} + \frac{\beta(1+\beta)\exp(-\gamma|\tau|)}{(\gamma-1)(\gamma-\beta)(1+\gamma+\beta)} + \frac{\gamma(\gamma+1)\exp(-\beta|\tau|)}{(\beta-1)(\beta-\gamma)(1+\gamma+\beta)}, \tag{3.36}$$

where $1 \leqslant \gamma \leqslant \beta$ so that

$$g(\omega) = \frac{2}{\pi}\frac{\gamma\beta(\gamma+\beta)(\gamma+1)(\beta+1)}{(1+\gamma+\beta)(1+\omega^2)(\gamma^2+\omega^2)(\beta^2+\omega^2)}. \tag{3.37}$$

This autocorrelation function is always positive, and for large γ and β it behaves like the simple exponential function, given by (3.32) with $\alpha = 1$. If γ and β are unity then

$$\rho(\tau) = (1+|\tau|+\tfrac{1}{3}\tau^2)\exp(-|\tau|) \tag{3.38}$$

so that

$$g(\omega) = \frac{2}{\pi}\frac{8}{3(1+\omega^2)^3}. \tag{3.39}$$

The existence of the four derivatives of the autocorrelation function at the origin is now guaranteed for the power spectral density function decays as ω^{-6}. Thus

$$D_2 = \frac{-\gamma\beta}{(1+\gamma+\beta)} \tag{3.40}$$

and

$$D_4 = \frac{\gamma\beta(\gamma+\beta+\gamma\beta)}{(1+\gamma+\beta)}. \tag{3.41}$$

Both D_1 and D_3 are zero and so all the discrete results obtained using the autocorrelation function given by (3.36) will in the limit be the same as the analogue results. This function corresponds to a surface with a model of type 1. A more versatile model is required to cover types 2 and 3.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Such a model may be obtained by modifying equation (3.34), the exponential cosine case, by adding a decaying periodic term and a variable one to give a new autocorrelation function. This model will cater for types 1 and 2 and some aspects of type 3. Thus

$$\rho(\tau) = (1-\overset{*}{C})\exp(-|\tau|)\cos 2\pi\theta\tau \frac{(1-\overset{*}{C})}{2\pi\theta_0}\exp(-\beta|\tau|)\sin 2\pi\theta_0\tau + \overset{*}{C}\cos 2\pi\gamma\tau, \qquad (3.42)$$

where
$$\beta = [\tfrac{1}{3}(1+4\pi^2\theta_0^2)-4\pi^2\theta^2]^{\frac{1}{2}}, \qquad (3.43)$$

where θ, θ_0 and γ are all measured by taking the exponent of the first term as unity. When $\overset{*}{C} > 0$ $\rho(\tau) \neq 0$ at $\tau \to \infty$ but it is a good approximation to many physical situations for small τ.

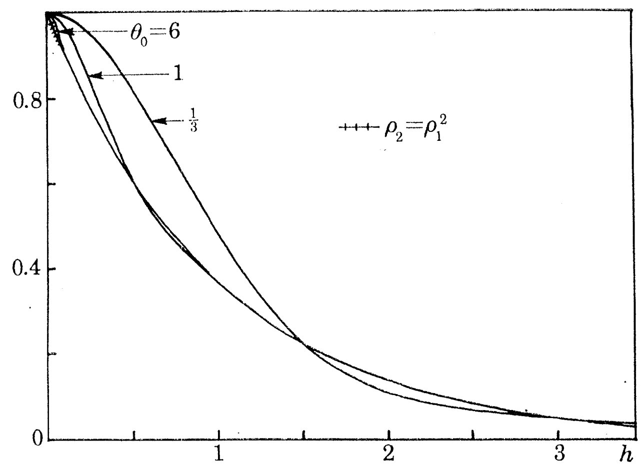

FIGURE 4. Shows how the shape of the model 1 autocorrelation function changes as θ_0 changes and how they differ from the exponential $\rho_2 = \rho_1^2$ at the origin.

When $\overset{*}{C} = 0$ this autocorrelation function will be referred to as model 1. This autocorrelation function is made up of three separate terms each having its use. The first term $(1-\overset{*}{C})\exp(-|\tau|)$ $\cos 2\pi\theta\tau$ takes account of types 1 and 2. The second term serves two purposes. Its main use is to make sure that the correlation function behaves in a theoretically acceptable way at the origin. The other use is to change the basic form of the autocorrelation function as is seen for different values of θ_0 in figure 4. The third expression in the equation takes account of surfaces of type 3. From the point of view of merely satisfying the practical types of surface equation (3.42) could be written as

$$\rho(\tau) = (1-\overset{*}{C})\exp(-|\tau|)\cos 2\pi\theta\tau + \overset{*}{C}\cos 2\pi\gamma\tau + \epsilon, \qquad (3.44)$$

where ϵ is the second term of the equation (3.42) with θ_0 chosen so that it only influences the behaviour of $\rho(\tau)$ for very small h. In fact θ_0 can always be chosen so that the effect of the second term is masked by the occurrence of any smoothing of the profile such as happens with the finite tip of the stylus in a stylus instrument. It can also be chosen so that it represents the instrumental distortion term present. Thus this autocorrelation function may provide very general coverage of surfaces found and measured in practice.

The power spectral density of the random terms of equation (3.42) when $\overset{*}{C} = 0$ is given by

$$g(\omega) = \frac{2}{\pi}\,\frac{[1+(2\pi\theta)^2][\beta^2+(2\pi\theta_0)^2][\beta^2+(2\pi\theta_0)^2-1-(2\pi\theta)^2]+\omega^2\{[\beta^2+(2\pi\theta_0)^2] + 4[\beta^2-(2\pi\theta)^2(2\pi\theta_0)^2]-[1+(2\pi\theta)^2]^2\}}{[1+(\omega+2\pi\theta)^2][1+(\omega-2\pi\theta)^2][\beta^2+(\omega+2\pi\theta_0)^2][\beta^2+(\omega-2\pi\theta_0)^2]}, \qquad (3.45)$$

which in the limit as $\omega \to \infty$ attenuates at a rate proportional to ω^6 and therefore satisfies the criterion laid down in the first part of this section.

Conceptually the first two expressions in equation (3.42) reflect the purely random components of the profile waveform whereas the additional cosinusoidal term reflects the deterministic. As such, for finishing processes the deterministic additive element is usually small when compared with the random element; it usually also has a shorter wavelength than the drop-off rate of the exponential term. An example of this type of surface is shown in the next section.

In equation (3.44) D_1 and D_3 are zero, the cosinusoidal additive term not contributing. Thus

$$\left.\begin{aligned} D_2 &= (1-\overset{*}{C})\,[1-2\beta-(2\pi\theta)^2]-4\pi^2\overset{*}{C}\gamma^2, \\ D_4 &= (1-\overset{*}{C})\,[1-6(2\pi\theta)^2+(2\pi\theta)^2+4\beta[2\beta^2-1+3(2\pi\theta)^2]]+16\pi^4\overset{*}{C}\gamma^4. \end{aligned}\right\} \quad (3.46)$$

The equation (3.42) therefore satisfies the requirements both theoretical and practical for a statistical surface model.

Depending on the values $\beta, \theta, \overset{*}{C}, \gamma$ this function can correspond to any of the three types of autocorrelation function set out earlier so:

$$\left.\begin{array}{llll} \text{if} \quad \beta > 1 & \text{and} \quad \overset{*}{C}, \theta = 0 & \rho(\tau) & \text{is type 1,} \\ \text{if} \quad \beta > 1 & \text{and} \quad \overset{*}{C} = 0, \quad \theta > 0 & \rho(\tau) & \text{is type 2,} \\ \text{if} \quad \overset{*}{C} \neq 0, & \beta > 1, \quad \theta = 0 & \rho(\tau) & \text{is type 3.} \end{array}\right\}$$

There are other less versatile possibilities for instance the gaussian autocorrelation function used by Beckmann & Spizzichino (1963) with

$$\rho(\tau) = \exp(-\tau^2), \quad (3.47)$$

$$g(\omega) = \pi^{-\frac{1}{2}} \exp(-\tfrac{1}{4}\omega^2); \quad (3.48)$$

another is the Lorentzian function used by Chandley (1976) with

$$\rho(\tau) = 1/(1+\tau^2) \quad (3.49)$$

and

$$g(\omega) = \exp(-\omega), \quad (3.50)$$

where (3.49) satisfies (3.3), when $\rho_1 = \rho(h)$ and $\rho_2 = \rho(2h)$.

