19

Experiments with gridded survey data

Mike Fletcher
Department of Mathematics, North Staffordshire Polytechnic

Dick Spicer
Research Centre for Computer Archaeology, North Staffordshire Polytechnic

19.1 Introduction

It is common nowadays to survey archaeological sites, whether topographically or electromagnetically, on a regular grid. With modern EDM machinery and the mathematical speed of today's micros, however, it is possible to use other methods of survey which are less labour-intensive.

The authors have been looking at some experimental techniques for examining and manipulating gridded and non-gridded survey data. Most of the programs which we have used for this work have been written for use on many different machines—indeed they will run on most microcomputers. The graphics programs used for output display are again written for portability, though the different capabilities of the many machines now on the market make it difficult to produce truly universal programs.

In this paper we offer one technique which we have found useful in our work of manipulating gridded data. It is a method of describing objectively the 'quality' of the survey, and is, in a sense, a measure of the way the survey fits the site: we call it the Gradient Coefficient (see section 19.2).

We arrive at two numbers by measuring the gradient in both directions (that is, x and y) at each datum point of a grid, and producing the mean of the gradients and the standard deviation of the gradients. The mean of these gradients simply tells us whether or not, and by how much, our site has a overall slope. This number will vary over a wide range depending both on the amount of slope and the size of the units used for measurement. The Gradient Coefficient, however, is a number limited in range from zero to one, and it expresses the overall amount of change between each datum and its neighbours. It is, to put it simplistically, a measure of the 'bumpiness' of the data. This is not to say that we are expressing the bumpiness of the site: if a site is surveyed well according to our criteria, then its Gradient Coefficient will be good (that is, small) and the data will, we claim, describe the site adequately.

The Gradient Coefficient can be described simply as the ratio of the variance of the gradients to the variance of the heights. Its derivation is given as follows:

Consider data values X_{ij} on a regular grid $n \times m$, with $1 \leq i \leq n$ and $1 \leq j \leq m$.

The variance of the X's can be found using the formula

$$\mathrm{Var}(X) = \frac{\sum X_{ij}{}^2}{nm} - \left(\frac{\sum X_{ij}}{nm}\right)^2$$

Now for each internal point $(2 \leq i \leq n, 2 \leq j \leq m)$ define two gradients:

$$G_{i,j,1} = X_{i,j} - X_{i-1,j}$$

$$G_{i,j,2} = X_{i,j} - X_{i,j-1}$$

The mean gradient, $\overline{G}$, is defined by:

$$\overline{G} = \sum \frac{(G_{i,j,1} + G_{i,j,2})}{2 * (n-1)(m-1)}$$

and the variance is defined by:

$$\mathrm{Var}(G) = \sum \frac{(G_{i,j,1}{}^2 + G_{i,j,2}{}^2)}{2 * (n-1)(m-1)} - \overline{G}^2$$

It can be shown that

$$0 \leq \mathrm{Var}(G) \leq 4\mathrm{Var}(X)$$

hence define the Gradient Coefficient GC by:

$$GC = \frac{\mathrm{Var}(G)}{4 * Var(X)}$$

so that

$$O \leq GC \leq 1$$

19.2 Use of the Gradient Coefficient

Let us now demonstrate the use of the coefficient by examining some simple artificially-generated geometric shapes. The first, most obvious, example is the flat surface, giving a GC of zero, and a mean gradient also of zero. An inclined plane would, of course, give a GC again of zero, but the mean gradient would give away the fact that it is a slope.

The most bumpy surface imaginable, we think, is one where the gradient direction reverses at each datum point. This is easily produced with a checkerboard pattern of, say, ones and zeros for our height data. The visual effect is a difficult one to show as a wire diagram (Fig. 19.1), for it produces regularity in all directions. When contoured, this data reveals a series of concentric circles, which alternate in vertical direction (Fig. 19.2). The GC of this data is one, and represents our worst possible case.

One would be very surprised to find such a value in real data; a better standard with which to make comparisons is a random data set, or 'white noise', which gives a GC of around 0.7 (Figs. 19.3 & 19.4).

As a further example, we have concocted a simple shape, which is an arbitrary surface seven points square (Figs. 19.5 & 19.6). It rises fairly gently at the middle, and gives a GC of 0.54 and a mean gradient of 0.24. We would describe this as a data set which is fairly bumpy, with a

