4 Development and application of computational methods (2)

4.1 Tests for the significance of oblique relationships

4.1.1 The theory of oblique planning

As outlined above*(2.1)*, a tentative idea was formed in the context of the hypothetical South Norfolk 'A' cadastre. This was that trigonometrical links might exist generally between oblique features and centuriated cadastres.

In order to see if this phenomenon was indeed more widespread, two publications were initially surveyed (Bradford 1957; Clavel-Lévêque 1983a). The result of this survey was reported more than four years ago (Peterson 1988a). In summary, 29 examples were found in which there appeared to be a link between oblique Roman features, mainly roads, and the cadastral grid. In 12 cases this link could be asserted with some subjective confidence. Only nine instances were seen of straight roads, some definitely and some only possibly Roman, that were definitely not related.

As proposed in the case of South Norfolk 'A' these relationships were such that the tangent of the angle between the feature and the grid was rational. In several cases it appeared that the relationship was brought about because the alignment passed through *termini*, and it was these cases which were most evidently the possible result of planning. This could have been by designing the road in such a way that it used the coordinates established by a pre-existing grid, or by positioning a cadastre so that a sequence of its grid points fell on the pre-existing road. In several other cases the same angles were observed, without the coincidence with *termini*, and, according to the theory about to be put forward, they too could be the result of planning.

It was remarked at the time (Peterson 1988a) that the observed angles were drawn from a small set, and that this limitation suggested a way in which the orientation may have been specified.

The Romans represented ratios less than unity as a whole number of smaller units (Smith 1951: 208). This restricts the number of ways in which a unit can be divided. For example, the only fractions into which the *as* (12 *unciae*) could be divided, without going to units smaller than the *uncia*, were multiples of divisors of twelve. The Romans could write (and, presumably, talk) about a fraction but they specified it in mathematical notation as a whole number of parts. So *quadrans*, one fourth of an *as* was notated as ' =- ' (three *unciae*). In this notation certain fractions could be expressed, but others, e.g. 2/5, could not, at least not in *unciae*.

The tangents of the angles of oblique features could have been defined in a similar way. The orientation of an oblique feature could have been specified by one side of a right angled triangle with its hypotenuse at the desired angle, given that the (implicit) other side was one grid distance (in this case 20 *actus*). For example, XV would indicate 15, i.e. 15:20 or 3:4.

This method of specification corresponds to that used by the Egyptians to specify the slope of pyramids at so-many palms and fingers per cubit, as documented in the Rhind papyrus (Dilke 1987b: 9); and it fits the observed data. All the angles with rational tangents observed by the author before 1988, and subsequently, belong to a set which can be expressed as "so many *actus* per 20 *actus*" *(figure 4.1)*.

It was noticeable, even in the original small sample of 29 possible relationships, that there was a marked preference for angles with simple ratios, i.e. those having a small sum of numerator and denominator when reduced to the lowest terms. Since simple ratios such as 1:1, 1:2, or 3:2 could perhaps arise by joining up *termini* which are quite near to each other, with no theoretical restriction, we have to consider the possibility that the set of angles produced by the oblique planning theory *(figure 4.1)* is unnecessarily large, and that the hypothetical specification method is unnecessarily complex.

Angle	Written as	Angle	Written as
1:10 (2:20)	II	9:10 (18:20)	XVIII
1:5 (4:20)	IV	1:1 (20:20)	XX
1:4 (5:20)	V	6:5 (24:20)	XXIV
3:10 (6:20)	VI	3:2 (30:20)	XXX
2:5 (8:20)	VIII	7:4 (35:20)	XXXV
1:2 (10:20)	X	9:4 (45:20)	XLV
3:5 (12:20)	XII	11:4 (55:20)	LV
7:10 (14:20)	XIV	3:1 (60:20)	LX
3:4 (15:20)	XV	7:2 (70:20)	LXX
4:5 (16:20)	XVI	9:2 (90:20)	XC

Figure 4.1 Rational tangent relationships in a 20 actus *grid.*

However, there are some oblique features which seem to use the ratio 11:4 (LV), which argues against this idea[70]. Surveying this angle directly on the ground would be awkward in practice since it would involve the use of remote termini. Also, why does this angle appear, when a simpler angle such as 3:7 does not? It seems more likely that the angles were normally constrained to certain values because they were specified in writing; indeed it is difficult to see how the layout of a large cadastre and its associated main roads could have been carried out unless written instructions were produced for the survey teams on the ground. The question is, what form these written instructions were likely to have taken.

It has been suggested (Peterson 1990a: 251) that in the case of one main road, the surveyors selected its alignment using a draft version of the *forma*. The first part of the Ermine Street connects London with the crossing of the River Lea at Ware. The main alignment of 30km passes through three intersections of the local cadastre at 11:4. Its ends are not visible from each other, nor is there a high point from which the whole route can be seen, so it seems most likely that, if the coincidence with major *termini* is not just pure chance, the road was planned using the pre-existing cadastral structure in some way.

---

[70] There were two ocurrences of the 11:4 relationship published in (Peterson 1988a: fig. 3), and two others had, by 1988, been seen in British contexts (Peterson 1988b: 173; 1990a: fig 12).

If so, the apparent choice of alignment is revealing. Although the road links London with a crossing of the river Lea at Ware, it does not take an absolutely direct route. A turn is made to the east at the end of the 30km straight section, on a hilltop about 2km south of the river crossing. This may indicate that the river crossing was pre-existing (although we have no other evidence that it was), and that would explain why the surveyors chose 11:4. We can imagine them consulting the draft *forma* and deciding which orientation to choose, starting from the major *terminus* nearest to London in approximately the desired direction. If the possible orientations are to be expressed under the constraints of the Roman number system *(figure 4.1)*, the choice is between 3:1, 11:4 and 5:2. The first angle would be too far to the west; the last would be too far to the east. 11:4 is almost exactly right, but the road still required a slight turn to the east on reaching the southern edge of the Lea valley. Under this interpretation, the very slight lack of directness is not a random effect. We can imagine that the surveyors knew precisely where they were going, since they had a survey grid in this area whose accuracy was comparable to those of modern times. What they lacked was our mathematical notation and computational tools. This constrained what they could specify, and hence instruct others to do.