Both of these functions have derivatives of all orders at the origin and therefore these are no theoretical problems in using them. It is also possible to modify these functions with a periodic term which in the case of the Lorentzian becomes

$$\rho(\tau) = (\cos 2\pi\theta\tau)/(1+\tau^2) \quad (3.51)$$

and

$$\left.\begin{aligned} g(\omega) &= \exp(-2\pi\theta)\cosh\omega, \quad 0 \leqslant \omega \leqslant 2\pi\theta, \\ g(\omega) &= \cosh(2\pi\theta)\exp(-\omega), \quad \omega \geqslant 2\pi\theta. \end{aligned}\right\} \quad (3.52)$$

Again this is well behaved at the origin: the only real difficulty is the practical one of fitting the different types of surface into this form.

3.6. *Summary*

From what has been said in this section it is clear that there are a number of theoretical and practical considerations which have to be met before a statistical model of a surface can be regarded as acceptable. The necessary criteria have been examined and some models which satisfy them developed.

4. Results

4.1. *Results obtained from the digital analysis of surfaces*

The first step in determining how the surface parameters change as the interval between the measured ordinates is varied is to show that the formulae to be used are correct in a practical sense. In §2 general expressions were developed relating many tribological parameters to two discrete measurements of the autocorrelation function ρ_1 and ρ_2 spaced h and $2h$ from the origin of the correlation function respectively. Change in h changes ρ_1 and ρ_2 in a way dependent on the shape of the correlation function. The changes in ρ_1 and ρ_2 therefore change the values of the surface parameters so that one can simulate the answers which would be obtained from a surface using data points spaced by h and measuring the parameters direct with the conventional formulae. For example, one could measure the mean peak height simply by scanning the data points obtained from the surface, selecting all the peaks and calculating the mean value from the distribution of peaks. Alternatively the technique suggested in this paper is to measure the autocorrelation function, mark off the values of ρ_1 and ρ_2 corresponding to the spacings of h and $2h$ and insert ρ_1 and ρ_2 into formula (2.18) which, with equation (2.20) may be written

$$\text{mean peak} = E[Y_0|T] = [(1-\rho_1)/\pi]^{\frac{1}{2}} \Big/ \frac{2}{\pi}\arctan\left[\left(\frac{3-4\rho_1+\rho_2}{1-\rho_2}\right)^{\frac{1}{2}}\right]. \tag{4.1}$$

This alternative method may well be more effective as a large number of tribological parameters can be considered from just one estimating situation, i.e. estimating the autocorrelation coefficients. If ρ_1 and ρ_2 are known, all the other parameters may, of course, be determined at the same time without examination of the data points, thus ensuring a very fast and simple way of measuring the surface. The only practical modification to all the formulae in §2 is that all terms involving ρ_1 and ρ_2 have to be multiplied by the (r.m.s.)^2 value of the surface because the autocorrelation function was normalized so that $A(0) = 1$. Hence in the mean peak height example above the working formulae would be (if σ is made equal to $(2/\pi)^{\frac{1}{2}}R_a$)

$$E[Y_0|T] = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} R_a \left(\frac{1-\rho_1}{\pi}\right)^{\frac{1}{2}} \Big/ \frac{2}{\pi}\arctan\left[\left(\frac{3-4\rho_1+\rho_2}{1-\rho_2}\right)^{\frac{1}{2}}\right] \tag{4.2}$$

the σ^2 cancelling out in the numerator and denominator in the arc tangent bracket.

To test the validity of the equations in §2 and hence whether the procedures outlined above could be considered equivalent, some real surfaces were examined. The experimental procedure follows.

A Rank Taylor Hobson Talysurf 4 stylus surface finish measuring instrument was connected on line to a Hewlett–Packard 2116 C computer in a manner described by Kinsey & Chetwynd (1973). Ten specimens were selected typical of the various processes found in practical engineering. Four of these are shown together with their autocorrelation functions in figure 5. The surfaces were measured digitally five times at an interval h corresponding to the stylus tip dimension, that is about 2 μm. For each of the tracks various tribological parameters were evaluated using three point analysis and using conventional formulae. The tribological parameters chosen were the peak density, the mean peak height, the average curvature at the peaks together with their standard deviations and the average slope. After this the interval between ordinates h was changed to a longer one and the parameters remeasured. This was repeated a further time making three different sample intervals for each track covering a range of 10 to 1 in all cases. It was felt that this number of parameters and sample intervals was sufficient to establish the validity or otherwise of the theory. At the same time that the parameters were measured the autocorrelation function of the profile trace was also measured, and values of ρ_1 and ρ_2 corresponding to the sample interval h and $2h$ of the digital data found. In fact the whole autocorrelation was not needed: only the two lag positions corresponding to h and $2h$. The measured values of ρ_1 and ρ_2 were then substituted into the relevant formulae taken from §2: in this case equations (2.15), (2.19), (2.20), (2.27), (2.28) and (2.14) respectively for mean peak height and standard deviation, peak density, mean peak curvature and standard deviation, and average slope. These values were then compared with the directly measured values. Table 1 shows the results obtained, expressed as a ratio of direct measurement to that obtained by using the estimated autocorrelation function. The spread values corresponded to the standard deviation of the data. It was found