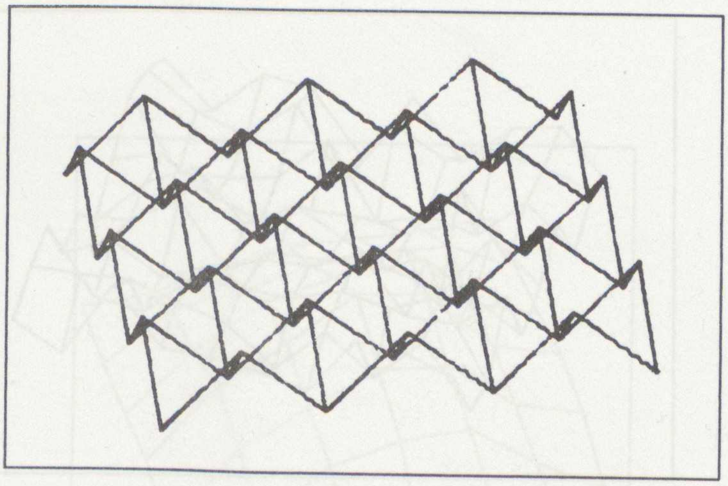

Fig. 19.1: 'Egg box': Wire diagram

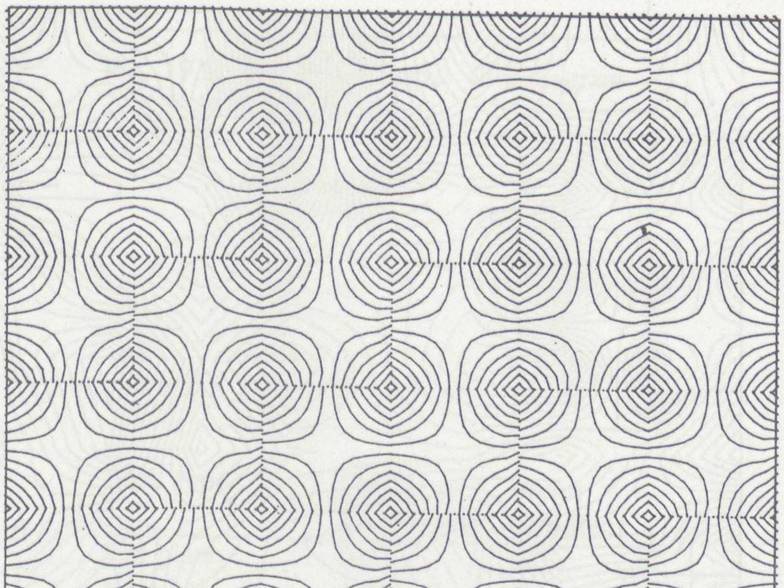

Fig. 19.2: 'Egg box': Contour plan

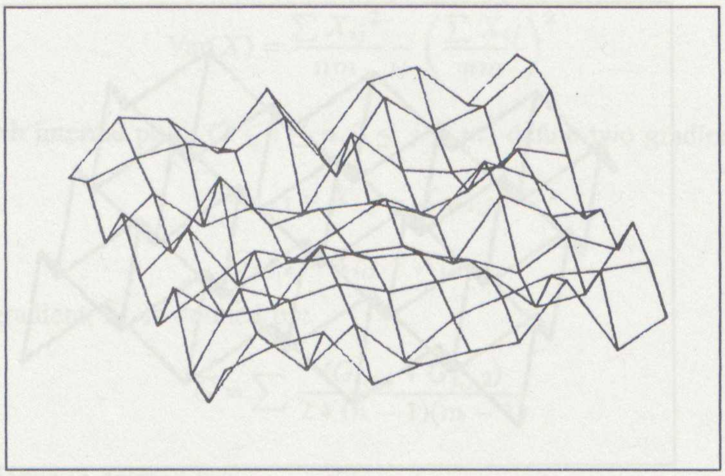

Fig. 19.3: Random noise: wire diagram

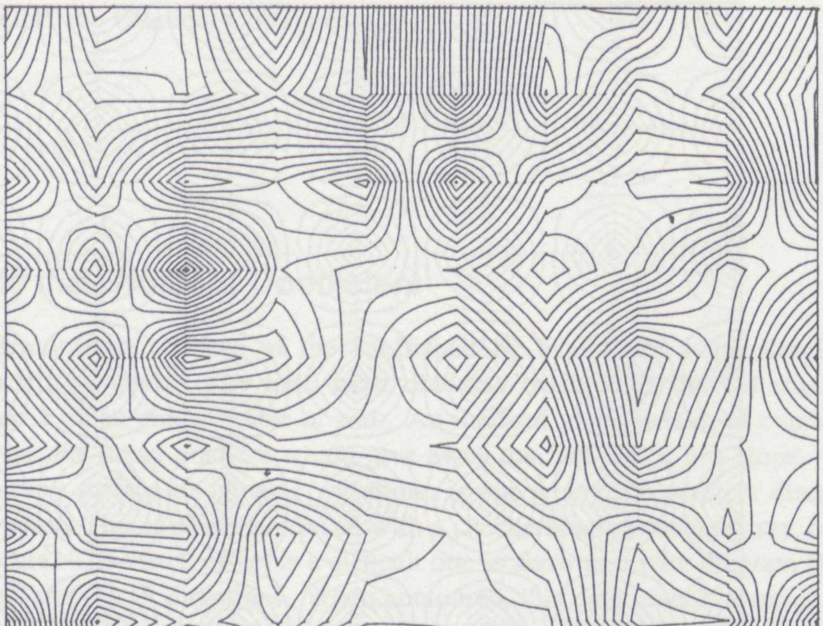

Fig. 19.4: Random noise: contour plan

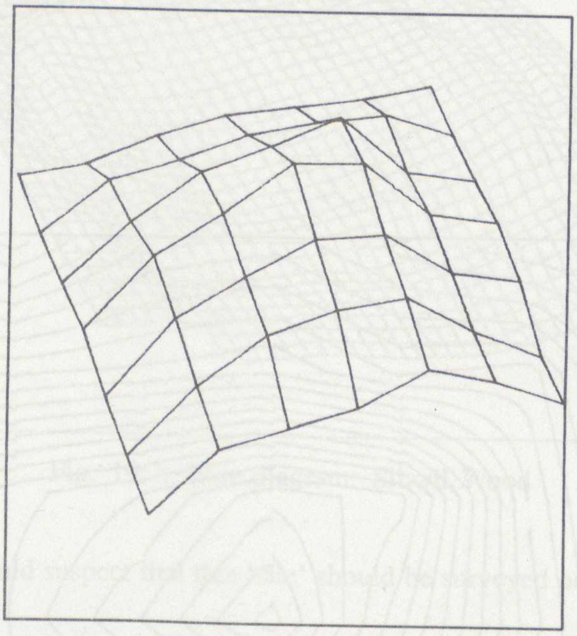

Fig. 19.5: Concocted shape: wire diagram

Data type	Gradient Coefficient	Mean Gradient
Flat plane	0.0	0.0
'Egg box'	1.0	0.0
White noise	0.71	2.4
Concocted shape	0.54	0.24

Table 19.1: Gradient Coefficient for simple examples

Fig. 19.6: Concocted shape: contour plan

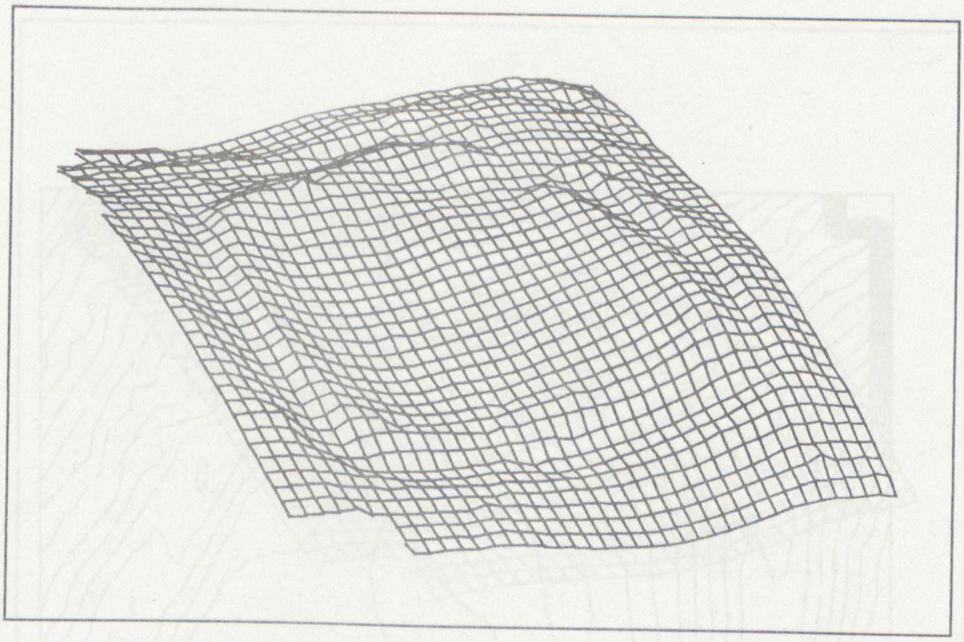

Fig. 19.7: Wire diagram: Sibwll Wood

slight slope, and we would suspect that this 'site' should be surveyed on a finer grid to produce a smaller GC.

Let us turn to real archaeological data. Data from our first site, Sibwll Wood, was kindly provided by Clwyd-Powys Archaeological Trust, and we deliberately accepted no prior information concerning the site, for reasons described in another paper (Spicer, this volume). It appears to be an enclosure on a fairly steep slope (Fig. 19.7). Its *GC* (of 0.06) tells us that it is not too variable in gradient, and we might provisionally say that the survey method was adequate to describe the site's features.

Using a simple and fairly rapid contour program (Fig. 19.8) gives us little information about the enclosure, and the obvious next step was to remove the overall slope to leave only the 'high frequency' information: namely the banks and/or ditches, together with noise. We subjected the data to a simple local averaging low-pass filter, which gave us a picture of just the slope (Figs. 19.9 & 19.10)—one might describe this flippantly as the 'pre-enclosure' surface. Subtraction of the two data sets produces a data set of residuals, and we see a better picture of the enclosure, this time on a flattish base (Figs. 19.11 & 19.12). Applying our test to these data (Table 19.2) shows that we now have only a slight slope, with a mean gradient value of -4.05, and the *GC* figure of 0.06 for both the original and filtered data confirms our belief that the survey fits the site well. The site would have needed many changes of survey station down the hill-side, each set-up contributing errors—or 'noise'—to the survey. This is amplified when the slope is removed from the data. Even so, the 'noise value' of the residuals (0.27) is still remarkably low.