Furthermore, this example makes it more likely that it was the angle which was being specified to the road surveyors, rather than the coordinates of any major *termini* which would lie on the road. In the case of 11:4 planning the next *terminus* would be about 8km distant, and there would be no guarantee that it could be seen. It would thus be easier for the road surveyors to proceed from a given point at a given angle, rather than from point to point. Such a procedure would be straightforward once a skeleton cadastral grid was in place on the ground. To maintain the correct angle it would only be necessary to maintain the correct proportions of the offsets in the two orthogonal directions defined by the cadastre.

This theory of oblique planning may not be correct in every detail, but its general predictions are clear. We should expect to see relationships between cadastres and oblique features, of the sort described and with the given angles.

This expectations has been fulfilled in a subsequent search of the literature, with the appearance of angles of 1:1, 3:2, 3:5 (Peterson 1992b). What is more, no data has emerged which contradicts the idea that the set of ratios (i.e. tangents of the angles) is constrained. Thus, so far, the theory stands up.

There are two interpretations of the fact that such relationships have become apparent to several workers independently. The first, sceptical, interpretation is that as we have studied more cadastres (all more or less hypothetical), we have increased the probability of seeing such relationships by pure chance. A test of this idea will be described in the next section.

The second interpretation is that they are real and were only waiting to be discovered. It is the development of interest in area, and the application of more accurate measurement techniques, which have revealed them.

If we adopt this second view, some most interesting research possibilities are open to us. These present themselves in two contexts. In the first we may have established the parameters of a cadastre (its orientation, module, and location) and we see oblique relationships between it and other Roman linear features. In the second no specific cadastral parameters have been specified, but we see Roman linear features whose angular relationships are those produced by oblique planning.

In the first context we may be able to suggest relative dating, by examining the possible chronological relationships which could have given rise to the observed topographical relationship. This must be done with care, and it may be helpful to suggest a number of principles.

Principle 1. The fact that a single segment of rectilinear road (or other linear feature) passes through *termini* does not necessarily imply that the survey of the road postdates the survey of the cadastre.

There are three cases to be considered:

(i) *The segment lies along a limes*. In this case the cadastre would normally be expected to have been based on the road, and later than it. This would normally be so, but would be less likely if another road segment also lies on a *limes*.

(ii) *The segment is parallel to a limes*. In this case the most likely relationship is that the road postdates the cadastre.

(iii) *The segment is oblique*. In this case it is still possible for the cadastre to have been based on the road. The relationship Béziers 'A' to the Via Domitia (Clavel-Lévêque 1983b; 1991) south west of Béziers at 1:2 is an example. Another example can be found east of Nîmes where there are two superimposed cadastres which are both related to another segment of the road, one at 1:10, the other at 1:3 (Chouquer 1983d; Fiches and Soyer 1983); Unless the superimposition of the grids was planned to allow for a later road to be obliquely related to both of them, which seems unlikely, the road cannot post-date both grids and so must predate at least one of them.

It may be significant that in all such cases so far observed, in which the oblique road definitely predates the cadastral grid, the relationship is of the form 1:*n*. This is what would be expected under the constraints of the Roman number system. If no grid has been established then its planned orientation with respect to the road cannot be expressed as a fraction of the number of units in the grid module *(figure 4.1)*. It can only be expressed as a unit fraction. It may be possible to construct other angles on the ground, but that is not the constraint that seems to be operating. As suggested above, the *mensor*-in-charge was obliged to issue written instructions, and thus his options were curtailed by the available mathematical notation.

Principle 2. If $n > 1$ segments of Roman road fit a grid simultaneously then it is likely that at least $n-1$ of them post-date the survey of the grid.

> This principle allows for the possibility that the grid may be based on one of the segments. It is not infallible since there may be instances where a grid may have been deliberately oriented and positioned to fit two pre-existing roads simultaneously and obliquely.[71] However, it is a reasonably positive indication that a grid exists and that some roads post-date it.

Principle 3. If a single segment of rectilinear road is oblique to a grid, not passing through *termini* and not at a theoretically significant angle, this indicates absolutely nothing about their relative chronology. The same applies for the whole length of a curvilinear road.

> There are numerous examples of such relationships arising when the road predates or postdates the grid. Some rectilinear roads, which may possibly have been based on early cadastres, will have segments oblique to later cadastres in the same area. In a Roman cadastre constructed on a large scale, even if the designer wanted to make the grid fit a segment of pre-existing road there would be other roads, and segments of the same road, which would inevitably be oblique.[72]. However, it must be added that if the road segment is oblique, does not pass through *termini*, but is at one of the theoretically significant angles, then some relationship may be suspected, although it may be indirect.

In the second context, when no specific cadastral parameters have been specified but we see a theoretically significant angle between

---

[71] This would have necessitated obtaining the solution to a Diophantine equation, see below.

[72] This principle contradicts the principle of "landscape stratigraphy", in which the oblique relationship is held to "prove" that the road postdates the field system. See below *(7.2.1)*.

Roman linear features, we may have been presented with a clue of the first importance.

Suppose, for example, that we saw an angle of 45° (1:1) between two Roman roads. If we exclude the idea that it occurred by chance, we could proceed with a limited number of hypotheses.