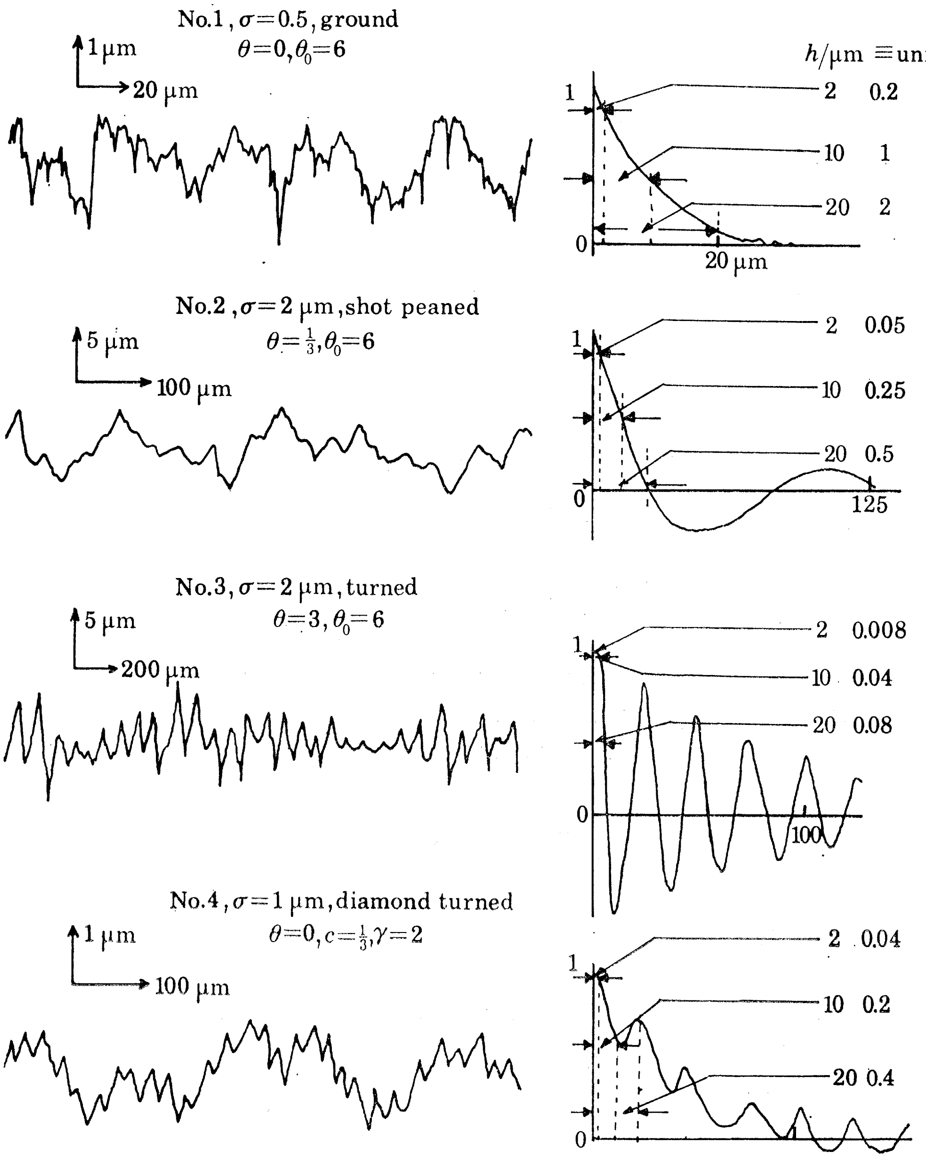

FIGURE 5. Shows four typical surfaces together with their autocorrelation functions. Three sample intervals are shown on each autocorrelation function. Estimates of how these intervals relate to the decay rate are shown. These estimated values of h' can now be inserted into figures 1–13 to show the values of the parameters.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

that neither mean values nor the spreads changed significantly for different sampled data intervals so the results were justifiably pooled for each parameter. These results clearly vindicated the theory and hence the use of the equations derived in § 2. The only problem encountered was the need for the autocorrelation function to be accurately estimated especially when ρ_1 and ρ_2 were close to unity as often happened when h the data spacing was at its smallest value.

By changing the sampling interval of the data for each trace it was possible to observe the changes in the actual values of the parameters measured (see for example table 2 and figure 5). These turned out to be very large as suspected. On the specimens chosen the mean peak height changed by a factor of about 2.5 : 1 as the sample interval was changed from 2 to 24 μm. In the case of the peak density it was about 4 : 1, the peak curvature almost 10 : 1 and the average slope

Table 1.

	peak height		peak density	peak curvature		average slope
	mean	std. dev.		mean	std. dev.	
ground surfaces	0.95 ± 0.03	0.94 ± 0.05	1.00 ± 0.02	0.93 ± 0.02	1.03 ± 0.04	0.95 ± 0.02
turned surfaces $\theta > 1$	1.10 ± 0.11	1.00 ± 0.07	0.93 ± 0.04	1.00 ± 0.06	1.18 ± 0.05	0.98 ± 0.03
grand averages	1.025 ± 0.08	0.97 ± 0.06	0.965 ± 0.04	0.965 ± 0.06	1.105 ± 0.05	0.965 ± 0.03

Table 2.

	peak density			peak height (μm)			peak curvature (mm⁻¹)		
spacing of ordinates/μm	$h = 2$	$h = 10$	$h = 20$	$h = 2$	$h = 10$	$h = 20$	$h = 2$	$h = 10$	$h = 20$
urface no. 1 in figure 5: $\sigma = 0.5$ μm									
equivalent spacing of ordinates	$h' = 0.2$	$h' = 1.0$	$h' = 2.0$						
value of correlation ρ_1 (h)	0.81	0.37	0.13						
value of correlation ρ_2 ($2h$)	0.67	0.13	0.02						
results	0.26	0.30	0.31	0.22	0.38	0.40	133	10.1	2.4
urface no. 2 in figure 5: $\sigma = 2$									
equivalent spacing of ordinates	$h' = 0.05$	$h' = 0.25$	$h' = 0.5$						
value of correlation ρ_1 (h)	0.95	0.67	0.30						
value of correlation ρ_2 ($2h$)	0.86	0.30	−0.18						
results	0.14	0.23	0.34	0.5	1.3	1.7	198	24.9	9.9
urface no. 3 in figure 5: $\sigma = 2$									
equivalent spacing of ordinates	$h' = 0.008$	$h' = 0.04$	$h' = 0.08$						
value of correlation ρ_1 (h)	0.999	0.97	−0.10						
value of correlation ρ_2 ($2h$)	0.997	−0.10	−0.76						
results	0.09	0.16	0.32	2.26	2.28	2.12	16.6	12.7	8.0
urface no. 4 in figure 5: $\sigma = 1$									
equivalent spacing of ordinates	$h' = 0.04$	$h' = 0.2$	$h' = 0.4$						
value of correlation ρ_1 (h)	0.98	0.6	0.73						
value of correlation ρ_2 ($2h$)	0.94	0.73	0.4						
results	0.166	0.36	0.23	0.24	0.49	0.64	60.2	16.4	2.8

ments on figures
.) h is the sample interval in μm from figure 5. h' is the value of h when the unit of length is the distance for decay of correlation ll to 37 %. h' is the value to insert in the abscissa of figures 6–13.
) To convert mean peak height from figure 9 to μm multiply by σ in μm for the surface.
) To convert peak curvature figure 12 to units of mm⁻¹ multiply antilog of value by σ (of surface) in mm and the ratio $(h'/h)^2$ re h is measured in mm.

over 2:1. These factors obviously depended to some extent on the type of specimen but the factors given are very typical. Obviously variations of this proportion are a series problem to the tribologist and the nature and extent of such variations must be known in order to be able to deal with them and so enable meaningful comparisons between workers to be carried out. Because of certain limitations on sampling rate and traverse lengths it was considered impracticable to extend the practical investigation on real surfaces: these restrictions increasing the possibility of missing some of the important features in the relationships between parameters and interval h. The method adopted to explore fully the sampled data problem was to use the autocorrelation model developed in §3, equation (3.42), by fitting different θ and θ_0 values into it to simulate different surfaces and then, using the equations developed in §2, see how the surface parameters changed as the sampling interval changed.

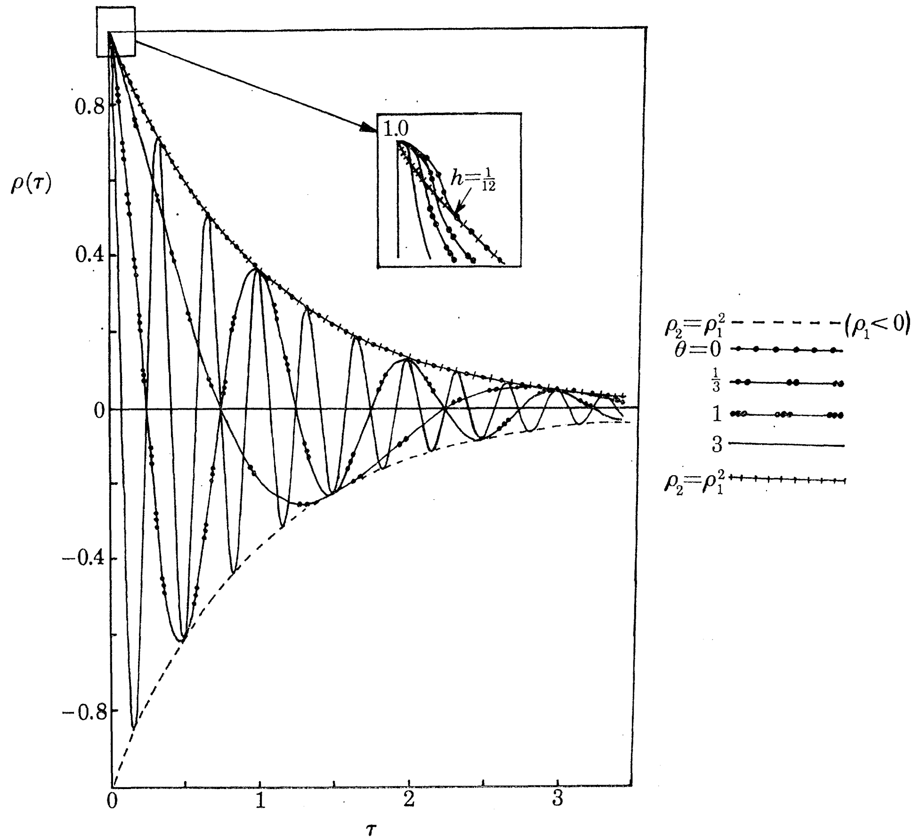

FIGURE 6. Autocorrelation models. This illustrates the behaviour of model 1 and the exponential autocorrelation function at the origin. $\rho(\tau) = e^{-|\tau|}\cos 2\pi\theta t + \dfrac{e^{-\beta|\tau|}}{2\pi\theta_0}\sin 2\pi\theta_0\tau$; $\theta_0 = 6$.

