Sampling in a two-dimensional plane

1985 J. Phys. A: Math. Gen. 18 2465

(http://iopscience.iop.org/0305-4470/18/13/023)

J. Phys. A: Math. Gen. 18 (1985) 2465-2477. Printed in Great Britain

Sampling in a two-dimensional plane

D J Whitehouse† and M J Phillips‡

† Department of Engineering, University of Warwick, Coventry CV4 7AL, UK
‡ Department of Mathematics, University of Leicester, LE1 7RH, UK

Received 26 November 1984

Abstract. The conventional way to measure the two-dimensional geometry of a surface is to map it using a rectangular grid pattern. This paper explores the trigonal method of mapping and compares the results with the rectangular as well as the theoretical values which should be obtained for a random surface.

1. Introduction

Surfaces are rough. This roughness dominates many of the applications of engineering surfaces. In recent years this awareness has given rise to great interest in measuring surfaces in order to predict the performance and preferably optimise it. Because the geometrical parameters of fundamental importance tend to be somewhat complicated the current trend is to measure the surface using digital techniques. The parameters of interest are rarely those used to control the manufacturing process such as R_a, the arithmetic average value. Functional parameters tend to be more associated with peaks and valleys on the surface. An example is the peak curvature in contact theory (Greenwood and Williamson 1966) and the slope of the surface in optical reflection (Welford 1977). Of course these parameters do not usually refer to those obtained from a single profile or cross section of the surface but to the complete surface: the two-dimensional characteristics. For a completely anisotropic surface measured across the lay the two are the same, but in general they are not. This has resulted in the growth of techniques which can scan over the whole surface to enable the true geometry to be revealed. Until recently this has been difficult for more than one reason; firstly the large storage necessary and secondly the accurate scan mechanism. To some extent both problems have been eased but there are still many things which require clarification. One of these is the pattern of scanning required for digital analysis and another is the change in parameter values which occur with different numerical models and yet another most significant effect is the way in which the parameters change with sampling. These problems were first tackled by Whitehouse and Archard (1970) and Whitehouse and Phillips (1978, 1982). The first and simplest question was that of determining the surface parameters from a single profile of the workpiece. Parameters such as peak height, curvature and surface slope were derived using the 3-point method to describe a peak. In this a peak is said to exist if the central ordinate (digital) measurement of three consecutive ones stood out as the highest. In the first paper a simple exponential correlation function model was assumed for the surface for two reasons: the first undeniably being ease of calculation, the second because by the very nature of random manufacturing processes the probability of a grit hitting any spot is

Poissonian i.e. first-order Markov, indicating at least an exponential envelope to the correlation function (Hamed *et al* 1978) and therefore relevant to current thinking on *fractal* surfaces (Jordan *et al* 1983). The case for a general autocorrelation function rather than an exponential was considered by Whitehouse and Phillips for profiles. This analysis was extended (Whitehouse and Phillips 1982) to two dimensions in which the 3-point analysis naturally extends itself to five point. Again a general well behaved autocorrelation function was considered and specific examples chosen to illustrate the variability of the parameters with sample interval and correlation model. In this case the whole surface is covered by a rectangular grid of measurements similar to that used in the measurement of straightness. Analytical expressions for the summit properties (two-dimensional peak) were worked out as before and were compared with the individual profile properties on the one hand and purely continuous properties on the other. For these comparisons correlation functions corresponding to a large number of surfaces were used. For simplicity the surfaces were assumed to be isotropic and Gaussian in height distribution to enable the theory of truncated random variables (Phillips 1984 and the appendix of Whitehouse and Phillips 1982) to be used. A more straightforward account of this area of work has recently been given by Greenwood (1984). It has to be noted here that future work in this subject will have to take into account cases where the autocorrelation function is not definable in the form given by equation (4.1) discussed later.

The two sampling patterns so far investigated are the obvious and simplest ones to use, but they are not necessarily the best. The objective in this paper is to examine another sampling pattern and investigate the practical and theoretical implications.

The various sampling plans can best be visualised by means of a circle whose centre is an ordinate, and around the circumference of which are k evenly spaced ordinates. For profile sampling k is two and the angular spacing is π. The rectangular grid sampling corresponds to $k=4$ and a spacing of $\pi/2$. The former is referred to as having diagonal symmetry and the latter tetragonal.

In this paper another value of k is considered, namely $k=3$ (trigonal symmetry). Analytical expressions are obtained for a number of surface parameters using the methods outlined before and some practical issues will be discussed, especially with respect to the measurement of fine surfaces.