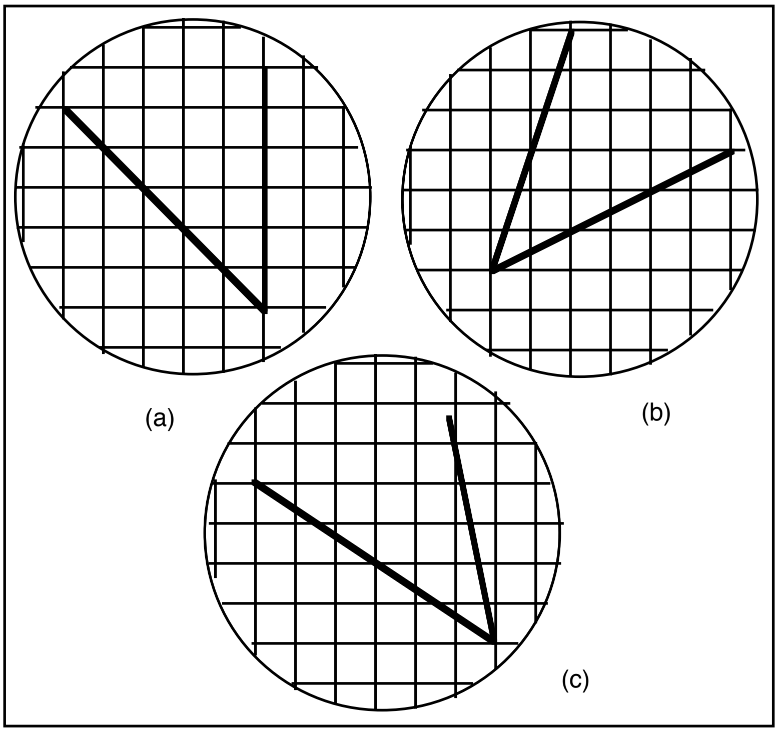


Figure 4.2 Three possible ways of generating a 1:1 relationship between linear features.

Naturally, one of the roads may be aligned with the cadastre and the other be at 1:1 to it *(figure 4.2 a)*, but there are other possibilities; for example the angle may be the difference between 3:1 and 1:2 *(figure 4.2 b)* or 5:1 and 2:3 *(figure 4.2 c)*.

In general the formula for calculating the tangent of the difference of two angles is

$$\mathrm{Tan}(\alpha - \beta) = \frac{\mathrm{Tan}(\alpha) - \mathrm{Tan}(\beta)}{1 + \mathrm{Tan}(\alpha).\mathrm{Tan}(\beta)}$$

and the same formula with the signs reversed gives the tangent of the sum of two angles.

In this case, with an observed outcome of 1:1, this equation can be satisfied by (for example) setting $\mathrm{Tan}(\alpha) = 5$ and $\mathrm{Tan}(\beta) = 2/3$, since this gives $\mathrm{Tan}(\alpha - \beta) = 1$ *(figure 4.2 c)*.

For any desired resultant angle, with rational tangent c, the problem can be restated as follows: given a ratio c, find two other ratios a and b that satisfy the equation

$$\frac{a - b}{1 + ab} = c$$

or $$a - (b + c) = abc$$

or, in words, "find three rational numbers such that the difference between one and the sum of the other two is equal to their product".

A modern mathematician would call this a Diophantine equation, and it is, when put into words, very similar to known examples of problems collected by Diophantus himself (Ore 1948: 165-208).

Thus, perhaps, this area of study is more than mere playing with numbers (Peterson 1992b). As we attempt to decipher the clues left by particular angular relationships, we need to be aware that there are in general several solutions to any given problem.

This does nothing to lessen the value of the evidence which such angular relationships provide. The inversion of the oblique planning hypothesis, arguing backwards from observed significant angles to the grid which may have produced them, could be a powerful tool, as examples in Kent *(6.2)* and at Lincoln *(6.3)* may show[73].

---

[73] This idea may have been anticipated 70 years ago. Francis Haverfield, one of this century's most distinguished Romanists, wrote as follows:

"I venture to suggest to antiquaries who have a taste for playing with instruments that they should measure the relations to the north point of any really straight pieces of Roman road which interest them, and note the deflection of each roadway from the north. I suspect that curious coincidences might be discovered, which would throw light on the roads and centuriation of Britain, and might also help to explain the process by which the Roman

4.1.2 The significance of oblique relationships

These oblique relationships, which appear to have been planned, are potentially important as an aid to an understanding of the links between cadastres and linear landscape features. It thus becomes essential to know if we have any grounds for believing that they did not occur by pure chance. The question is, how many of these relationships would we observe if we placed a straight line completely at random on a square grid?

A computer-based Monte-Carlo simulation can provide us with some answers. Without loss of generality, we can select a starting point for the line in the unit square bounded by the x and y axis, and an angle for the line between 0° and 90°. However, the choice of a random line length was based on empirical evidence of the lengths of segments of Roman roads. It was judged that, as an adequate approximation, it could be assumed that they would be distributed uniform randomly in the range 2 to 16km. For squares of 700m this would be 3 to 23 times the grid size. It was recognised that some segments of Roman road are considerably longer than this, but, since such longer lengths are less likely to fit the grid, the chosen values are likely to overestimate, rather than underestimate, the fraction of road lengths which would fit by chance

For a given line, it would be decided that it was a hit if it passed within a given distance of at least two grid intersections, and did not miss intersections it ought to hit *(figures 4.3)*.

---

roads were laid out so straight - a process which I think has not yet been fully solved, but of which I cannot here treat in detail." (Haverfield 1921: 125).

Unfortunately he was never able to give this a detailed treatment, but we are left with the suspicion that he saw the possibility of angular relationships between roads and centuriations, and also how the cadastral grid could be used to keep the roads straight.