4.2. *Results obtained from analysis of surface models*

The surface likely to cause the largest problems are those of type 2: those in which there is a strong oscillatory component; for this reason surfaces of this type were singled out for the most extensive examination. Surfaces of type 3 do not need to be investigated separately because their behaviour is dominated at small spacing intervals by the exponential component, i.e. ρ_2 is near to the value of ρ_1. Very large additive periodic components are usually confined to rough turning or milling and as such are not often used in tribological experiments.

Figure 6 shows autocorrelation functions obtained using the chosen correlation model (equation (3.42)) referred to here as model 1 for a variety of values of θ and θ_0. The abscissa is the lag or sampling interval h measured in units of the exponential decay. This means that when the abscissa reads unity the exponential envelope of the oscillating correlation function has fallen to e^{-1}, i.e. 37% of its original value. Whitehouse & Archard (1970) chose to call ordinates 'independent' if $\rho_1 = 0.1$ which in their case meant that h was about 2.3 ($= \ln 10$),

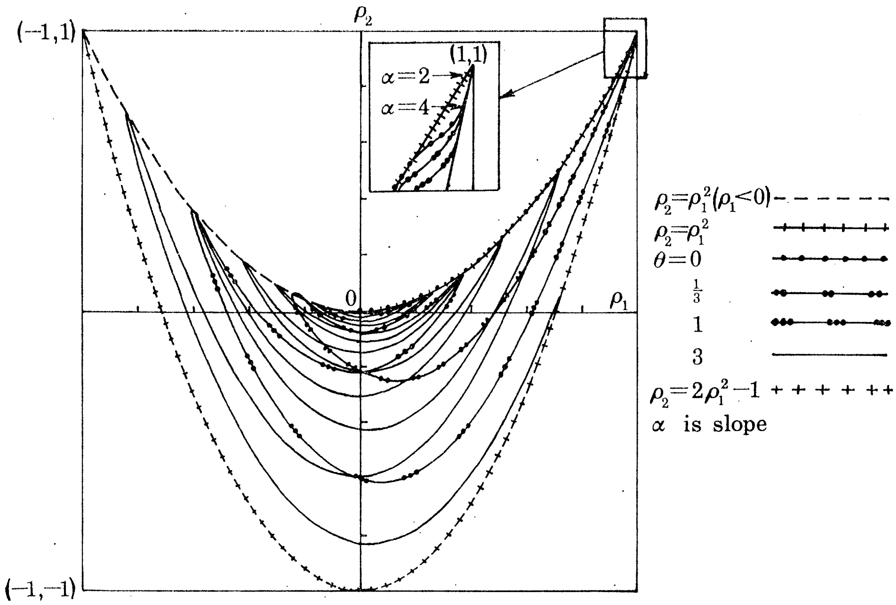

FIGURE 7. Comparison of locus of $\rho_1 = \rho(h)$ $\rho_2 = \rho(h)$ for four variations of the derived autocorrelation model 1 with the exponential autocorrelation function. $\rho(\tau) = e^{-|\tau|}\cos 2\pi\theta t + \dfrac{e^{-\beta|\tau|}}{2\pi\theta_0}\sin 2\pi\theta_0\tau$; $\theta_0 = 6$.

Four types of manufacturing process are simulated in what follows. If $\theta_0 = 6$, the value of $\theta = 0$, corresponding to the near exponential, corresponds to grinding and honing, $\theta = \frac{1}{3}$ corresponds to shot blasting or shot peaning and $\theta = 1$ and $\theta = 3$ are typical of diamond turning. In figures 8–13 different tribological parameters are plotted as a function of h the sampling interval. These figures are intended only to show the nature of the relationship between the parameters and h for various processes in a convenient form. If an experimenter wants to know how sensitive his surface is to sample interval or to determine the parameter values corresponding to a specific scale of size of asperity – as determined by h – he must first measure the autocorrelation function. Then he should measure ρ_1 and ρ_2 from it, corresponding to the sample interval h and $2h$, and insert them into the relevant formula in §2. Figures 6–13 are only meant to give a clue as to the nature of the variations for various types of surface. It might be that his surface correlation function corresponds exactly in which case he need not use the formulae but can read the parameter variations directly off the graph after taking care to match the horizontal scale. An example of how to do this is shown in figure 5 and table 2.

Because, in the formulae of §2, only two parameters, ρ_1 and ρ_2 are involved, it is possible to see how they relate to each other as h is varied for the different types of surface mentioned above. This is shown in figure 7 with ρ_1 as one axis and ρ_2 as the other. The four curves corresponding to the four θ values. All start at (1, 1) where $h = 0$ and progress to (0, 0) where h is infinite; the ordinates are independent. For $\theta = 0$ the curve is almost quadratic. As θ increases the curve oscillates from ρ_1 positive to ρ_1 negative with increasing frequency and requires many more oscillations before the curve approaches (0, 0). For very large θ the curve approaches the boundary $\rho_2 = 2\rho_1^2 - 1$ the reason for which is described in §2.

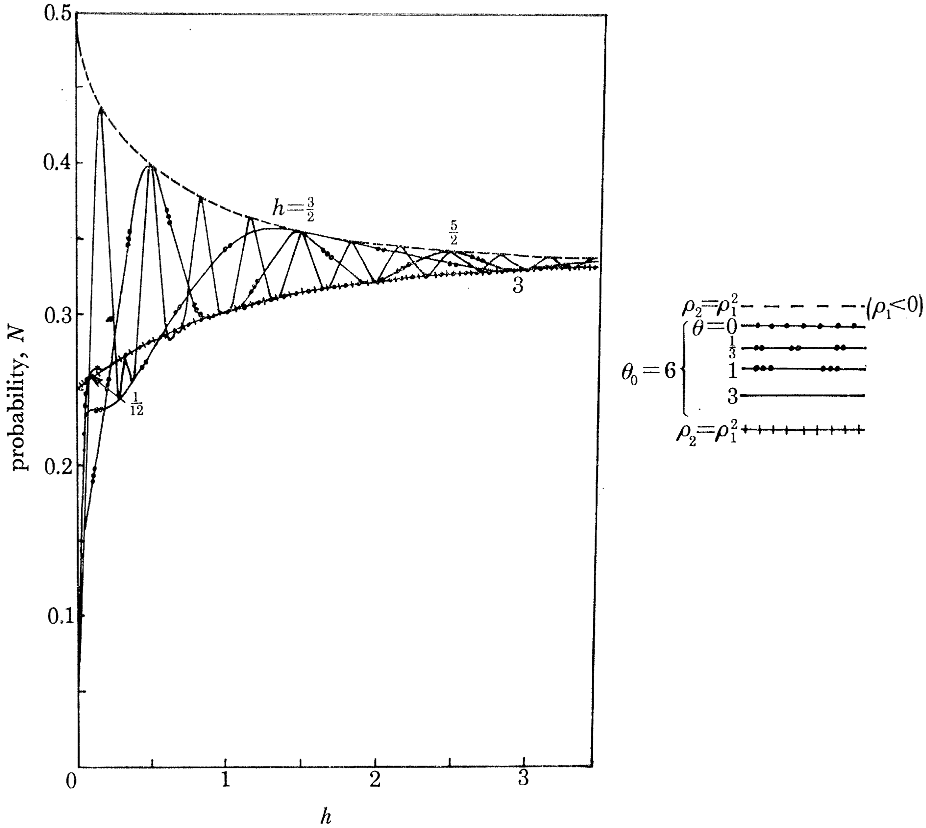

FIGURE 8. The way in which the density of peaks varies with the spacing h between ordinates. Five correlations are shown and the figure illustrates how all peak densities associated with model 1 reduce to zero at $h = 0$.