2. Sampling schemes in a two-dimensional plane

A number of possible sampling schemes for sampling in a two-dimensional plane will now be presented. They are illustrated in figure 1. In all cases the distance between measurements is h.

Firstly there is sampling along a straight line (from a *profile* of the surface). This sampling scheme only takes measurements in *one* dimension of the plane. However, it is presented for completeness and because it was the first case considered by Whitehouse and Archard (1970) and Whitehouse and Phillips (1978). It is illustrated in figure 1(a) with $k=2$ and $\theta=\pi$.

Secondly a sampling scheme could be used which would take measurements at the vertices of a hexagonal grid. The summit properties could be defined using four ordinates, i.e. the measurement at a vertex and the three adjacent ordinates at a distance h from the chosen vertex. This would be the case when $k=3$ and $\theta=\frac{2}{3}\pi$, and will be referred to as the *trigonal* symmetry case.

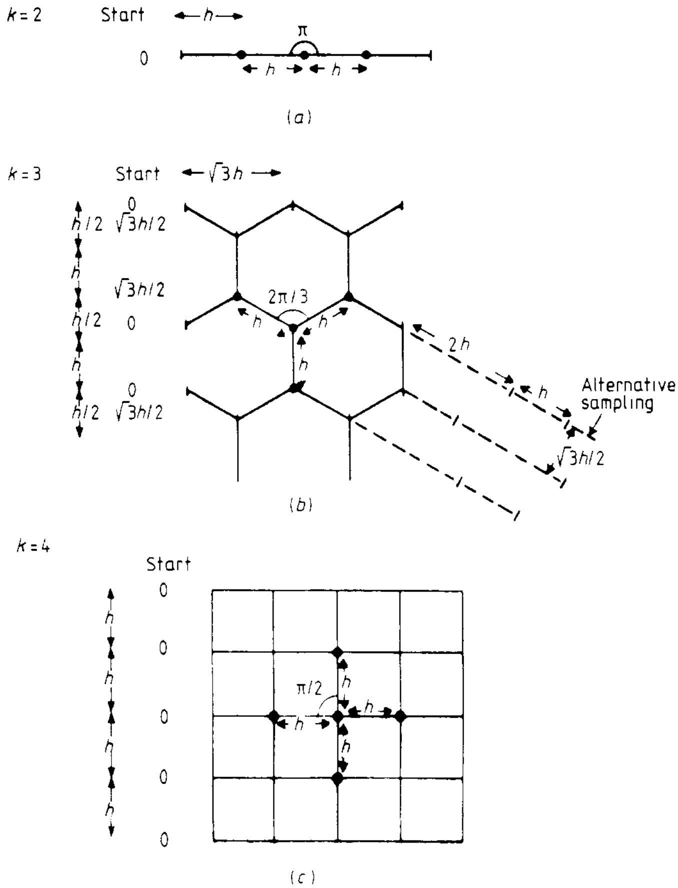

Figure 1. Sampling patterns for the (*a*) 3-point (digonal), (*b*) 4-point (trigonal) and (*c*) 5-point (tetragonal) numerical models with a spacing *h* between ordinates.

In order to produce such a hexagonal grid it would be necessary to sample along parallel lines separated alternately by a distance $\frac{1}{2}h$ and h. The spacing between ordinates along a line would be $\sqrt{3}h$ but the position at which the first ordinate is measured would be 0 for the $(4j-3)$th and $4j$th lines and $\frac{1}{2}\sqrt{3}h$ for the $(4j-2)$th and $(4j-1)$th lines, for $j \geq 1$. This is illustrated in figure 1(*b*). Alternatively one could sample along parallel lines a distance $\frac{1}{2}\sqrt{3}h$ apart, but this would involve a different position for the first ordinates and the spacing between ordinates would alternately be h and $2h$.

Thirdly there is sampling on a square grid. This was considered by Whitehouse and Phillips (1982) and Greenwood (1984) and will be referred to as the *tetragonal* symmetry case. It is illustrated in figure 1(*c*) with $k=4$ and $\theta=\frac{1}{2}\pi$. The sampling scheme requires sampling along parallel lines separated by a distance h and with a spacing between ordinates along a line of h.

In this paper results will be given for the hexagonal grid or trigonal symmetry. For purposes of comparison the cases when $k=2$, 3 and 4 will be considered. The notation of these papers will be that used below. If the m random variables $X=(X_1, X_2, \ldots, X_m)$

have a joint multivariable Gaussian (normal) distribution with mean μ and variance-covariance matrix V then this is denoted by $X \sim N\ [\mu, V]$. Also the convention of using an upper case letter for a random variable and a lower case letter for a realisation of the random variable will be followed.