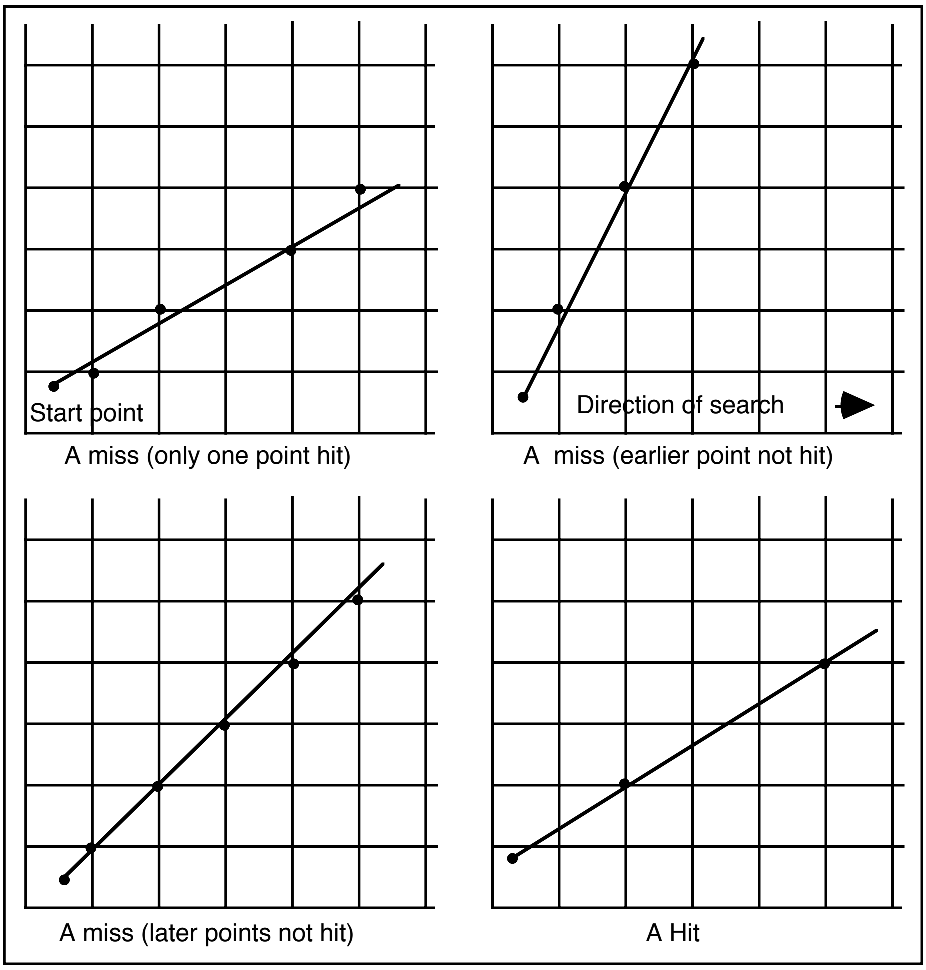


Figure 4.3 Lines and grids, hits and misses.

This definition was turned into an algorithmic form by conducting a search, ordinate by ordinate, from left to right *(figure 4.4)*. This algorithm was incorporated into several programs designed to investigate the behaviour of the simulation model.

The likelihood that the oblique line will fit depends upon the tolerance selected for the grid points. This tolerance is the 'radius' of the point. If the line passes closer to the point than the tolerance, then it hits. If the tolerance is zero, the chance of a line hitting is zero.

The relationship between the tolerance and the percentage length of lines hitting was investigated by a Basic program which tested

```
Logic of the grid simulation program

1) Select a random starting point within the unit square bounded by the y-axis to the left and the x-axis below.
2) Select a random orientation between 0 and 90° .
3) Select a length between 3 and 23 times the grid size.
4) Test for all lines at intervals of one unit parallel to the y-axis
        which are intersected by the oblique line, and
if the oblique line hits no intersections or only one intersection, add the length of the line to the total line length,
or if the oblique line hits two intersections, but misses an earlier or
        later intersection which it ought to hit, add the length of the line to the total line length,
or if the oblique line does none of these things (it fits), add the
        length of the line to (a) the total line length and (b) the total length of line which fits, and
        if the ratio of the x and y (or y and x) intercepts made by the
                oblique line is in the list of 'specifiable' angles, add the length of the line to the total length of lines which fit.
                at specifiable angles.
```

Figure 4.4 Program logic for grid simulation.

10,000 lines for each of 41 tolerance values from zero to 20% of the grid size *(figure 4.6)*.

It may seem surprising that the percentage of lines which have a rational fit does not tend towards 100% for large tolerances, but if we consider how the algorithm works, it is explicable. We can see *(figure 4.5)* that in certain cases the algorithm will decide from two points that the orientation is, for example, 0:1. However, this is an

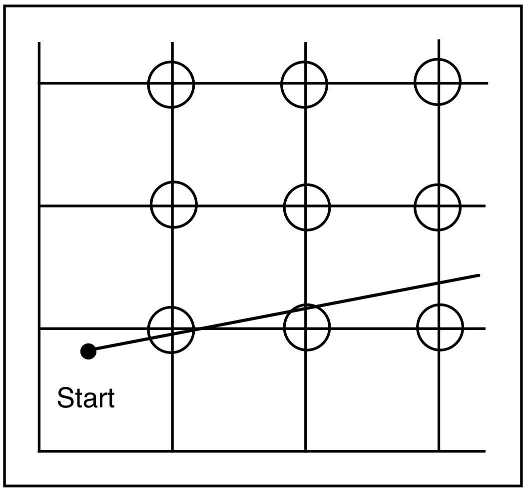


Figure 4.5 Example of simulation behaviour for large tolerances.