So far we have demonstrated this technique as it were retrospectively: that is, we have applied it to data after the whole site has been surveyed. If the value of *GC* were found to be unsatisfactory, then all we could do is learn from our mistakes in time for our next survey. But let us see if it is possible to use our method to predict results. We turn back again to artificial data—a mathematically-generated cone, sitting, or rather floating, on its side in a puddle. A

Fig. 19.8: Sibwll Wood: contour plan

Object/site name	Gradient coefficient	Mean gradient
Sibwll Wood	0.06	−28.29
Sibwll filtered	0.06	−22.20
Sibwll residual	0.27	−4.05

Table 19.2: GC for data from Sibwll Wood

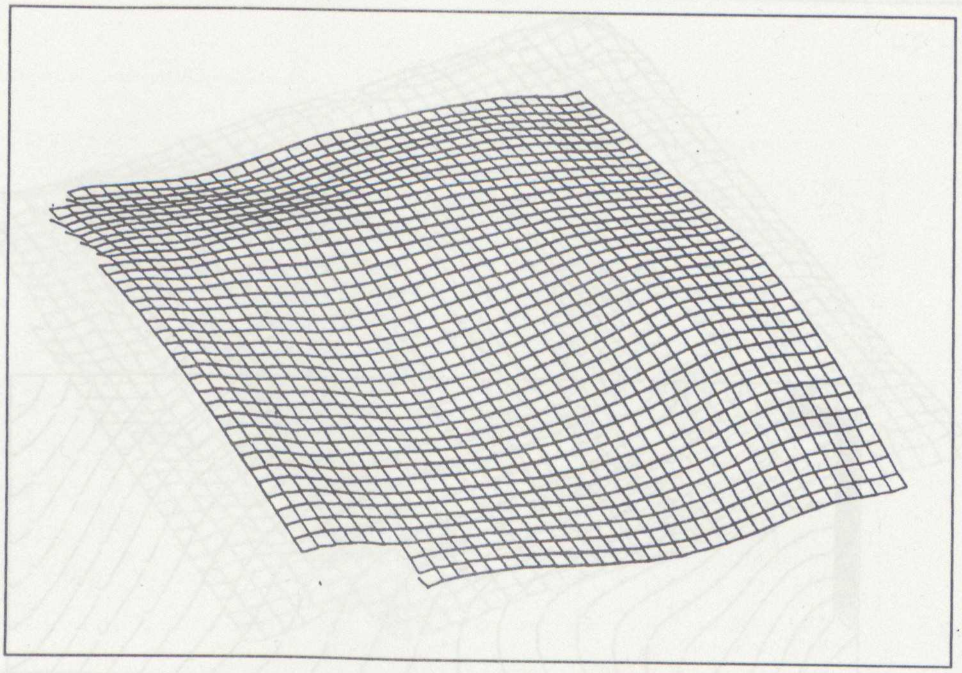

Fig. 19.9: Sibwll Wood: wire diagram filtered

Object/site name	Gradient coefficient	Mean gradient
Cone perfect	0.08	0.0
Cone random 1	0.67	0.0
Cone random 2	0.39	0.0
Cone random 3	0.24	0.0

Table 19.3: GC for cone data

control data set was derived on a grid 25 by 50 (Figs 19.13 & 19.14).

For this analysis we effectively sampled an infinitely large data set randomly—not on a grid—by mathematically calculating the height value, using different sample quantities. They are illustrated as contours in Figs. 19.15 to 19.17 and as surfaces in Figs. 19.18–19.20. Cone 1 was derived from 100 random points, cone 2 from 200, and cone 3 from 800 points. We reconstituted the shape by recreating a 10 by 20 grid from the random points using a simple plane-fitting method. The results of deriving their Gradient Coefficient are shown in Table 19.3. When we take a look at the diagrams, we see how unlike the original most cases are. Indeed, the diagram and the *GC* figure for the first random sample reveal that it is much closer to pure noise than to our cone. Provisional results from this recreation process indicate a rule of thumb that the number of sampled points must at least equal the number of grid squares required for the final result in order to get any recognisable shape. Work is currently continuing to pursue this investigation further.

Our formula for the Gradient Coefficient is so easily applied by a prospective surveyor that one might use it to judge a number of small sample surveys taken as a preliminary investigation in order to discover what grid interval most suits the site. So we are in a position now to examine a 'what if?' scenario, by taking a look at a real archaeological site—Symon's Castle,