Consider the parameters in turn; the probability that the central ordinate of a triplet is a peak (given by (2.2)) will be examined first as shown in figure 8. The four types of surface are shown, and compared with the purely exponential. The first point to notice is that no matter what the type of surface (any θ) the probability N converges to $\frac{1}{3}$ as h tends to infinity; this is a consequence of three-point analysis; if all ordinates are independent the chance of one being larger than any other two is $\frac{1}{3}$. However, as h approaches zero with the exponential autocorrelation function, N approaches $\frac{1}{4}$. The reason for this difference is that in the exponential case the derivatives D_2 and D_4 are not defined as was explained in §§3.3, 3.4 and 3.5. The value of zero obtained with the model 1 surfaces is entirely consistent with the need for the discrete distributions to converge in the limit to the continuous (analogue) distributions.

As θ increases the probability N (equivalent to the peak density) oscillates wildly. There is an upper envelope for N and this converges to 0.5 as h approaches zero. This is the result to be expected when sampling a sinusoidal signal with randomly varying amplitude and phase. So the situation can arise with this type of surface that within a very short range of sampling interval h values the number of peaks actually counted from a set of ordinates could vary from close to zero to almost half the ordinates! In the practical results given in §4.1 variations of four to one were indeed found. Notice that in this figure the probability N follows the exponential autocorrelation function for $\theta = 0$ until at $h = \frac{1}{12}$ it drops suddenly to zero, a consequence of $\theta_0 = 6$.

The mean peak height has a similar type of behaviour as is shown in figure 9. Here again the purely exponential curve behaves differently from the surfaces of model 1; near to the origin it drops to zero whereas the more realistic surfaces tend to a non-zero value. In fact it can be shown that this limiting value is given by

$$\lim_{h\to 0} E[Y_0|\,T] = \frac{1}{2}\frac{(\frac{1}{2}\pi)^{\frac{1}{2}}\,n_0}{m_0}, \tag{4.3}$$

where m_0 is the density of peaks (related simply to N) and n_0 is the density of zero crossings. This is obtained from equations (3.7), (3.9) and (3.10).

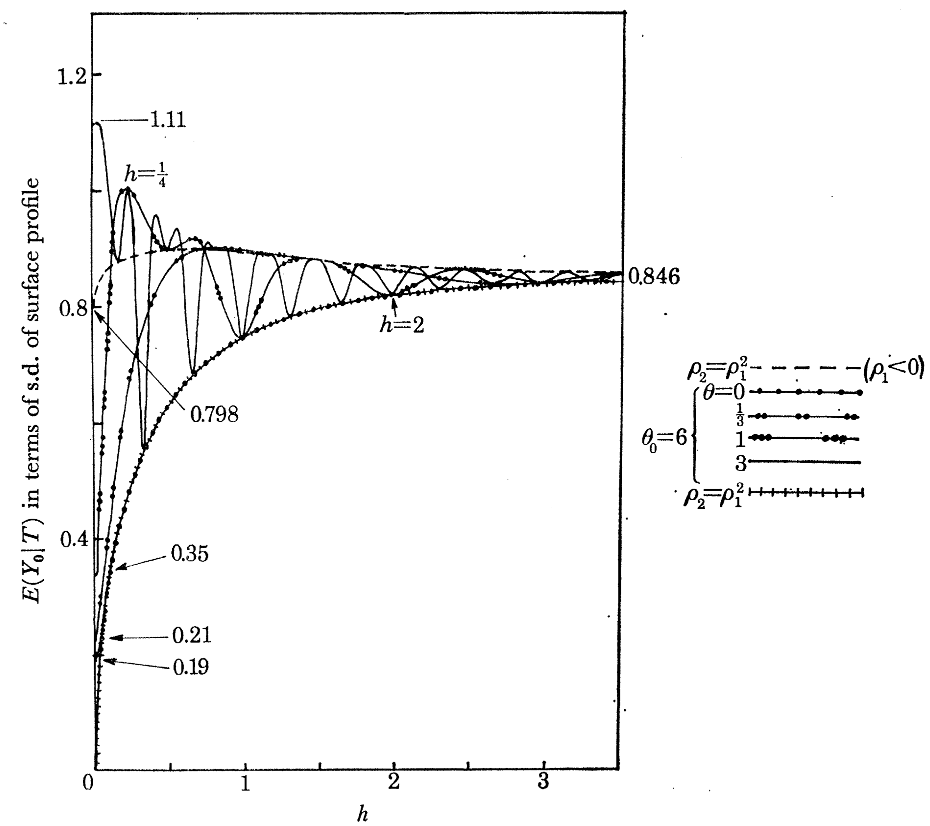

FIGURE 9. The way in which mean peak height changes with the spacing h between ordinates. The figure gives the limiting values at the origin for four variations of model 1 autocorrelation function and the asymptotic value for h.

As the surfaces become more periodic (θ increasing) the mean peak height oscillates as h tends to zero. The upper envelope approaches the value of $(\frac{1}{2}\pi)^{\frac{1}{2}}$ as h approaches zero. The essential message of this figure is that very large variations are possible especially with those surfaces belonging to type 2, the oscillatory case. Very similar results are obtained with the cosine–Lorentzian model for a surface.

The behaviour of the standard deviation (r.m.s.) of the peak height is shown in figure 10. Again there is a common limit for all autocorrelation functions as h tends to infinity. Unlike the other parameters, however, the consequences of the sampling interval changing are not so serious. Variations of only about 40 % of the r.m.s. value of the surface are possible but this is much smaller than with the other parameters. This cannot be said for the correlation coefficient of peak height and curvature which is shown in figure 11. The envelopes and hence possible values can be between zero and unity. The main feature of the mean peak curvature $E[C|T]$ shown in figure 12 and the r.m.s. of the peak curvature shown in figure 13 is that for the real surfaces illustrated by the four embodiments of model 1 the values of $h = 0$ are finite. This is not so for the exponential autocorrelation function whose limiting value is infinity: a consequence of the fact that the differentials at the origin do not exist.