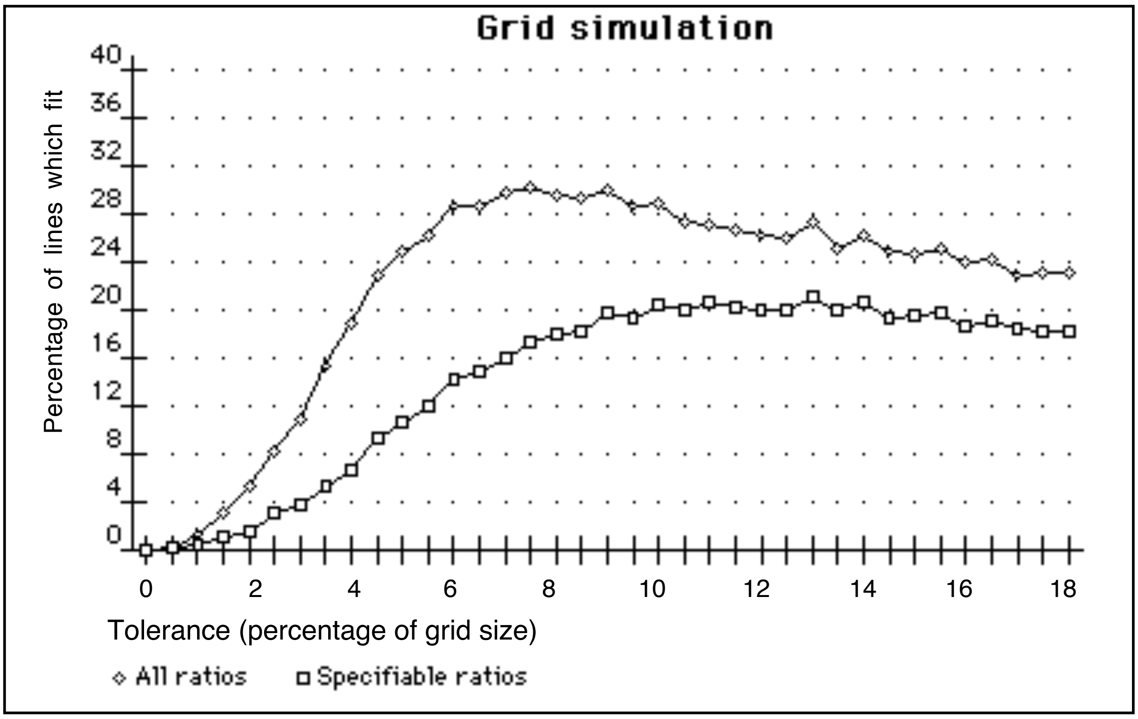


Figure 4.6 Simulation of the random fit of oblique lines to a grid.

incorrect decision, since the next point is missed and the line is accordingly rejected. If the tolerance had been smaller this particular line would have stood a chance of being accepted as a 1:5.

If we are studying the relationship of linear features to cadastral grids on a map, the question is what tolerance to assume. As stated above *(3.1.1)*, 0.4mm on the map is a reasonable figure to use. On 1:50,000 maps, this is equivalent to 20m. It gives coordinate points on the map with a diameter of 0.8mm, which is large when compared to the width of the conventional representation of a single carriageway road, which varies between about 0.5mm and 0.7mm. We can thus be confident that, with this tolerance, a fit or (more importantly) a lack of fit will be determined without excessive ambiguity. If we use a radius of 20m, this represents a tolerance of 2.8%.

From the simulation result we can see that, for this tolerance, under 4% of random lines will fit. This may be examined more closely by a further simulation.

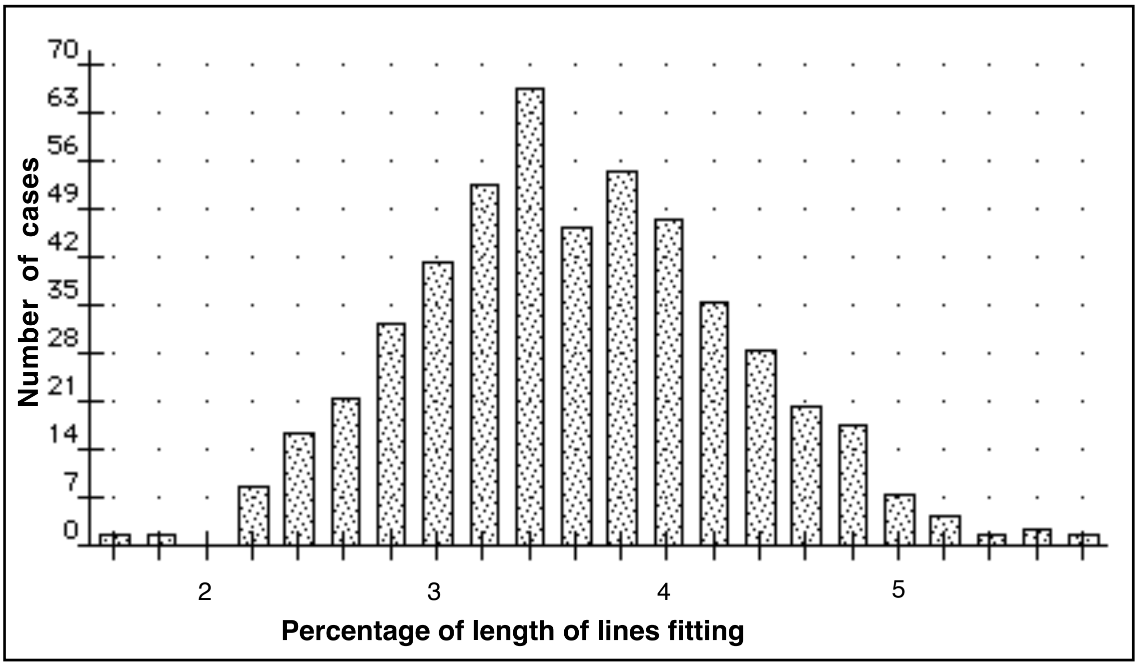


Figure 4.7 *Distribution of percentages fitting (500 trials of 1000 lines with a tolerance 2.8%).*

The simulation was performed 500 times to produce a histogram of the percentage fitting in each trial, in classes of 0.2% *(figure 4.7)*. If we take this to approximate a normal distribution, its statistics are:

mean:	3.65
standard deviation:	1.28

Thus there is less than a 16% probability that more than 5% of any sample of lines will fit, and there is only a 0.003% (1 in 33,000) chance of more than 9% fitting. These figures represent respectively: the mean plus one standard deviation and the mean plus four standard deviations (Moroney 1956: 108-119).

The conclusion that we can draw from the simulation study is that we cannot neglect the possibility that a supposed fit of an oblique feature to a cadastral grid has occurred by chance. About 1 in 25 segments of roman road will fit a grid positioned in a completely arbitrary fashion.

We can only be prepared to bet that the associations are not chance when we have a reasonably large sample, say 25 segments, and more than 9% of the total length of the segments is in segments which fit.

Unfortunately this is hardly ever achieved in practice. There may not be enough segments of road, and the situation is inevitably complicated by the presence of roads which were established earlier than the presumed date for the establishment of the cadastre. They are not expected to fit, except by chance, so if they are included in the sample they may unfairly bias the result against the hypothesis that the cadastre exists.

Nevertheless, so long as we accept these caveats, we can say that there is only about a 4% chance that a length of road, chosen at random, will fit a grid to the tolerance given here.

This implies that if we see such a thing in an otherwise arbitrary grid, we are looking at what would be, on the basis of chance, a fairly rare phenomenon.