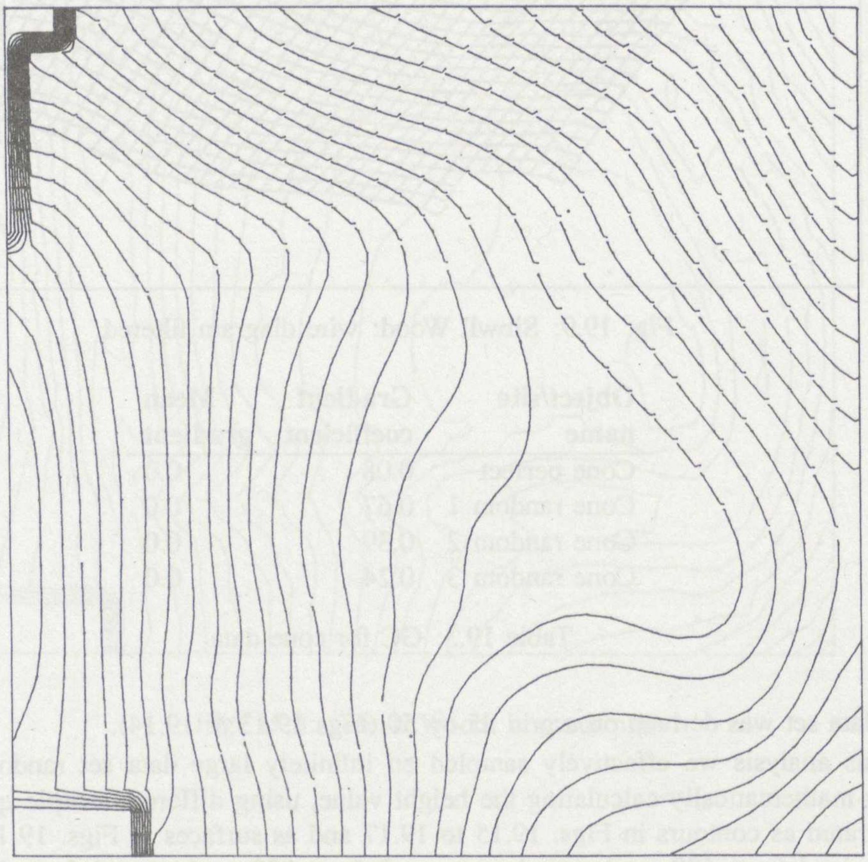

Fig. 19.10: Sibwll Wood filtered: contour plan

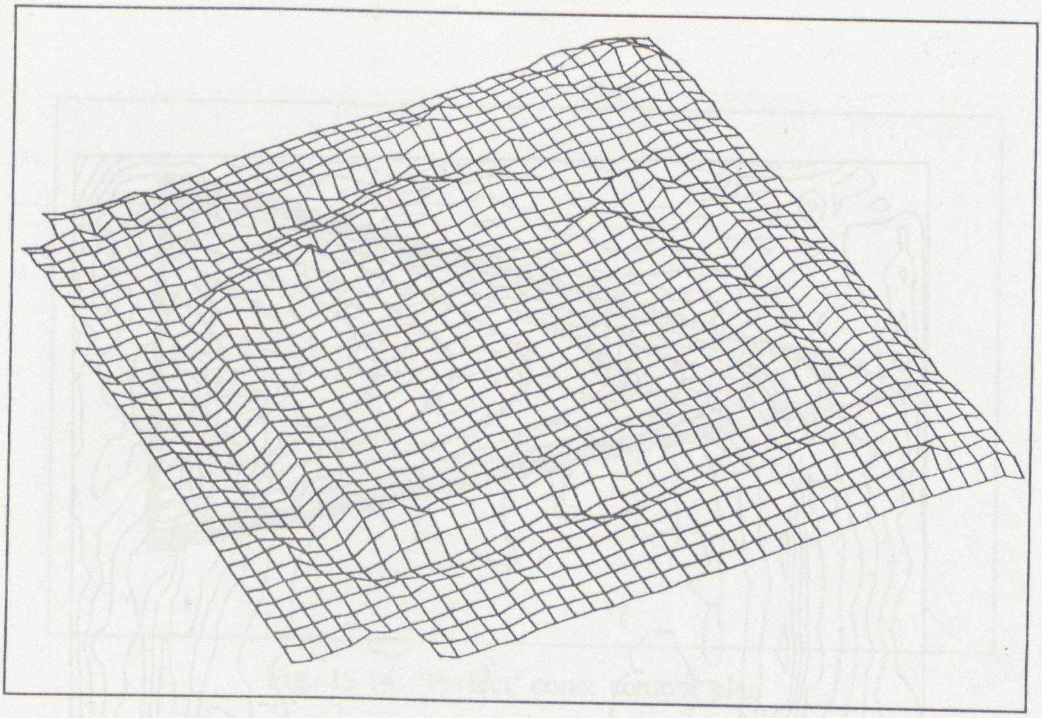

Fig. 19.11: Sibwll Wood residuals: wire diagram

Object/site name	Gradient coefficient	Mean gradient
Symon's Castle	0.16	8.06
Symon's filtered	0.16	8.03
Symon's residual	0.63	-0.04

Table 19.4: GC for cone data

Powys (the data for which has been provided by Jeremy Huggett and Chris Arnold). We find it has a GC of 0.16, a low value. How can we test to see if this reflects the quality of the survey?

If the data is smoothed, the new GC is still 0.16. Fig. 19.21 shows the site, using panels lit as if from a low sun, and we observe the shape of the motte clearly, with more than a hint of the raised defence encircling the flat top. Large features like the ditch and the bailey are clearly shown, but so are smaller details, like the low rectangular earthwork system in the next field, noted on the ground at the time of the survey, but shown only as a complete rectangle by this display.

When we remove the smoothed surface, and leave the residuals, the small features are still clearly visible (Fig. 19.22), though the now flat surface has acquired extra 'noise', a fact borne out by a considerable rise in the Gradient Coefficient (Table 19.4). We would expect a large-scale site survey of this kind to contain errors. There are a number of trees, and an awkward and dangerous quarry slope to contend with, as well as the normal problems of maintaining level readings over a large height difference. Noise of this kind superimposed on a steep slope would not be revealed by the Gradient Coefficient, hence the need to perform residual filtering.

Supposing, now, that this survey, seventy or so metres square, and representing over 5000 readings (and about 250 surveyor-hours), had been taken at two-metre intervals instead—