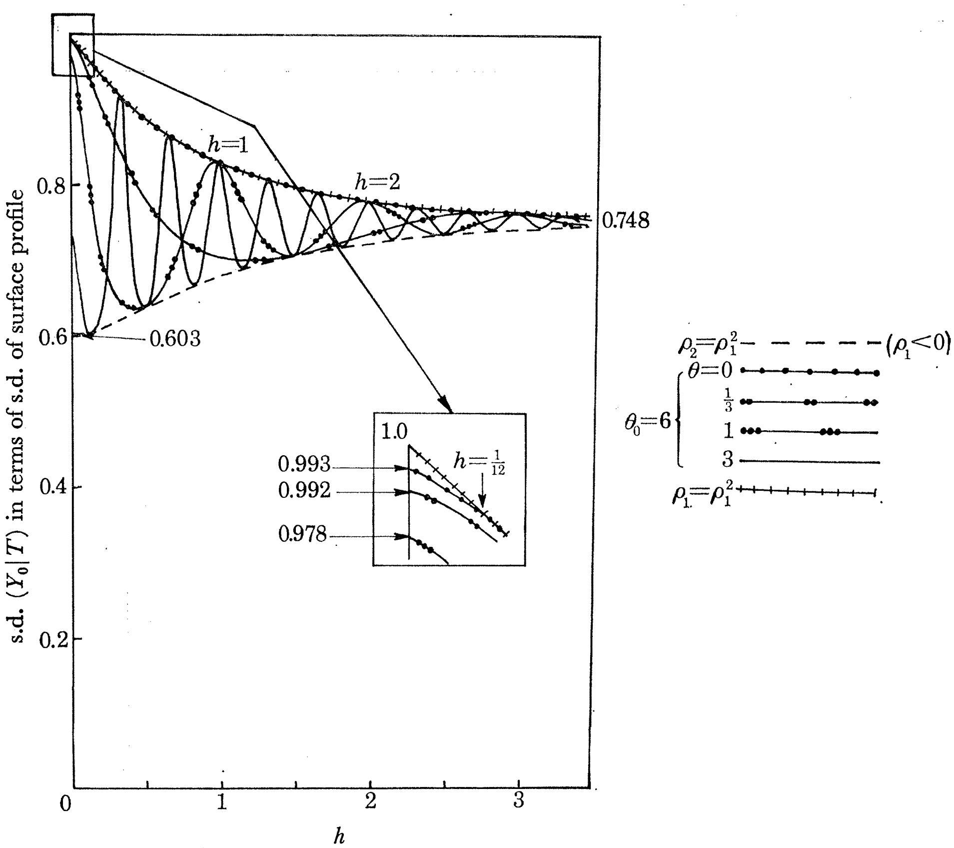

FIGURE 10. The way in which the standard deviation of peak height changes with the spacing h between ordinates. The figure illustrates the difference in behaviour at the origin between four variations of model 1 autocorrelation function and the exponential autocorrelation function.

4.3. *Summary*

Practical results have verified the formulae developed in § 2 for the tribological parameters in terms of the two correlation coefficients. In addition to this large possible variations in parameters are demonstrated. The simulation of real surfaces using the model of a surface developed in § 3 has amplified the conclusions on the sensitivity of the oscillatory type of surface and has emphasized the rate at which such changes can occur with only small changes in the sampling interval.

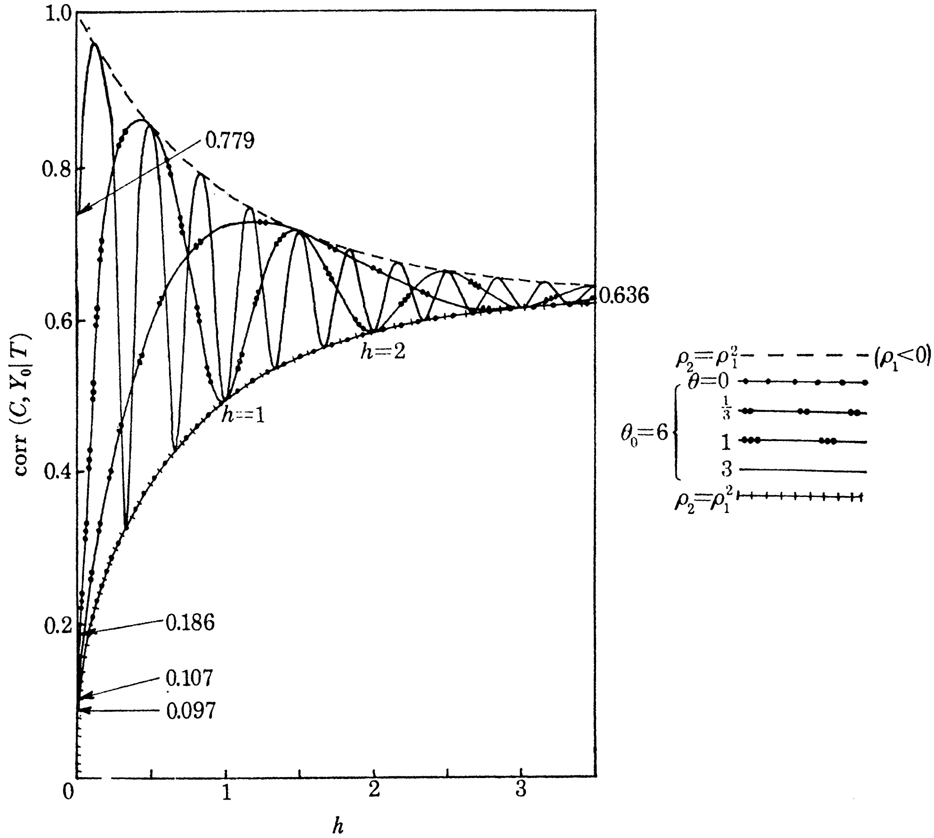

FIGURE 11. The way in which the correlation between peak height and curvature varies with h. The figure shows the limiting values of the four variations of model 1 autocorrelation function at the origin.

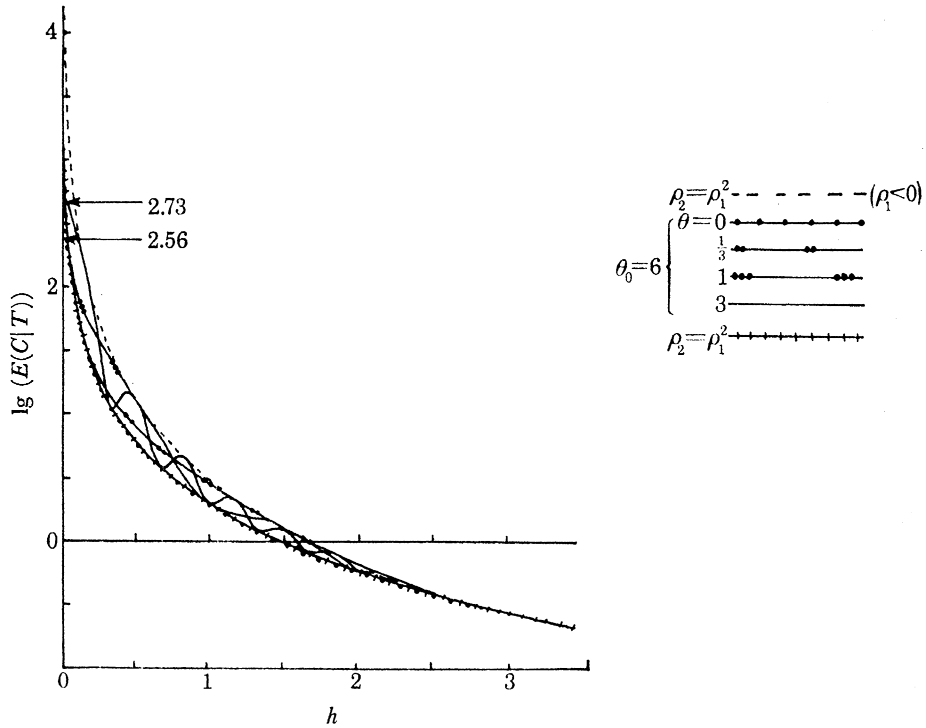

FIGURE 12. Shows the way the mean peak curvature changes with h. The mean value is expressed on a logarithmic scale.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

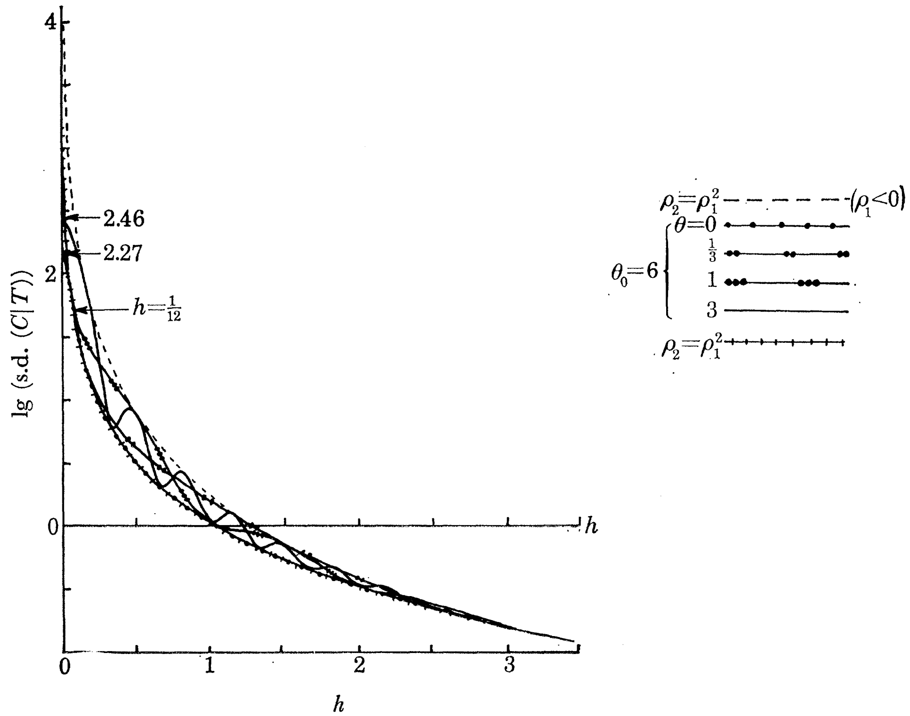

Figure 13. Shows how the standard deviation of peak curvature changes as h changes. It illustrates the breakpoint where the curve for $\rho_2 = \rho_1^2$ diverges from the curve corresponding to model 1 autocorrelation with $\theta = 0$.