There are also some cases, in real cadastres, in which oblique segments meet at grid square corners. On the basis of pure chance this should be an even more rare event. A simulation of this was carried out, using the same tolerance as before (2.8%), taking pairs of lines with length ten units, starting at a fixed point (0.999, 0.999) very close to a corner of the unit square, with a random orientation. The results *(figure 4.8)* showed that for each 10,000 lines (5,000 pairs) there were about 100 cases in which both lines fitted the grid at a specifiable angle.

Tries	Hits
10000	96
20000	195
30000	293
40000	390
50000	494
60000	581
70000	679
80000	789
90000	885
100000	989
110000	1073
120000	1185
130000	1287
140000	1388
150000	1484
160000	1576
170000	1686
180000	1783
190000	1875
200000	1977

Figure 4.8
Double hits.

Given that the intersection of the two lines, defined to a tolerance of 0.028, has a chance of 0.0025 of falling at a given point in the unit square, there is a chance of about 1 in 40,000 that two lines will meet at an intersection and both fit.

4.2 Application of Geographic Information System software

4.2.1 GIS software

Although at this stage of development of the technology it is probably pointless to attempt to formulate precise definitions, we would use the term 'Geographic Information System' to refer to a type of Land Information System (Dale and McLaughlin 1988: 8-15) designed to fulfil an organisation's needs for the handling of geographic, and particularly spatial, information. GIS software would then be the term used to refer to the integrated software components which allow such a socio-technical system to operate.[74]

Further reading on GIS software may be found in Peuquet and Marble (1990), an introduction together with a case study applied to archaeology in Gaffney and Stancic (1991), and further archaeological case studies in Allen, Green and Zubrow (1990). It is thus unnecessary to give detailed descriptions here. Nevertheless it may be helpful to state that one of the essential features of GIS software is the ability to combine information about a particular location or set of locations and to produce new information. For example, for any given point we may obtain an attribute value, slope, which is derived from altitude values of neighbouring points. Much more complex calculations are also possible. This processing function distinguishes a system supported by GIS software from other types of land information system, such as cadastres (ancient or modern), in which little processing of the spatial data takes place.

---

[74] It is necessary to draw a distinction between a GIS and GIS software, particularly when the term GIS has been used to refer what would here be called GIS software, for example by Marble (1990: 10).

There are two different ways in which GIS software represents space:
(a) Vector representations, which are a natural development from automated cartography systems, and which can have topological information associated with the representations of lines as sequences of coordinate points.
(b) Raster representations consisting of pixels, which are analogous to, and sometimes derived from, images such as those produced by satellites. This form of representation allows for the easy creation of new images by map algebra.

IDRISI (Eastman 1990), the system used for this research, runs on IBM PCs and compatible machines. It can input and output vector data, and convert between both forms of representation, but it is essentially a raster system. The raster facilities allow for the easy creation of new images, or layers, by arithmetical operations on corresponding raster cell values, using add, subtract, multiply, etc. (the 'overlay' function). Other facilities are: 'reclassification', which changes values of cells (for example to concatenate two areas); 'assignment' which creates a new layer which has values assigned to cells according to a look-up table; the 'distance' function which calculates euclidean distance from non-zero target cells; and other functions.

IDRISI is designed to provide inexpensive access to computer-assisted geographic analysis technology and it was originally intended as a research and teaching tool (Eastman 1990: 1). As such, the system is not designed to fully support Geographic Information Systems, as defined above. However, it has proved to be suitable for the research undertaken here.

As an application of GIS software, this research has limited and specific aims. The intention was not to establish any sort of multi-purpose database, but to use the analysis facilties of the software, in conjunction with spreadsheets, to study features in relatively restricted areas.

4.2.2 The relationship between possible cadastral traces and soil type in Romney Marsh

From a consideration of the trigonometrical relationships between Roman roads, it appears that a centuriated cadastre, Kent 'A', may have been implemented in east Kent. Further reasons for pursuing this line of investigation will be given below *(6.2)*.

GIS software can be used in an attempt to investigate a possible relationship between possible traces of this cadastre and the general division of soil types in Romney Marsh, an area lying at Kent's southern extremity.

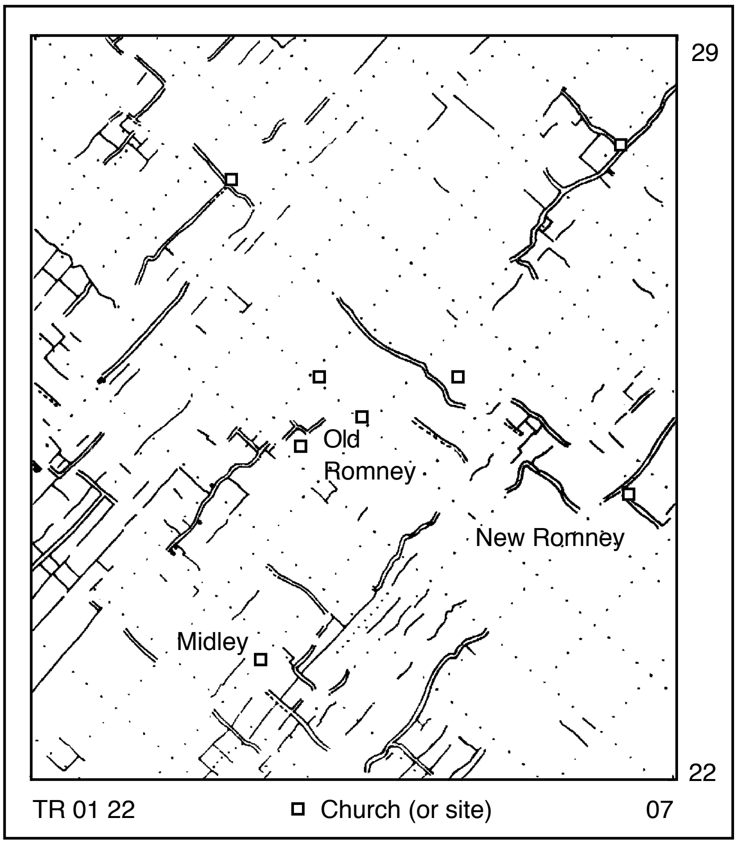
In this area there is a basic distinction, made by Green (1968), between soils of the Old Marsh, essentially dating back a least as far as the Roman period, and soils of the New Marsh, which were flooded in the middle ages and later[75].