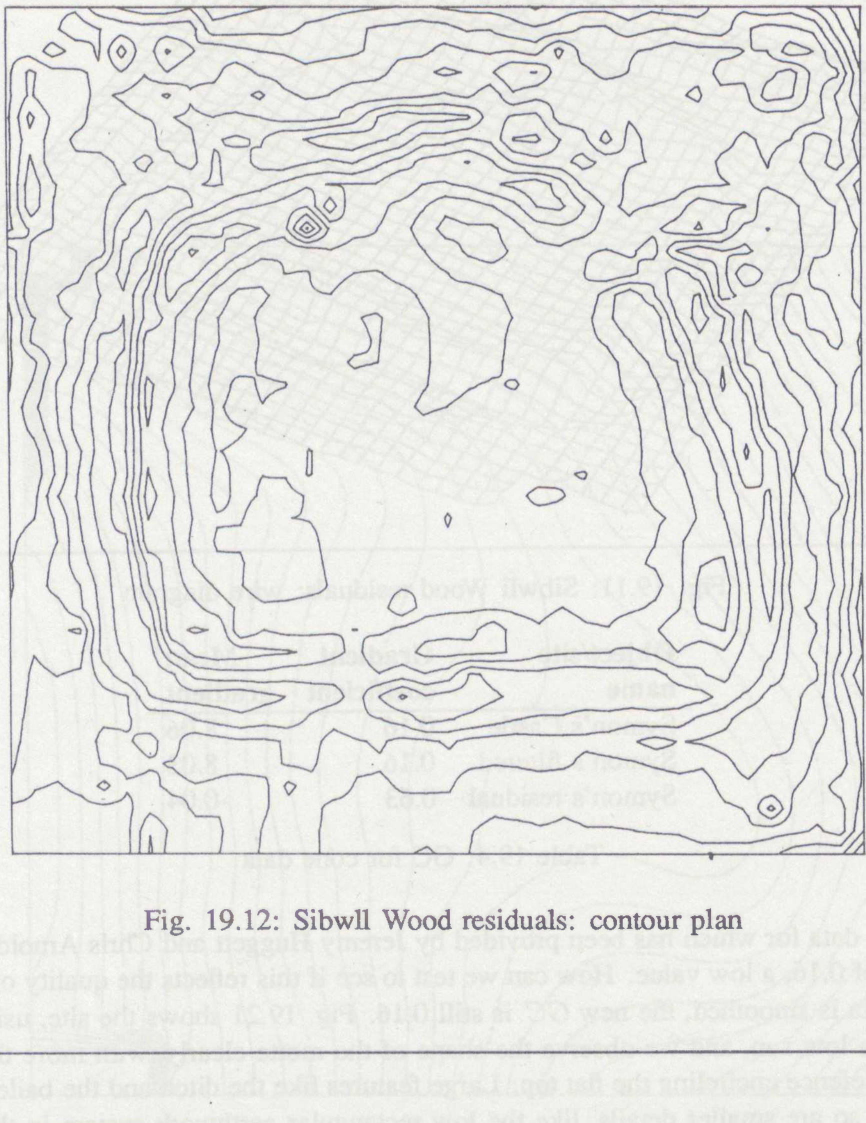

Fig. 19.12: Sibwll Wood residuals: contour plan

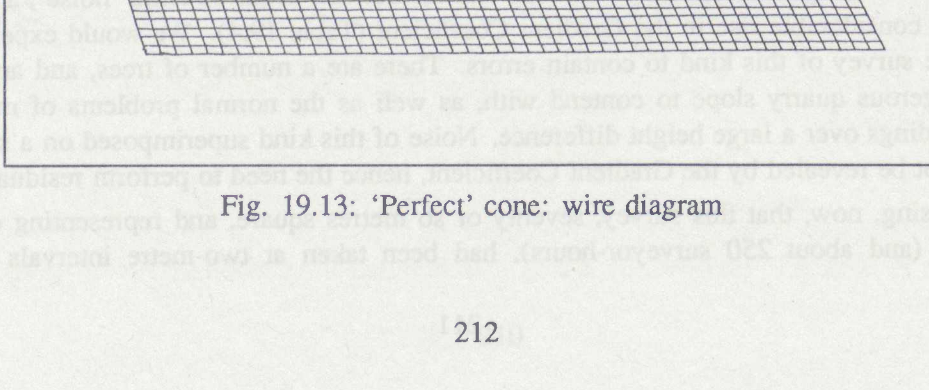

Fig. 19.13: 'Perfect' cone: wire diagram

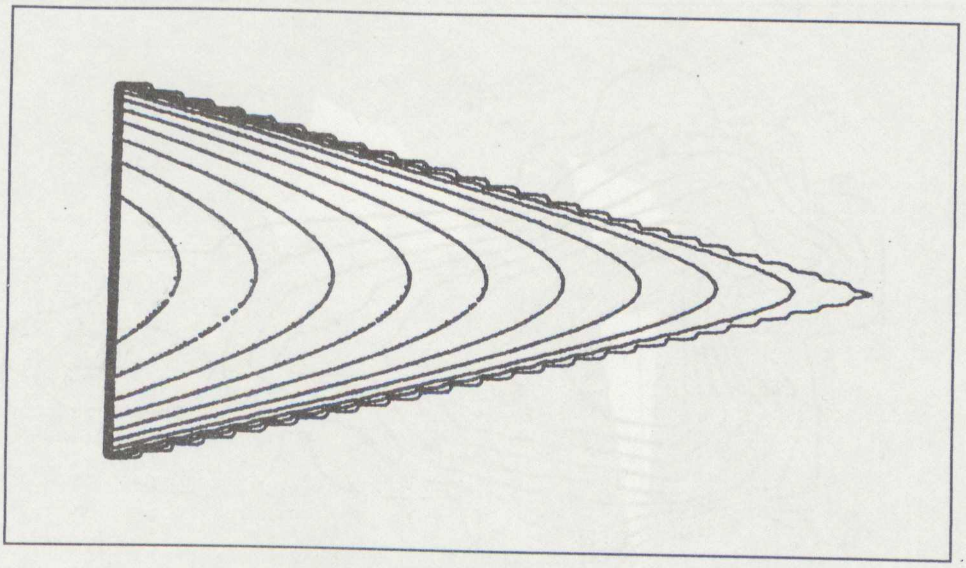

Fig. 19.14: 'Perfect' cone: contour plan

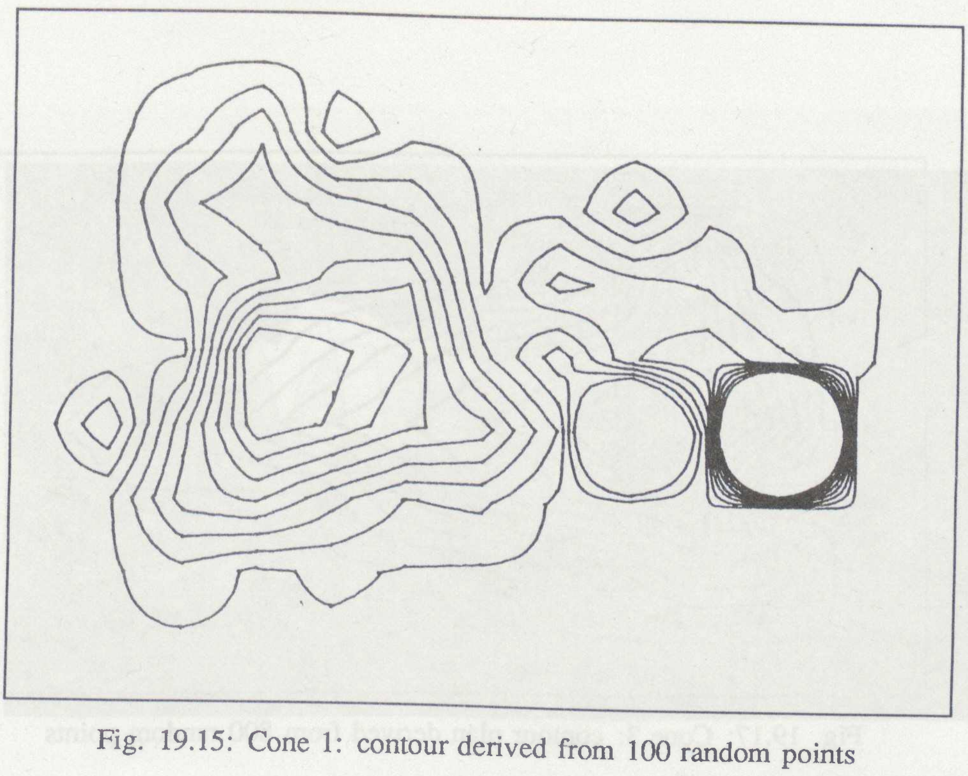

Fig. 19.15: Cone 1: contour derived from 100 random points

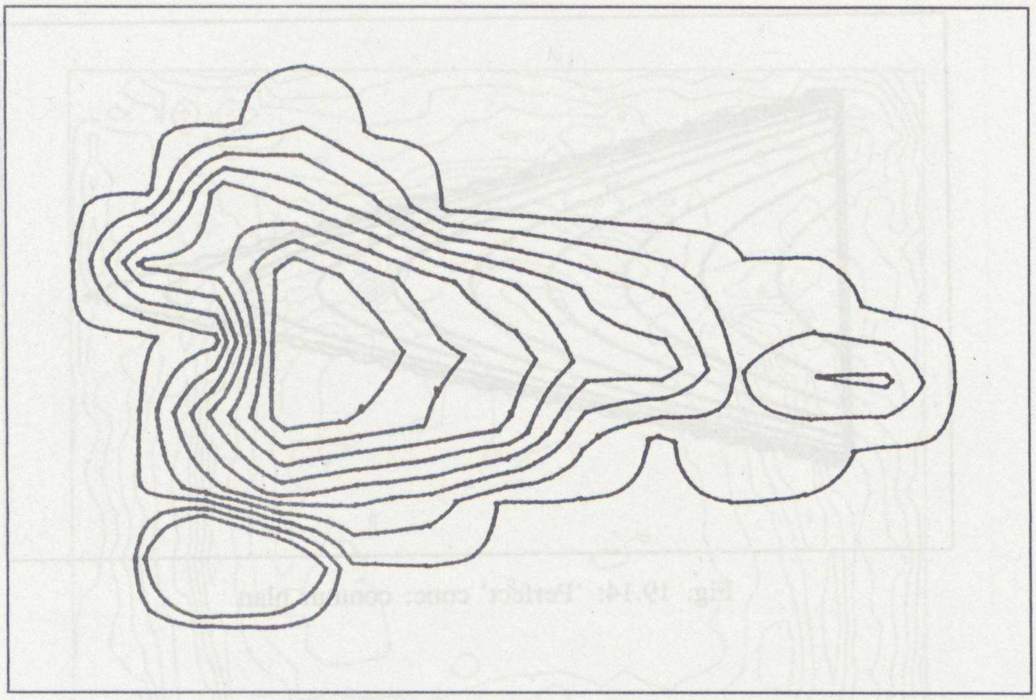

Fig. 19.16: Cone 2: contour plan derived from 200 random points

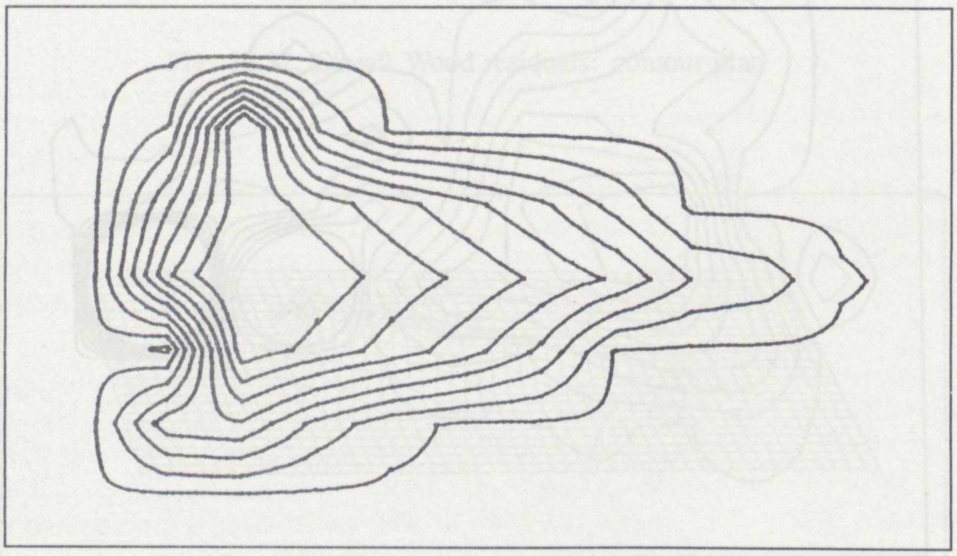

Fig. 19.17: Cone 3: contour plan derived from 800 random points

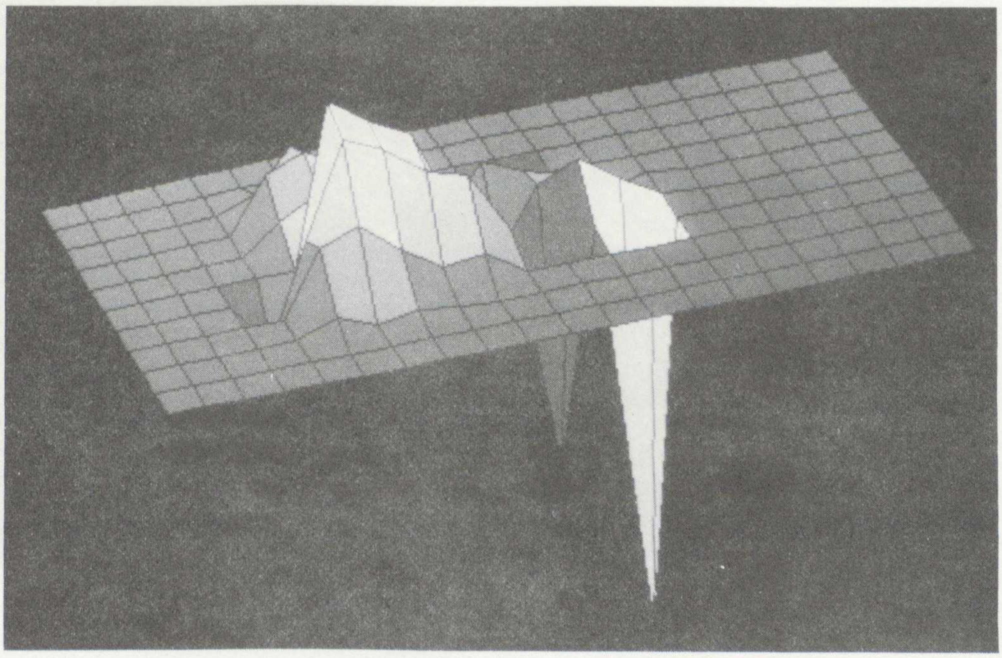

Fig. 19.18: Cone 1: panelled surface derived from 100 random points

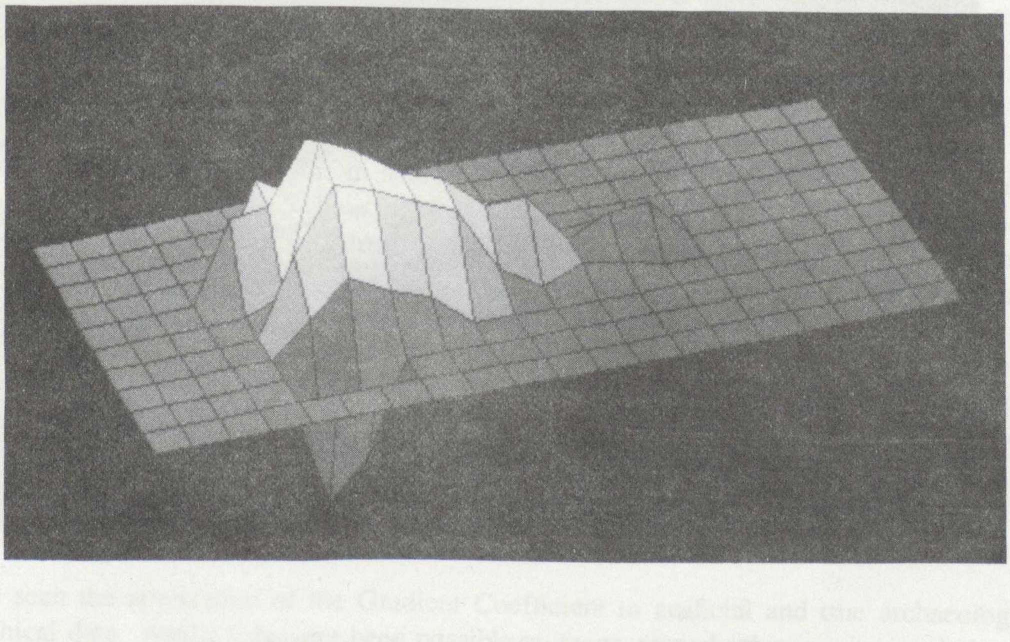

Fig. 19.19: Cone 2: panelled surface derived from 200 random points