5. Discussion

General expressions have been derived which show how each 'scale of size of asperity' contributes to an overall parameter of a surface such as peak height. This has been done by regarding the waveform as a set of sampled data and 'freezing' the scale of size of asperity being evaluated by keeping the sample interval fixed. The whole range of asperities which make up the surface can therefore be investigated by changing the sample interval. This technique using sampled data analysis is considerably simpler than one using a conventional filtering method. The general expressions developed here considerably extend those obtained by Whitehouse & Archard (1970), who limited their treatment to surfaces having a simple exponential correlation function.

It has been demonstrated that the expressions developed for these discrete properties converge in the limit to the theoretical results obtained for continuous waveforms provided that the first four derivatives of the autocorrelation function at the origin exist. Furthermore, using this limiting technique it has been possible to obtain many geometrical characteristics of the surface not previously evaluated. Perhaps the most important are those concerned with the peak curvature distribution. Mean peak curvature used by Greenwood & Williamson (1966) can be expressed very simply as $E(C|T)$ obtained from (3.20), (3.29) and (3.10). Thus

$$E(C|T) = \pi^3 m_0 n_0 (2/\pi)^{\frac{1}{2}}, \tag{5.1}$$

where $(2/\pi)^{\frac{1}{2}}$ is the mean deviation of the profile height, m_0 is the peak density and n_0 is the zero crossing density, all of which can be measured from a profile graph. Similar expressions have been found for the variance of peak curvature, the peak height – curvature correlation and so on.

Indeed it could possibly be argued that derivation of expressions useful in tribology is easier using this limiting technique than by any other method.

During the investigation of how parameters changed with sample interval a general statistical model of a surface has been developed which for the first time satisfies both practical and theoretical requirements. Thus

$$\rho(\tau) = (1-\overset{*}{C})\exp(-|\tau|)\cos 2\pi\theta\tau + \frac{(1-\overset{*}{C})}{2\pi\theta_0}\exp(-|\beta|\tau)\sin 2\pi\theta_0\tau + \overset{*}{C}\cos 2\pi\gamma\tau. \qquad (5.2)$$

Surfaces manufactured by different processes and under different conditions can as a result of this development be modelled effectively. One of the most important results has been the disclosure of the large range of variation of surface parameters as a function of sample interval. Variations of more than two to one are not uncommon for surfaces even within one type. The quantitative extent and nature of these variations has been found by simulation using the surface models described by the autocorrelation function in equation (5.2). These results have been verified by varying the sample interval on real surfaces and measuring the surface parameters directly.

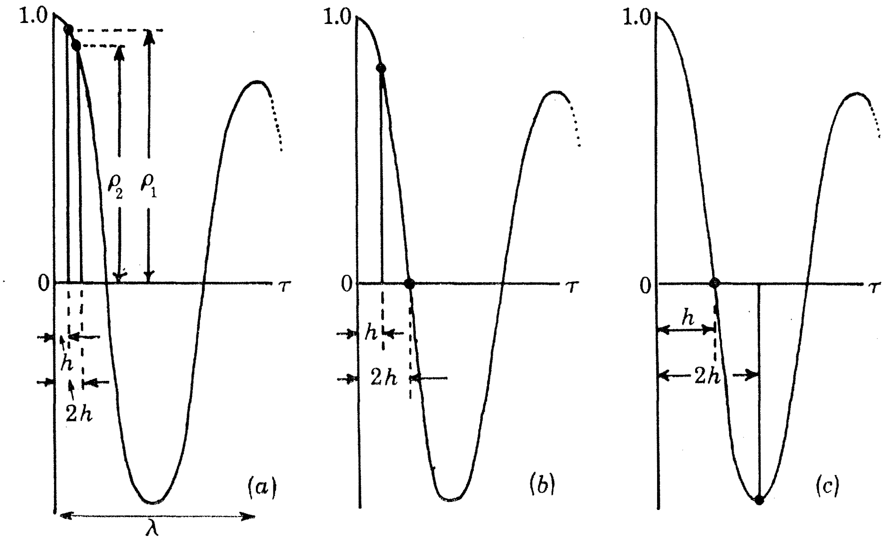

FIGURE 14. The worst cases for parameter variations are shown. The autocorrelation function shows three possible sample intervals each nominally within the Nyquist criterion. It is clear that only in (*a*) will there be any real stability because ρ_1 and ρ_2 are not changing rapidly relative to each other.

Those surfaces which have a highly oscillatory autocorrelation function of the kind which is modulated (that is, where the random term is multiplied by the deterministic) give rise to most problems. This is obvious from figures 7–13. If the sampling occurs as shown in figure 14 (*a*) and (*b*) considerable differences will occur in the value of the parameters. Indeed, if as is suggested in figure 14 (*c*) the sampling interval is just a quarter of the periodic wavelength and therefore well within the Nyquist criterion of two samples per period the rate of change of ρ_2 relative to ρ_1 is near to its maximum and will be exceedingly sensitive to small changes in h. It appears from this that if the criterion for sampling is one of stability of the measured parameters then h should be made either much smaller, or alternatively, more nearly equal to half the period, so that ρ_1 and ρ_2 both lie on the peaks of the cosinusoid. From the point of view of the possible presence of harmonics the first criterion would be preferable, that is sampling should be between $\lambda/8$ and

$\lambda/16$ where λ is the period of the oscillation in the autocorrelation function corresponding roughly to the tool feed mark. This in effect means that for the purpose of stability of results the sampling could be four or more times shorter than that required by the Nyquist criterion. There is another way of making the observation that three-point analysis becomes less acceptable when the differences between ρ_1 and ρ_2 become large and that is in terms of the surfaces themselves. In this case it is only accurate to use three-point analysis for measuring peak properties such as curvature when fourth-order central differences are small compared with first-order differences between the heights of ordinates. Thus

$$2y_0 - (y_{+1} + y_{-1}) \gg \frac{\delta^4}{12} y_0. \tag{5.3}$$

This same criterion does not apply to the density of peaks or to the peak height parameters (Whitehouse 1978, to be published).

It is clear from these considerations of the effect of the spacing between data points that a lot of trouble can be expected when fitting values of mean peak height, curvature into contact, friction or lubrication formulae. The results certainly show why tribologists and other surface researchers are finding it difficult to get agreement with digital data obtained from the same surface! It may well be that variations obtainable on one surface due to this effect could be larger than those variations between surfaces. It therefore suggests that in order to determine the best interval for the sampling of a practical surface one needs to have some idea of the scale of size of the asperity which is considered most important and then fit the sample interval to encompass it adequately in the data.