This distinction had absolutely no influence on the choice of position and orientation for the hypothetical Kent 'A', because it was initially thought that such a system, if it existed at all, was restricted to the area including Canterbury, Dover and Richborough, some 30km away. It was also thought that Romney Marsh was an area in which the Romans had no great interest.

It is therefore surprising to see, principally in the Old Marsh, an extremely regular layout of roads and ditches, many of which correspond in both orientation and position to the major *limites* of the grid, and to the subdivision of 10 *actus (figure 4.9)*. Such subdivision of a centuriation is frequently encountered in other

---

[75] Nicholas Brooks (1988) agrees. He says that "Fundamental to any understanding of Romney Marsh in the early Middle Ages is Green's distinction between the 'Calcareous' or New Marshland, which has been subjected to inundation by the sea within historic times, and the 'Decalcified' or Old Marshland from which the calcium has largely leached away after centuries of natural drainage." But the use of the word "natural" is questionable.



Based on the 1960 Ordnance Survey 1:25,000 map with the permission of the Controller of her Majesty's Stationary Office © Crown Copyright

Figure 4.9 *Possible traces of Kent 'A' cadastre in the centre of Romney Marsh.*

badly drained areas, such as the Po valley[76], and the effect is to produce a structure resembling a grid of 10 *actus*. This is what seems to appear here but an opinion on the likelihood that these traces represent the remains of a Roman cadastre is still a matter of judgement, based upon knowledge of genuine degraded cadastres in similar environmental conditions. The morphology of the system, by itself, leaves the question of the presence or absence of the cadastre open to unfruitful debate.

We have seen *(3.3)* that one way to add objectivity to this debate may be to test the location of indicators of medieval settlement, the churches and courts, against the 355m grid. However, tests of this sort often meet with a sceptical response from those who, as a condition of belief, insist upon seeing some physical connection. It may therefore be useful to attempt another approach.

We can use a computational approach to test our subjective perception that traces corresponding to the 10 *actus* grid appear to be much more prominent in the Old, rather than the New, Marsh.

---

[76] See, for example, the reproductions of maps from the Catasto Boncompagni of 1780 (Romano & Vivanti 1976: figs 66-70).

If the cadastral "traces" are genuine, one would expect the areas of Marsh which were definitely flooded in the Middle Ages to show less of them[77]. It could thus be revealing to measure their relative densities on the two types of soil, and to see if detailed variation in density suggests any model for the process of formation of the present landscape.

4.2.2.1 IDRISI procedures

A simplified soil map, based on those of Green (1968: figs 14 and 16), was digitised using a Summagraphics Microgrid III digitiser linked to a Viglen 386 personal computer running DIGIT II software. In order to check the completeness of the digitisation, the individual areas of soil type were digitised as polygons. The DIGIT II output was converted to ARC/INFO format using the TOARC software. This produced a list of pairs of coordinates for each arc as ASCII text [78].

Since there was no conversion software available to create IDRISI polygons, the sequences of coordinates making up the arcs were transferred to a Microsoft Works spreadsheet and each one was sequence numbered. The sequences of coordinates could then be

---

[77] However, we see that traces are not completely absent in the New Marsh. There may be for several reasons:
i) Error in the soil survey
ii) The chance occurrence of a topographic trace on a *limes (Appendix 3)*. This is the inevitable background to genuine traces.
iii) The physical preservation of the *limes*, for example as a bank, despite the flooding.
iv) The conceptual preservation of the *limes*, as a line between landmarks which are still visible, or as a record of a property boundary which can be reconstructed when the flooding is past.
v) The construction of new land divisions in reclaimed land along the same lines as the divisions on the neighbouring dry land.

[78] For Romney Soils this took ROMNEY.CNA (DIGIT II polygon output) to ARC/INFO format in ROMNEYAR.LIN, ROMNEYAR.PNT, ROMNEYAR.LAB.

copied, reversed if necessary and joined together to make IDRISI polygons, with IDRISI headers and trailers added[79].

Command (input file)	**Values**	**Output file**
Initial	267 rows, 294 columns. Cell size 75x75m. Value 0.	ROMRAS
Polyras w ROMPOL	Min x 92000 Max x 114000 Min y 15000 Max y 34999	ROMRAS
Reclass ROMRAS	Background 0 Old Marsh 2 New Marsh 3 Shingle/sand 4	ROMSOILS (Palette ROMSOILS)
Initial	As above	TRACERAS
Reclass TRACERAS	Background 0 Traces 1	TRACE1
Overlay TRACE1 on ROMSOILS	Cover (replace cells in ROMSOILS by cells with value 1 in TRACE1)	TRACES

Figure 4.10 Initial IDRISI processing of digitised data.

The features corresponding to a grid of 355 metres were also digitised and converted by TOARC to ARC/INFO format. This could be read directly by the IDRISI 'arcidris' function into a line vector file ROMTRACE.VEC. These vector files were processed as shown *(figure 4.10)*. The result is the image, TRACES *(figure 4.11)*.

Using the 'color' display module of IDRISI, we can see an image similar to that seen in hand-traced maps, but in less detail.

---

[79] Given that the simplified soil map contained only 8 polygons, this semi-manual method probably took less time than would have been required to write a conversion program. Such a program would undoubtedly have been needed for the larger South Norfolk soil map *(4.2.3)*.