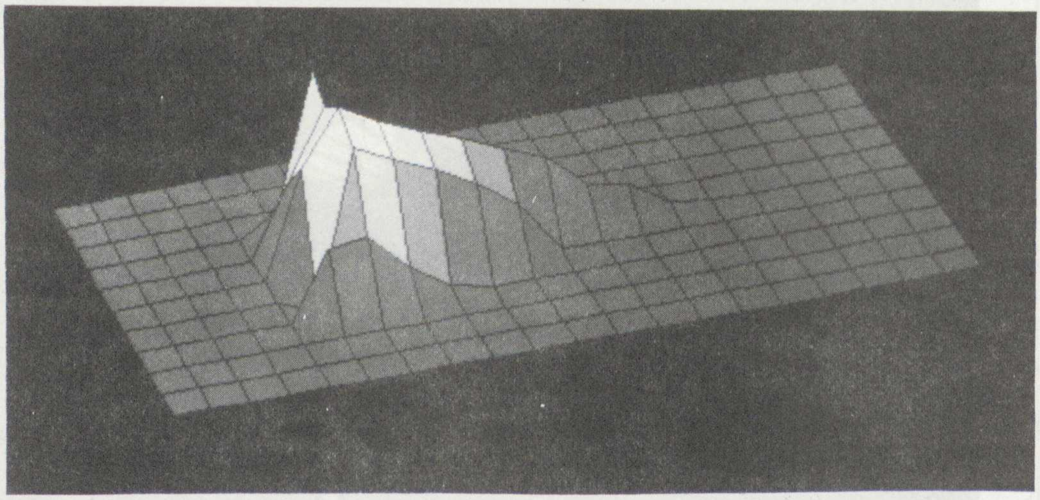

Fig. 19.20: Cone 3: panelled surface derived from 800 random points

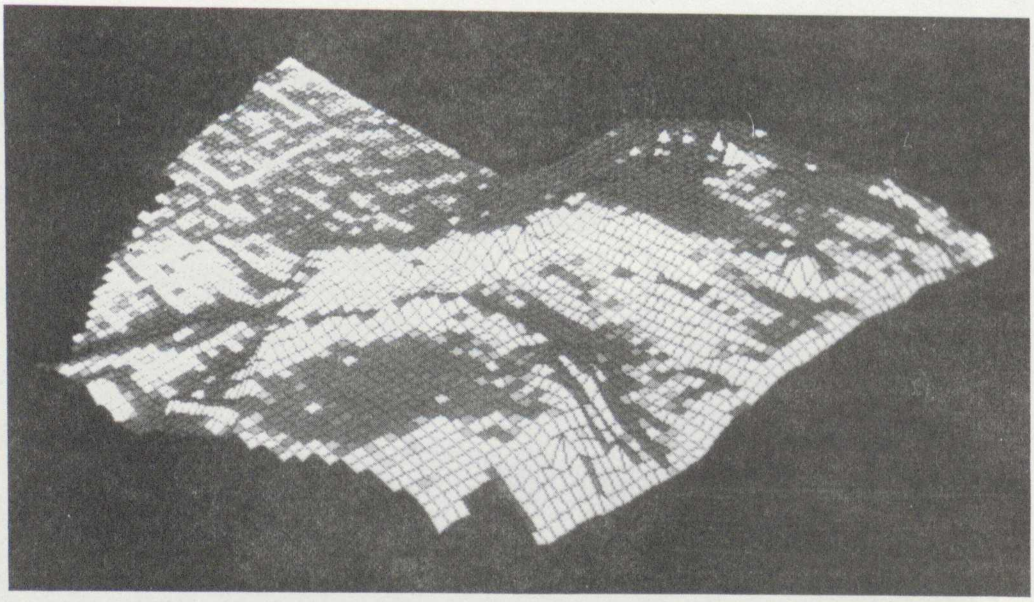

Fig. 19.21: Symon's Castle: lit panelled surface

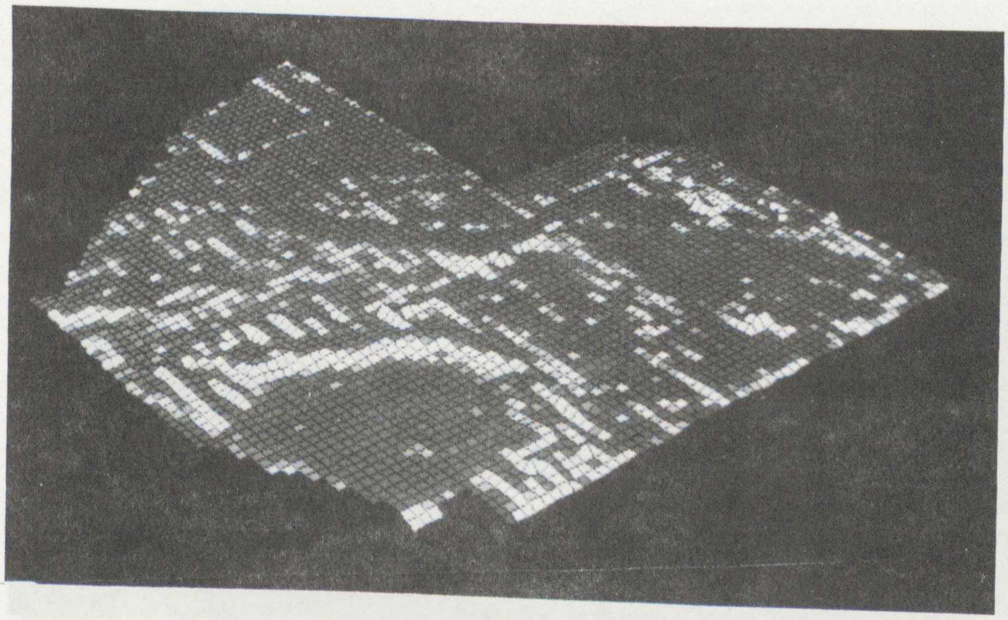

Fig. 19.22: Symon's Castle residuals: lit panelled surface

Object/site name	Gradient coefficient	Mean gradient
Symon's div 2	0.23	8.06
Symon's div 3	0.28	14.89
Symon's div 5	0.35	20.16
Symon's div 10	0.55	39.37

Table 19.5: GC for sampled data from Symon's Castle

thus quartering the time and effort in surveying. How well can we now see these features? Figs. 19.23–19.26 and Table 19.5 show the results of sampling the data at the following fractions: halves, thirds, fifths and tenths. Already at two-metre intervals the *GC* has risen, though the small features are still visible, indicating perhaps that two-metres would be the absolute maximum interval to use. At three-metre intervals the rectangular feature is unrecognisable (though some activity is indicated by different-coloured panels in the vicinity). Note how the mean gradient as well as the Gradient Coefficient rises progressively as we reduce the resolution of the data. What we are seeing here is the steady removal of the ditch, and the appearance of the overall slope between motte, bailey, and surrounding fields.

19.3 Discussion

We have seen the application of the Gradient Coefficient to artificial and true archaeological topographical data. Whilst it has not been possible so far to give absolute recommendations as to what constitutes a good or bad figure, we feel that even as a relative measure it is valuable