3. The hexagonal grid in the trigonal symmetry case

Results have been obtained for the probability density function and expectation (mean) of peak (or summit) height, the density of summits and the expected peak (or summit) curvature in the cases when $k=2$ by Whitehouse and Phillips (1978) and when $k=4$ by Whitehouse and Phillips (1982). The results for the hexagonal grid ($k=3$) in the trigonal symmetry case will now be given. These can be obtained from the general results of truncated random variables given in Phillips (1984) and in the appendix of Whitehouse and Phillips (1982).

For measurements with four ordinates let z_0 be the height of the central ordinate and s_1, s_2 and s_3 be the differences between this ordinate and the three adjacent ordinates at a distance h. The ordinate z_0 will be defined to be a 4-point summit if s_1, s_2 and s_3 are all positive. By analogy with the 3-point and 5-point definitions of peak (or summit) curvature the discrete definition of 4-point curvature is

$$c = 2(s_1+s_2+s_3)/3h^2. \tag{3.1}$$

Assuming that the surface height measurements have a multivariate Gaussian distribution and that the surface is isotropic then

$$(Z_0, (2-2\rho_1)^{-1/2}(S_1, S_2, S_3)) \sim N[0;\ V_4], \tag{3.2}$$

$$V_4 = \begin{bmatrix} 1 & d & d & d \\ d & 1 & a & a \\ d & a & 1 & a \\ d & a & a & 1 \end{bmatrix} \tag{3.3}$$

with

$$d = (\tfrac{1}{2}-\tfrac{1}{2}\rho_1)^{1/2} \tag{3.4}$$

and

$$a = \tfrac{1}{2}(1-2\rho_1+\rho_{\sqrt{3}})/(1-\rho_1), \tag{3.5}$$

where $\rho_1 = \rho(h)$ and $\rho_{\sqrt{3}} = \rho(\sqrt{3}h)$ and $\rho(t)$ is the correlation coefficient between ordinates a distance t apart.

If T_4 is the event $(S_1>0,\ S_2>0,\ S_3>0))$ then the distribution of 4-point summit height is the conditional distribution of Z_0 given that T_4 has occurred. This can be obtained using the results of the appendix of Whitehouse and Phillips (1982) with $m=3$,

$$X = (2-2\rho_1)^{-1/2}(S_1, S_2, S_3)', \tag{3.6}$$

$$d = (\tfrac{1}{2}-\tfrac{1}{2}\rho_1)^{1/2} \tag{3.7}$$

and

$$V = V_3 = \begin{bmatrix} 1 & a & a \\ a & 1 & a \\ a & a & 1 \end{bmatrix}, \tag{3.8}$$

so that

$$\lambda = 1 + 2a. \tag{3.9}$$

Then the probability density function of the 4-point summit height distribution is given by

$$f(z_0 | T_4) = \frac{\Phi^{(3)}(z_0[(1-\rho_1)/(1+\rho_1)]^{1/2};\ \boldsymbol{V}_c)\phi(z_0)}{\Phi^{(3)}(0;\ \boldsymbol{V}_3)}, \tag{3.10}$$

where

$$\boldsymbol{V}_c = \begin{bmatrix} 1 & a_c & a_c \\ a_c & 1 & a_c \\ a_c & a_c & 1 \end{bmatrix} \tag{3.11}$$

and

$$a_c = (\rho_{\surd 3} - \rho_1^2)/(1 - \rho_1^2). \tag{3.12}$$

($\Phi^{(n)}(y'$: $\boldsymbol{V})$ is the cumulative distribution function at y of the n-dimensional multivariate Gaussian distribution with zero expectation and variance-covariance matrix $\boldsymbol{V}$. $\phi(x)$ is the probability density function of the univariate standard Gaussian distribution.)