Unquestionably the most important result of this paper is the demonstration that by estimating just two points on the autocorrelation function and knowing the r.m.s. of the surface profile one is able to predict all the tribological parameters that could be measured from a profile using the same interval; the need to estimate them directly is therefore removed. The problems of measuring many parameters is therefore reduced to evaluating two points on an autocorrelation function and inserting them into the relevant formulae in §2. The only qualifications to this are that the surfaces are nominally gaussian and that the sampling is not too large significantly to reduce the accuracy of the three-point method. This latter qualification is only a safeguard, the technique will always give the same results as the direct use of three-point analysis on the profile; the question is whether or not the three-point analysis is sufficiently accurate. Since most surfaces on critical components in tribological situations involve finishing processes, they are likely to be gaussian in character. In wear situations, however, this method is likely to apply only to initial conditions but it is often the original finish which dominates subsequent behaviour (Whitehouse 1971). In other functional situations such as lubrication, contact, electrical, thermal and optical behaviour the use of the two values of the autocorrelation function appear to be generally applicable because the surface is not severely changed during the functional operation. In fact provided the skew of the distribution is within about ± 1, results from non-gaussian surfaces still agree with the theory.

The only restriction on the type of autocorrelation function for which this technique will be satisfactory is that $2\rho_1^2 - 1 < \rho_2 < 1$; this can only be violated for a purely deterministic waveform such as a cosinusoid and consequently is not a practical restriction.

Measurement of the autocorrelation function in two places is straightforward nowadays with the availability of digital correlators. It is also possible to estimate the autocorrelation function to a good degree of accuracy by hand from the profile graph; the two points could be evaluated in a few minutes (Whitehouse 1976) so that pilot estimations of the tribological parameters could be obtained very simply without the use of expensive equipment. Obviously in cases which involve, for example, the control of the manufacturing process (where surfaces are often severely non-gaussian) more refined apparatus will always be needed.

One considerable advantage of specifying a surface in terms of the three correlation coefficients $\rho(0) = \text{r.m.s.}^2$, $\rho(h)$ and $\rho(2h)$ is that they are additive from surface to surface. This means that the properties of the gap can be found by first adding the coefficients of the mating surfaces. Thus if σ_A^2 is the r.m.s. value of surface A and $\rho_A(h)$, $\rho_A(2h)$ are its coefficients and similarly for surface B. Then the gap coefficients are

$$\left.\begin{aligned} \sigma_g^2 &= \sigma_A^2 + \sigma_B^2, \\ \rho_g(h) &= (\sigma_A^2 \rho_A(h) + \sigma_B^2 \rho_B(h))/\sigma_g^2, \\ \rho_g(2h) &= (\sigma_A^2 \rho_A(2h) + \sigma_A^2 \rho_B(2h))/\sigma_g^2. \end{aligned}\right\} \tag{5.4}$$

Tribological gap properties such as given in §2 are obtained by inserting the gap correlation coefficients into the relevant formulae.

Use of density of peaks and zero crossings instead of correlation coefficients is not an attractive alternative because they are much more sensitive to instrumental error and because they are not simply additive. Thus using the nomenclature

$$H\rho = \frac{1}{\pi}\sqrt{\frac{D_2}{D_0}}, \quad D = \frac{1}{2\pi}\sqrt{\frac{D_4}{D_2}}$$

we get for comparison

$$\left.\begin{aligned} \sigma_g^2 &= \sigma_A^2 + \sigma_B^2, \\ H_g &= ((\sigma_A^2 H_A^2 + \sigma_B^2 H_B^2)/\sigma_g^2)^{\frac{1}{2}}, \\ D_g &= \left(\frac{\sigma_A^2 H_A^2 D_A^2 + \sigma_B^2 H_B^2 D_B^2}{\sigma_A^2 H_A^2 + \sigma_B^2 H_B^2}\right)^{\frac{1}{2}}. \end{aligned}\right\} \tag{5.5}$$

From (5.5) it is clear that the errors are cumulative and because of instrumental problems, such as frequency response, are more likely to occur.

To extend the theory to cover five-point analysis is possible but more difficult and hardly justifiable. The next step is the three-dimensional aspect when the surfaces are, and are not, isotropic.

Summarizing, this paper has been an attempt to bridge the gap between the purely random process analysis of surfaces and their measurement. In the past tribological parameters have only been related theoretically to the autocorrelation function behaviour at the origin (or the moments of the spectrum which are notoriously difficult to measure) (Nayak 1973) whereas in this paper the theory has been taken a step further to enable the practical evaluation of tribological parameters in terms of two points on the measured autocorrelation function. It should therefore help researchers to measure parameters without much involvement in digital analysis or random process theory.

We wish to thank the Directors of Rank Taylor Hobson for permission to publish and for their continued support for work of this kind.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

References

Archard, J. F. 1957 *Proc. R. Soc. Lond.* A **243**, 190.

Beckman, P. & Spizzichina, A. 1963 *The scattering of electromagnetic waves from rough surfaces.* Oxford: Pergamon.

Bendat, J. S. 1958 *Principles and applications of random noise theory.* New York: Wiley.

Bowden, F. P. & Tabor, D. 1954 *The function and lubrication of solids.* Oxford University Press.

Chandley, P. J. 1976 Surface roughness measurements from coherent light scattering. *Optics and Quantum Elec.* **8**, 323.

Cox, D. R. & Miller, H. D. 1965 *The theory of stochastic processes.* London: Methuen.

Davenport, W. B. & Root, W. L. 1958 *An introduction to the theory of random signals and noise.* New York: McGraw Hill.

Finney, D. J. 1962 Cumulants of truncated multinormal distributions. *J. R. Statist. Soc.* B **24**, 535.

Greenwood, J. A. & Williamson, J. B. P. 1966 Contact of nominally flat surfaces. *Proc. R. Soc. Lond.* A **295**, 300.

Greenwood, J. A. & Williamson, J. B. P. 1977 *Development in the theory of surface topography.* Leeds/Lyon Conf. Trib. Lyon.

Johnson, N. L. & Kotz, S. 1970 *Distributions in statistics: Continuous univariate distributions*, vol. 1. Boston, Mass.: Houghton Mifflin.

Kinsey, D. & Chetwynd, D. G. 1973 Some aspects of the use of digital computers to the on-line measurement of surfaces IMEKO, Dresden June 1973.

Lawley, D. N. 1943 A note on Karl Pearson's selection formulae. *Proc. R. Soc. Edinb.* **62**, 28.

Longuet-Higgins, M. S. 1957 Statistical analysis of a random moving surface. *Proc. R. Soc. Lond.* A **249**, 321.

Mitchell, L. A. & Rowe, M. D. 1968 *Proc. Inst. Mech. Engrs* **182**, Pt 3K (conference on the properties and metrology of surfaces), p. 101.

Nayak, P. R. 1971 Random process model of rough surfaces. *Trans. ASME J. of Lub. Tech.* July 1971, p. 398.

Nayak, P. R. 1973 Random process model of rough surfaces in plastic contact. *Wear*, **26**, 305.

Peklenik, J. 1967–68 *Proc. Inst. Mech. Engrs* **182**, Pt 3K (conference on the properties and metrology of surfaces), p. 108.

Regier, M. H. & Hamdan, M. A. 1971 Correlation in a bivariate normal distribution with truncation in both variables. *Aust. J. Statist.* **13**, 77.

Rice, S. O. 1944 Mathematical analysis of random noise. *Bell Syst. tech. J.* **23**, 282.

Rosenbaum, S. 1961 Moments of a truncated bivariate normal distribution. *J. R. Statist. Soc.* B **23**, 405.

Tallis, G. M. 1961 The moment generating functions of the truncated multinormal distribution. *J. R. Statist. Soc.* B **23**, 233.

Weiler, H. 1959 Mean and standard deviations of a truncated normal bivariate distribution *Aust. J. Statist.* **1**, 73.

Whitehouse, D. J. & Archard, J. F. 1970 The properties of random surfaces of significance in their contact. *Proc. R. Soc. Lond.* A **316**, 97.

Whitehouse, D. J. 1971 Ph.D. thesis. Leicester University.

Whitehouse, D. J. 1974 Theoretical analysis of stylus integration. *Annals CIRP.* **23**, 181.

Whitehouse, D. J. 1976 Approximate methods of assessment of surface topography parameters *Annals CIRP* **25**, 461.

Whitehouse, D. J. 1978 The digital measurement of peaks on surface profiles *J. mech. Engrng Sci.* **20**, no. 4 ,221.