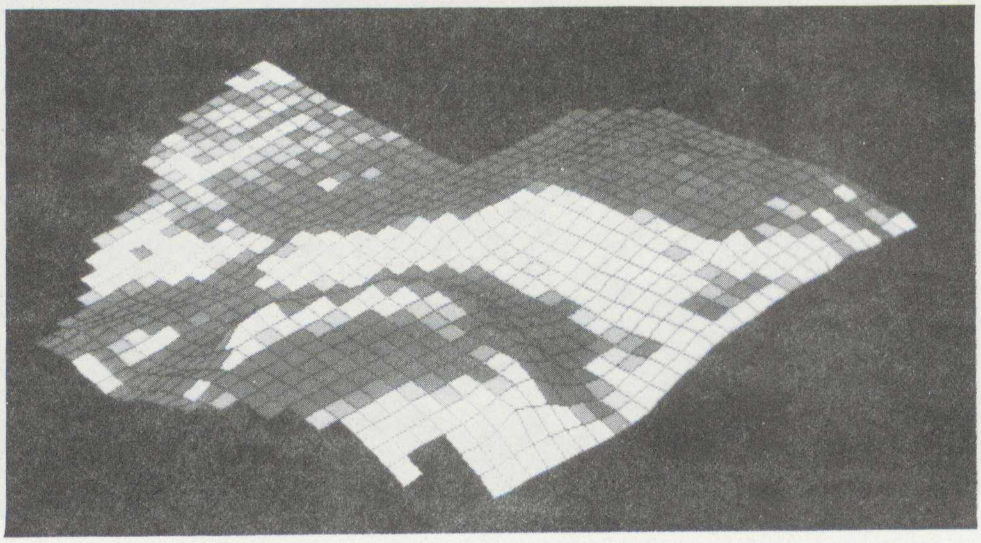

Fig. 19.23: Symon's Castle: lit panelled surface—2 metre grid

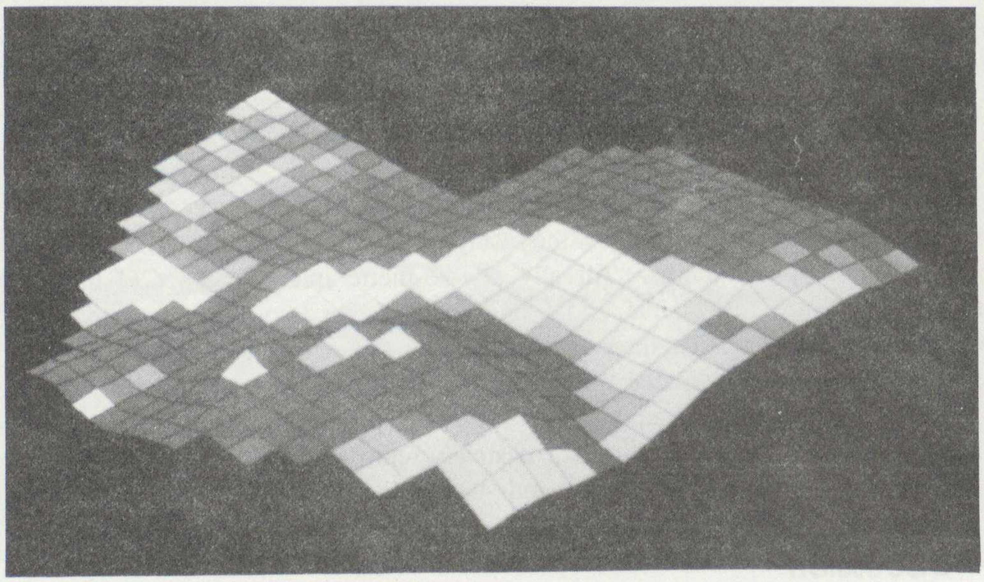

Fig. 19.24: Symon's Castle: lit panelled surface—3 metre grid

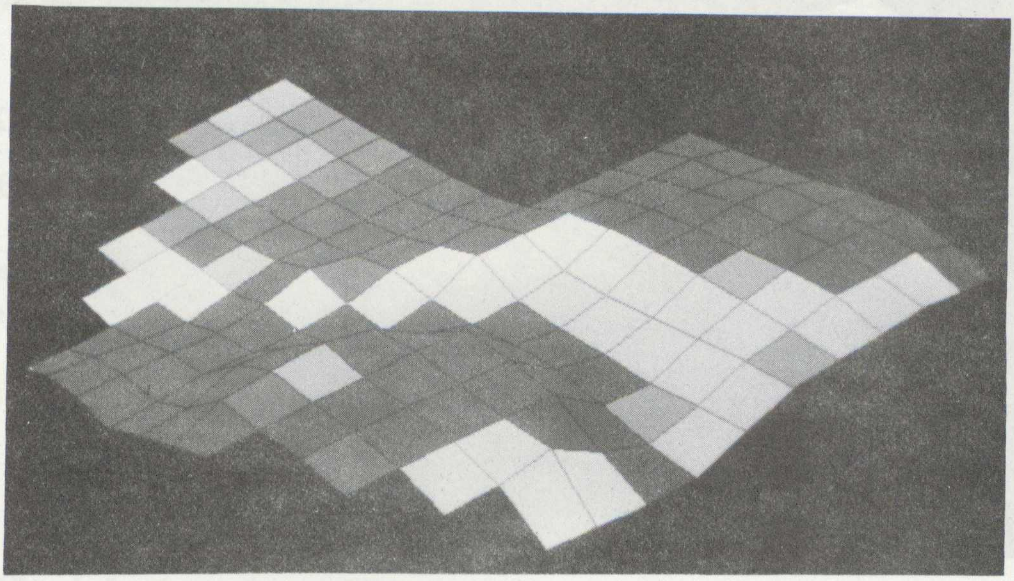

Fig. 19.25: Symon's Castle: lit panelled surface—5 metre grid

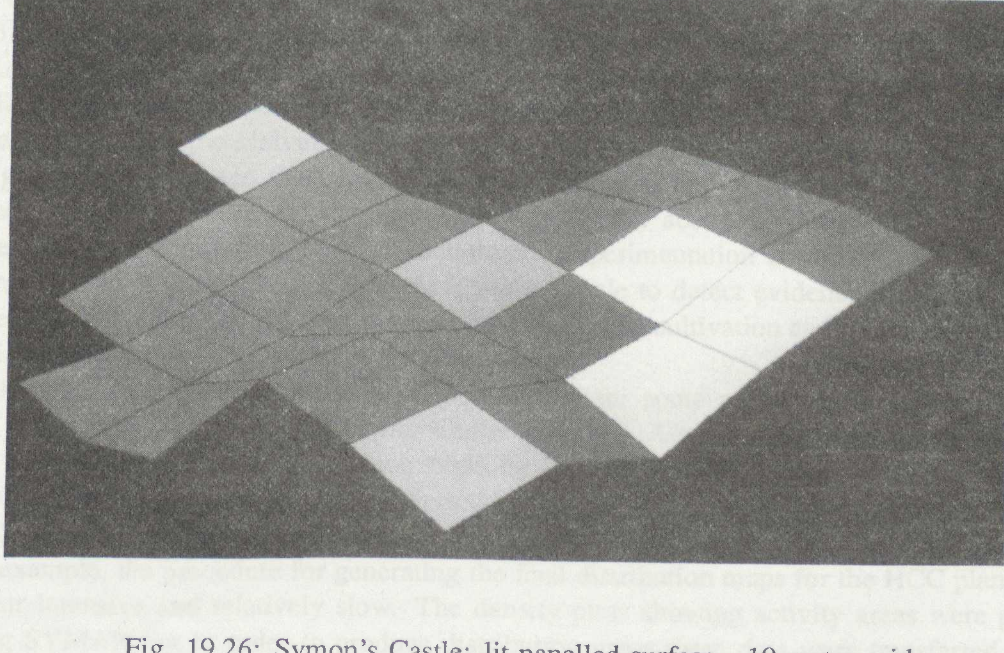

Fig. 19.26: Symon's Castle: lit panelled surface—10 metre grid

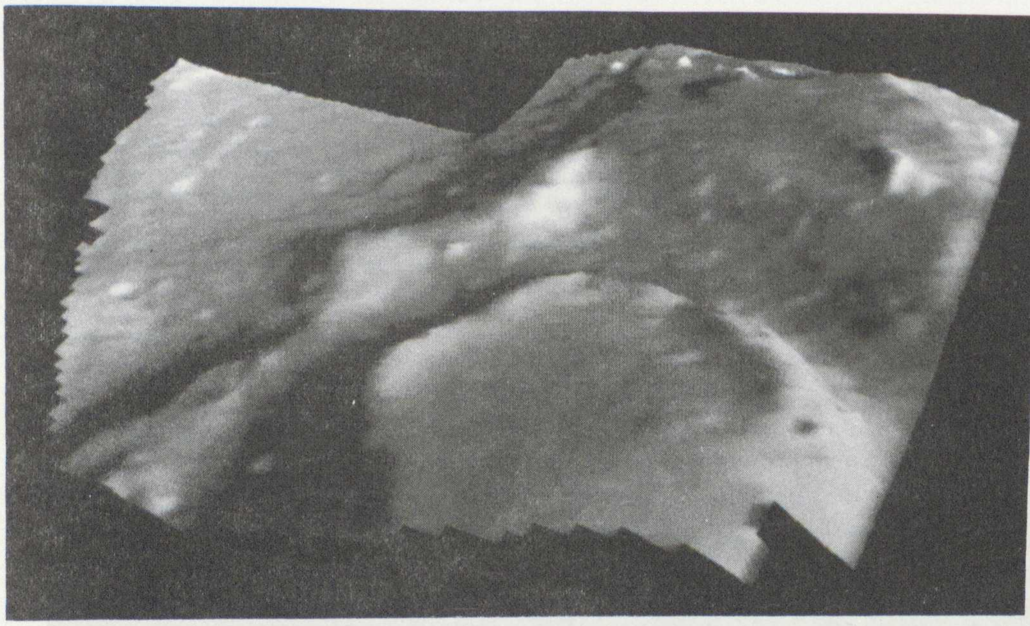

Fig. 19.27: Symon's Castle: lit modelled surface

in assisting the choice of size. For the last site discussed, the value of 0.16 for the GC is small enough to indicate a satisfactory survey, which is demonstrated by the detail reproduced in Fig. 19.27.