The denominator of (3.10) is the orthant probability which gives the probability that an ordinate is a 4-point summit using the nomenclature of Cheng (1969). Hence, from David (1953),

$$pr(T_4) = \Phi^{(3)}(0;\ \boldsymbol{V}_3) = \tfrac{1}{2} - \tfrac{3}{4}(\pi)^{-1}\cos^{-1}(a). \tag{3.13}$$

The expected (mean) 4-point summit height is given by

$$E(Z_0 | T_4) = \frac{3(1-\rho_1)^{1/2}\Phi^{(2)}(0;\ \boldsymbol{B}_3)}{2\sqrt{\pi}\Phi^{(3)}(0;\ \boldsymbol{V}_3)}, \tag{3.14}$$

where

$$B_3 = \begin{bmatrix} 1-a^2 & a(1-a) \\ a(1-a) & 1-a^2 \end{bmatrix}. \tag{3.15}$$

Hence

$$E(Z_0 | T_4) = 3(1-\rho_1)^{1/2}\left[\frac{1}{4} + \frac{1}{2\pi}\sin^{-1}\left(\frac{1-2\rho_1+\rho_{\surd 3}}{3-4\rho_1+\rho_{\surd 3}}\right)\right] \times \left\{2\sqrt{\pi}\left[\frac{1}{2} - \frac{3}{4\pi}\cos^{-1}\left(\frac{1-2\rho_1+\rho_{\surd 3}}{2-2\rho_1}\right)\right]\right\}^{-1}. \tag{3.16}$$

The distribution of the height Z_0 of a 4-point summit conditional on a curvature C is Gaussian with an expectation given by

$$E(Z_0 | C = c,\ T_4) = \frac{3h^2(1-\rho_1)c}{4(2-3\rho_1+\rho_{\surd 3})} \tag{3.17}$$

and variance given by

$$\mathrm{Var}(Z_0 | T_4) = \frac{1-3\rho_1+2\rho_{\surd 3}}{2(2-3\rho_1+\rho_{\surd 3})}. \tag{3.18}$$

This is the same as the distribution of the height Z_0 of an ordinate conditional on the 4-point curvature *but not* conditional on the ordinate being a summit. This is a result which holds for the three values of $k=2$, 3 and 4. This is because V_k is of the form

$$V_k=\begin{pmatrix}1 & d1' \\ d1 & \bar{V}\end{pmatrix}, \tag{3.19}$$

where $\bar{V}$ is a correlation matrix with a constant row (column) sum. This result enables the expected $(k+1)$-point peak (or summit) curvature to be obtained from the expected $(k+1)$-point peak (or summit) height.

Hence the expected 4-point summit curvature is given by

$$\begin{aligned}E(C\mid T_4)&=\frac{4(2-3\rho_1+\rho_{\surd 3})}{3h^2(1-\rho_1)}\cdot E(Z_0\mid T_4)\\&=2(2-3\rho_1+\rho_{\surd 3})\left[\frac{1}{4}+\frac{1}{2\pi}\sin^{-1}\left(\frac{1-2\rho_1+\rho_{\surd 3}}{3-4\rho_1+\rho_{\surd 3}}\right)\right]\\&\quad\times\left\{h^2[\pi(1-\rho_1)]^{1/2}\left[\frac{1}{2}-\frac{3}{4\pi}\cos^{-1}\left(\frac{1-2\rho_1+\rho_{\surd 3}}{2-2\rho_1}\right)\right]\right\}^{-1}.\end{aligned} \tag{3.20}$$

It is also possible to obtain the following simple connection between the variances of $(k+1)$-point peak (or summit) height and curvature:

$$\mathrm{Var}(Z_0\mid T_{k+1})=\mathrm{Var}(Z_0\mid C,\,T_{k+1})+\left[\frac{E(Z_0\mid C=c,\,T_{k+1})}{c}\right]^2\cdot\mathrm{Var}(C\mid T_{k+1}). \tag{3.21}$$

This relation was shown by Whitehouse and Phillips (1978) for $k=2$ (with $Z_0\equiv Y_0$) and by Greenwood (1984) for $k=4$.

So by the application of the theory of Gaussian truncated random variables it has been possible to obtain connections between the expectations and variances of 4-point summit and curvature ($k=3$).

4. The effect of sampling interval and limiting results

It is vital to investigate the variation of parameters with h because it is due to the large number of possible differences in sampling interval that the scatter of measured values of parameters occur between investigators.

The distributions of 4-point summit height and curvature ($k=3$) have been derived in terms of correlation coefficients between ordinates. These two correlation coefficients are ρ_1, for ordinates a distance h apart, and $\rho_{\surd 3}$, for ordinates a distance $\sqrt{3}h$ apart. If the surface is isotropic and the autocorrelation function is $\rho(t)$, then $\rho_1=\rho(h)$ and $\rho_{\surd 3}=\rho(\sqrt{3}h)$. So ρ_1 and $\rho_{\surd 3}$ will vary as h varies, depending on the shape of the autocorrelation function of the surface.

Results for the summit height have been obtained by Nayak (1971) for the continuous surface. So it is possible to compare his results with those obtained for the discrete results of § 3 as the sampling interval h converges to zero.

To do this it is necessary to make assumptions about the behaviour of the autocorrelation function $\rho(h)$ near the origin. It will be assumed that

$$\rho(h)=1+D_2h^2/2!+D_4h^4/4!+o(h^4) \tag{4.1}$$

where $D_2 < 0$, $D_4 > 0$ and

$$\eta = -D_2(D_4)^{-1/2} < (\tfrac{5}{6})^{1/2}. \tag{4.2}$$

D_2 and D_4 are the second and fourth derivatives of the autocorrelation function at the origin.

Comparison will be made for estimates of parameters measuring peak and summit properties of the surface. This will be done for the three cases of 3-, 4- and 5-point estimates corresponding to $k = 2$, 3 and 4, respectively.

The first parameter that will be considered is the density of peaks or summits. These parameters are known for a continuous random Gaussian surface and were given for peaks as

$$D_{\text{peak}} = \tfrac{1}{2}\pi^{-1}(D_4/-D_2)^{1/2} \tag{4.3}$$

by Rice (1944) and for summits as

$$D_{\text{sum}} = (6\pi\sqrt{3})^{-1}(D_3/-D_2) \tag{4.4}$$

by Nayak (1971).

The density of the peaks or summits is the number of peaks per unit length or summits per unit area, using the $(k+1)$-point definition of peak for $k = 2$ and summit for $k = 3$ and 4. The expected density of peaks or summits is given by the product of $\text{pr}(T_{k+1})$ and the density of ordinates, where T_{k+1} is the event $(S_1 > 0, \ldots, S_k > 0)$ and S_1 to S_k are the differences between the central ordinate and the k adjacent ordinates at a distance h.

The limiting behaviour of $\text{pr}(T_{k+1})$ as h tends to zero, the density of ordinates and the limit of the expected density of peaks (or summits) are given in table 1. The limits are given in terms of the limiting results for a continuous surface given by (4.3) and (4.4). It is seen that the density of peaks (when $k = 2$) converges to the continuous limit. This is not the case for summits (when $k = 3$ and 4). In both cases the density would be overestimated by 73% and 31% respectively.

The second parameter which will be considered is the average peak (or summit) height. The results are known for a continuous random Gaussian surface and were given for peaks as

$$E(Z\,|\,\text{continuous peak}) = (\pi/2)^{1/2} \cdot \eta \tag{4.5}$$

by Rice (1944) and Whitehouse and Phillips (1978) and for summits as

$$E(Z\,|\,\text{continuous summit}) = 4\pi^{-1/2} \cdot \eta$$
$$= 1.801(-\tfrac{1}{2}\pi)^{1/2}\eta \tag{4.6}$$

by Nayak (1971). So the average summit height is 80% higher than the average peak height.

Again the expected height of peaks (when $k = 2$) converges to the continuous limit for peaks on a profile. However, this is not the case for summits (when $k = 3$ and 4) as is seen in table 2. In both cases the expected summit height is underestimated by 13% for the 4-point case and by 6% for the 5-point case.

Because the conditional distribution of height given curvature is Gaussian with a mean which is a linear function of curvature, for all values of k, the expected summit curvature will converge in the same manner as the expected summit height (see equation (3.17) and the discussion following equation (3.18)).

Table 1. Expected summit (peak) density.

	k	$\mathrm{pr}(T_{k+1})$ Limiting behaviour as $h\to 0$	Density of ordinates	Expected density: Limit as $h\to 0$	Expected density: Limiting behaviour as $h\to\infty$
Three points	2	$\frac{1}{2\pi}\left(\frac{D_4}{-D_2}\right)^{1/2} h$	$\frac{1}{h}$	$\frac{1}{2\pi}\sqrt{\left(\frac{D_4}{-D_2}\right)} = D_{\mathrm{peak}}$	$\frac{1}{3h}=\frac{0.333}{h}$
Four points	3	$\frac{\sqrt{3}}{8\pi}\left(\frac{D_4}{-D_2}\right)h^2$	$\frac{4}{3\sqrt{3}h^2}=\frac{0.770}{h^2}$	$\sqrt{3}\cdot D_{\mathrm{sum}}=1.732D_{\mathrm{sum}}$	$\frac{1}{3\sqrt{3}h^2}=\frac{0.192}{h^2}$
Five points	4	$\frac{[\pi+2\sin^{-1}(\frac{1}{3})+4\sqrt{2}]}{24\pi^2}\times\left(\frac{D_4}{-D_2}\right)h^2$	$\frac{1}{h^2}$	$\frac{\sqrt{3}[\pi+2\sin^{-1}(\frac{1}{3})+4\sqrt{2}]}{4\pi}\times D_{\mathrm{sum}}=1.306D_{\mathrm{sum}}$	$\frac{1}{5h^2}=\frac{0.2}{h^2}$

To study the effect of the change of the sampling interval h on the digital measurements of an isotropic surface it is necessary to specify a model for the autocorrelation function of the surface. For the model to fit in with observed autocorrelation functions of surfaces it would be desirable to have a negative exponential function with a multiplicative periodic function. Whitehouse and Philips (1978) 'smoothed' the exponential-cosine function to give a function which was smooth at the origin. An alternative approach was used by Whitehouse and Phillips (1982) which replaced the negative exponential function by another function that is smooth at the origin but behaves like the negative exponential function for large t. Both of these correspond with the autocorrelation functions of many typical practical surfaces.

Table 2. Expected summit (peak) height $E(Z_0|T_{k+1})$.

	k	$h=0$	$h=\infty$
Three points	2	$\left(\frac{\pi}{2}\right)^{1/2}\eta = 0.555\frac{4}{\sqrt{\pi}}\eta$	$\frac{3}{2\sqrt{\pi}}=0.846$
Four points	3	$2\left(\frac{3}{\pi}\right)^{1/2}\eta=1.559\left[\left(\frac{\pi}{2}\right)^{1/2}\eta\right]=0.866\left[\frac{4}{\sqrt{\pi}}\eta\right]$	$\frac{3}{2\sqrt{\pi}}\left[1+\frac{2}{\pi}\sin^{-1}(\frac{1}{3})\right]=1.029$
Five points	4	$\frac{8\sqrt{(2\pi)}}{[\pi+2\sin^{-1}(\frac{1}{3})+4\sqrt{2}]}\eta=1.688\left[\left(\frac{\pi}{2}\right)^{1/2}\eta\right]=0.938\left[\frac{4}{\sqrt{\pi}}\eta\right]$	$\frac{3}{2\sqrt{\pi}}\left[\frac{10}{3}-\frac{5}{\pi}\cos^{-1}(\frac{1}{3})\right]=1.163$

This autocorrelation function is given by

$$\rho(h)=\operatorname{sech}[\tfrac{1}{2}\pi h A(\theta)]\cos[2\pi\theta h A(\theta)], \tag{4.7}$$

where

$$A(\theta)=\operatorname{sech}(2\pi\theta)+2\sum_{r=0}^{\infty}\frac{(-1)^r\theta}{2\pi\{\theta^2+[\frac{1}{4}(2r+1)]^2\}\sinh[\pi(2r+1)/8\theta]}, \tag{4.8}$$

where h is the sampling interval. The values of θ used are 0 and $\frac{1}{2}$. For this autocorrelation function

$$D_2=-(\tfrac{1}{2}\pi)^2[1+(4\theta)^2][A_2(\theta)]^2 \tag{4.9}$$

and

$$D_4=(\tfrac{1}{2}\pi)^4[5+6(4\theta)^2+(4\theta)^4][A_2(\theta)]^4. \tag{4.10}$$

The expected density of summits is given in figures 2 and 3 and the expected height of peaks or summits is given in figures 4 and 5 for the autocorrelation function for $\theta=0$ and $\frac{1}{2}$.

The expected 4-point and 5-point density of summits differ little as the sampling interval h exceeds one correlation length. For smaller sampling intervals the 4-point expected density of summits exceeds that for the 5-point expectation.

In contrast to the expected density the 4-point and 5-point expected height of summits differ fairly consistently as the sampling interval varies. The limits of the 5-point expectation are greater than the 4-point expectation as the sampling interval approaches both zero and infinity.

5. Discussion

The technique of using discrete measurements has application in fields where it is expensive or time consuming to obtain large amounts of data. The reason for this

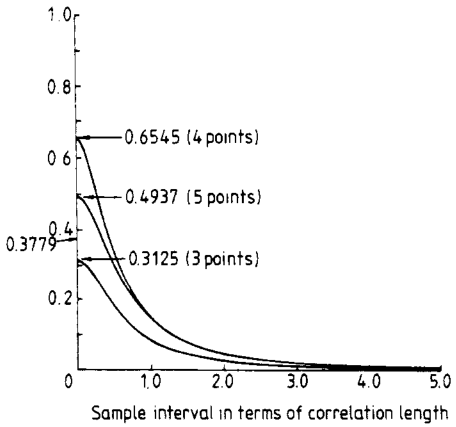

Figure 2. The variation of the expected density of 3-point peaks and 4- and 5-point summits with a spacing h between ordinates. The autocorrelation function is shown with $\theta=0$.

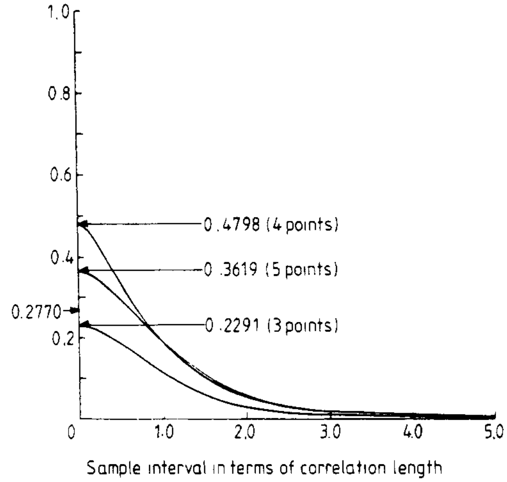

Figure 3. The variation of the expected density of 3-point peaks and 4- and 5-point summits with a spacing h between ordinates. The autocorrelation function is shown with $\theta = \frac{1}{2}$.

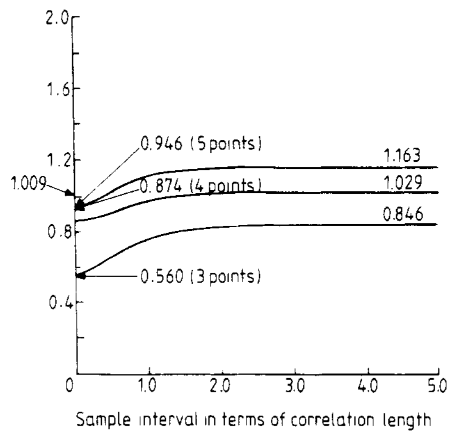

Figure 4. The variation of the expected height of 3-point peaks and 4- and 5-point summits with a spacing h between ordinates. The autocorrelation function is shown with $\theta = 0$.

paper was to try and see whether taking measurements using a non-conventional sampling scheme would produce any advantages to outweigh the disadvantage of complexity. The advantages considered were less information to collect, easier analytical derivation of theoretical results and simpler numerical methods.

The sampling schemes that were considered all had the property that the information could be collected by sampling along parallel straight lines with a fixed sampling interval. (It might be necessary however to have a variable starting point, though this would follow a regular pattern.) This ensured that if a measurement (ordinate) was chosen when using a particular scheme it would always have the same number of adjacent ordinates at a distance h (the chosen sampling interval), provided the chosen ordinate is not on the boundary.

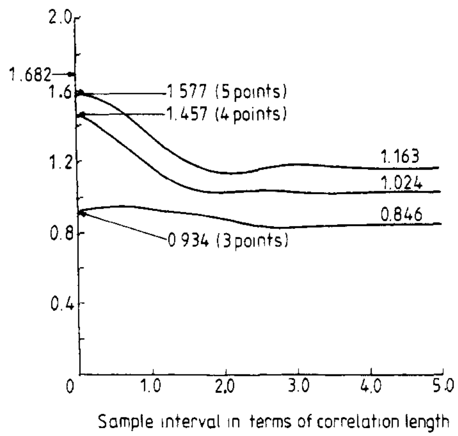

Figure 5. The variation of the expected height of 3-point peaks and 4- and 5-point summits with a spacing h between ordinates. The autocorrelation function is shown with $\theta = \frac{1}{2}$.

From the point of view of simplicity of sampling mechanism the square grid ($k = 4$) in the tetragonal case is the best. In this case the spacing between the lines is constant and equal to the sampling interval h along the line. Also the starting points for the sampling all lie along a straight line. However other schemes do have advantages to offset their complexity.

The trigonal ($k = 3$) case has the advantage that measurements of slope can be taken in three directions as opposed to two for the tetragonal ($k = 4$) case. Though the theoretical results have been restricted to the consideration of *isotropic* surfaces it may still be of *practical* value to be able to check the assumption of isotropicity in more than two directions.

The trigonal ($k = 3$) case can be obtained by an alternative sampling method but this involves alternating the sampling interval from h to $2h$. This alternative method is equivalent to rotating the grid through $\pi/6$.

From the point of view of collecting digital information the trigonal ($k = 3$) case is preferable as 'less' information is collected. The density of ordinates is $(4/3\sqrt{3})h^2$ ($= 0.77/h^2$) compared with $1/h^2$ for the square grid ($k = 4$). So in the same area 23% less ordinates would be needed. The advantage of this would need to be weighed against the disadvantages.

Another advantage of the trigonal ($k = 3$) case is that fewer ordinates are used when defining properties of the extremities. To check the definition of a 4-point summit only three conditions have to be obeyed, as opposed to four conditions for the 5-point summit. It should also be noted that some properties of the discrete defined random variables, such as the limiting value of $(k+1)$-point summit (or peak) height as the sampling interval tends to infinity, are simply a function of the numerical definition and are independent of the surface being measured.

Any discrete measurement of a surface must lose information compared with a complete 'map' of the surface. This is inevitable! However, ideally, any discrete measurement should produce results which converge to the results for the continuous surface as the sampling interval h tends to zero.

For sampling along a straight line ($k = 2$) it is seen that the discrete results do converge to those for the continuous profile. They do not, however, converge to the results of the two-dimensional surface. For example $D_{\text{peak}}^2 = 0.83\ D_{\text{sum}}$, so that assuming independent measurements at right angles would produce a limit which is 17% too small.

For two-dimensional measurements when sampling with $k=3$ or 4, the limiting results for expected summit density and expected summit height do not converge to the results for the continuous surface. In the case of expected summit density the limit is 73% too large for $k=3$ and 31% too large for $k=4$. Again for expected summit height the case $k=3$ is worse than for $k=4$ but the differences are not so large. This suggests that some surface parameters may be estimated by discrete methods fairly well but others may not. For the case of average profile slope all three sampling schemes agree (for $k=2$, 3 and 4) but this is, of course, an essentially one-dimensional parameter.

In order to consider the merits of sampling schemes it is necessary to study their theoretical properties. By doing this it is possible to obtain new insights into the general problem. This is possible only by using models which lead to tractable mathematics. The three sampling schemes with $k=2$, $k=3$ and $k=4$ considered in this paper have been chosen because they have a common property which enables them to be investigated using analytical results previously obtained in theoretical statistics. Using the trigonal ($k=3$) symmetry case leads to a simpler mathematical model than for the tetragonal ($k=4$) symmetry case, as this reduces the dimension by one. However, taken as a whole it may be that a hexagonal sampling plan where $k=6$ offers the maximum benefit in terms of the three criteria mentioned above. One message which has emerged from this exercise is that the conventional grid pattern method of sampling is not necessarily the best. The implications of this in general pattern recognition and image analysis scanning systems are likely to be significant.

The work presented here has concerned primarily the effect of sampling patterns on the values of parameters obtained from measured surfaces. It has not therefore been aimed at investigating the actual nature of surfaces in general. Well behaved correlation functions have been assumed and certain specific examples have been used to give the researcher an idea of the value of parameter changes that might be expected to occur on typical measured surfaces. This has been justified by the fact that to some extent all instruments used for obtaining the data have a short wavelength filter incorporated whether it be a stylus or a light spot which tends to force the correlation function at the origin to be smooth. However, there can be no denying that the very nature of random manufacture encourages the presence of exponential and other misbehaved characteristics in the correlation function. The effect of sampling patterns on such fractal (multiscale) surfaces will be the subject of further work.

References

Cheng M C 1969 *Ann. Math. Stat.* **400** 152–61
David F 1953 *Biometrika* **40** 458–9
Greenwood J A 1984 *Proc. R. Soc.* A **393** 133–57
Greenwood J A and Williamson J B P 1966 *Proc. R. Soc.* A **295** 300–19
Hamed M S, Whitehouse D J and Buttery T C 1978 *Ann. CIRP* **27** 449–504
Jordan D L, Hollins R C and Jakeman E 1983 *Appl. Phys.* B **3** 179–86
Nayak P R 1971 *Trans. ASME J. Lubr. Technol.* **93** 398–407
Phillips M J 1984 *Math. Meth. Appl. Sci.* **6** 248–6
Rice S O 1945 *Bell Sys. Tech. J.* **24** 46–156

Welford W T 1977 *Opt. Quantum Electronics* **9** 269-87
Whitehouse D J and Archard J F 1970 *Proc. R. Soc.* A **316** 97-121
Whitehouse D J and Phillips M J 1978 *Phil. Trans. R. Soc.* A **290** 267-98
—— 1982 *Phil. Trans. R. Soc.* A **305** 441-68