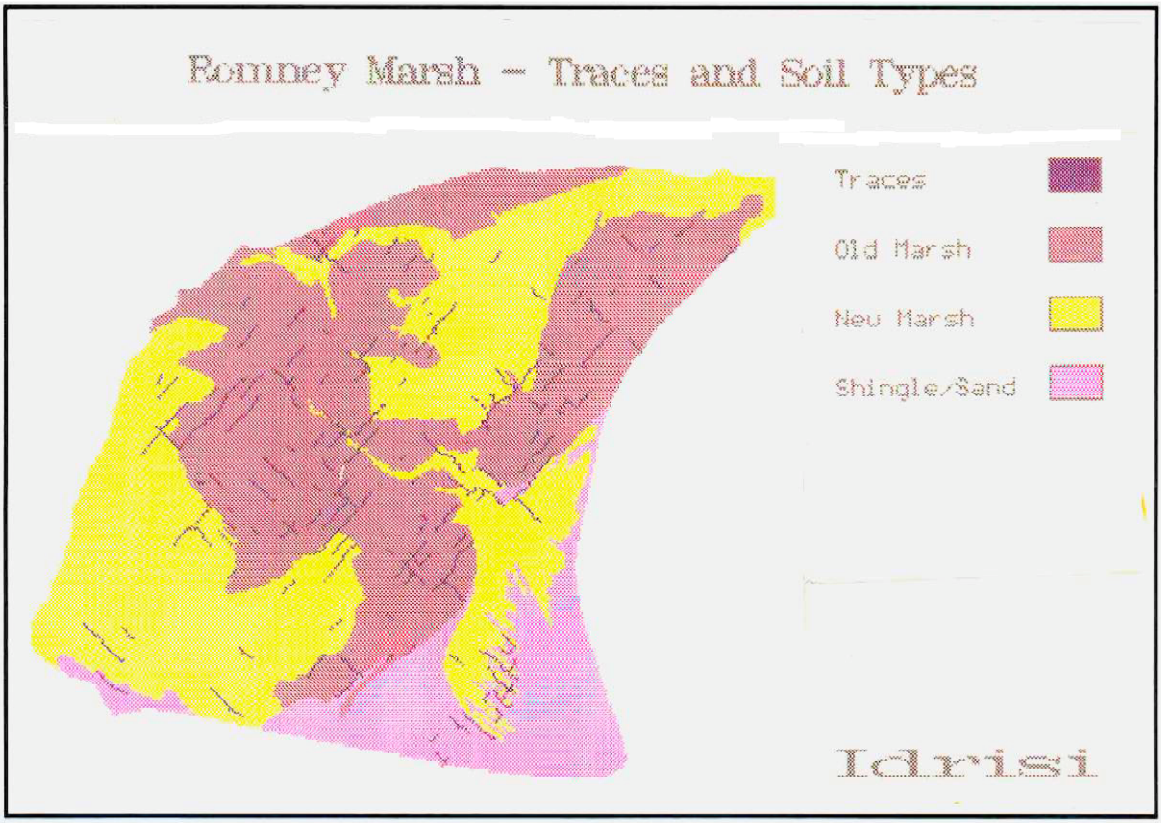


Figure 4.11 Romney Marsh - generalised soil types and possible cadastral traces

This was the result of the choice of a raster grid of 75x75m squares, a figure that was arrived at after some experimentation with different cell sizes. Limits on available elapsed time for processing and disk space imposed by the available equipment[80] did not make it feasible to use a finer grid.

Nevertheless, even with this grid size it is possible to see the distribution of the traces and (much more important) it is possible to treat them quantitatively. Looking at the image it appears that the density of traces is low in the New Marsh but that it increases towards its boundary with both the Old Marsh and shingle/sand areas. It can also be observed that in several places a trace forms part of this boundary. This subjective impression can be tested. We

---

[80] Most of the work was done on an Amstrad PC1640 HD20. Some processes, such as distance calculation, had an elapsed time of several minutes.

can use the GIS software to measure how the density of traces changes with distance from the new marsh edge.

The IDRISI 'distance' function was used, in conjunction with other operations, to compute the values for the distances of raster cells from the New Marsh boundary*(figure 4.12)*.

Command (input file)	**Values**	**Output file**
Reclass ROMSOILS	1 in new soils, else 0.	NEWSOILS
Distance NEWSOILS		DNEW

Figure 4.12 IDRISI calculation of distance values.

The equivalent images OLDSOILS and DOLD were also produced. NEWSOILS and OLDSOILS were then 'overlaid', by adding the values of corresponding cells, to produce an image, MASK, with value 1 in the Marsh area and 0 elsewhere[81]. Cell values in MASK and DOLD were then multiplied, using 'overlay', to produce an image with positive distance values only in the New Marsh area, and zero elsewhere. The 'histo' function then produced a histogram of cell values. This showed that the lowest non-zero value was 75m, which is the cell width[82]. The highest value was 3624.9m. This image was then processed further with DNEW *(figure 4.13)*, to produce an image with non-zero values only in the Marsh area, increasing from a rounded value of 1 in cells within the New Marsh furthest from the boundary, to 3551 and 3626 on each side of the New/Old Marsh

---

[81] This was also a useful check on the operation of IDRISI. Since there were no cells with value 2, and no cells with value 0 between the Old and New Marsh areas. Thus, as you would expect, the boundary between Old and New was defined without ambiguity.

[82] Could this minimum value affect the result of statistical tests, such as Kolmogorov-Smirnov? There may be less than expected within a distance equal to the cell width.

boundary, and then increasing further with greater distance from the boundary within the Old Marsh.

Command (input file)	**Values/parameters**	**Output file**
Reclass ROMSOILS	1 in new soils, else 0.	OLDSOILS
Distance OLDSOILS		DOLD
Overlay OLDSOILS NEWSOILS	add	MASK
overlay DOLD MASK	multiply	DOLDA
Initial	267 rows, 294 columns. Cell size 75x75m. Value 3626.	VAL3626
overlay VAL3626 DOLDA	Subtract (cell values in first minus cell values in second)	DOLDB
Initial	As above. Value 3551.	VAL3551
Overlay DNEW VAL3551	Add	DNEWA
overlay DNEWA OLDSOILS	multiply	DNEWB
overlay DNEWB DOLDB	cover (non-zero cells of DNEWB replace cells of DOLDB)	DIST
overlay DIST MASK	multiply	DISTANCE

Figure 4.13 IDRISI generation of DISTANCE image.

This gives a set of values which is lowest in the centre of the new marsh areas and which continues to increase across the boundary, giving a distance value for the boundary, at a point between adjacent cells 75m apart, of about 3588.

The 'color' module shows this DISTANCE image *(figure 4.14)* and allows us to see the relationship between the traces, overlaid from the vector file, and the band containing the edge, shown in white.

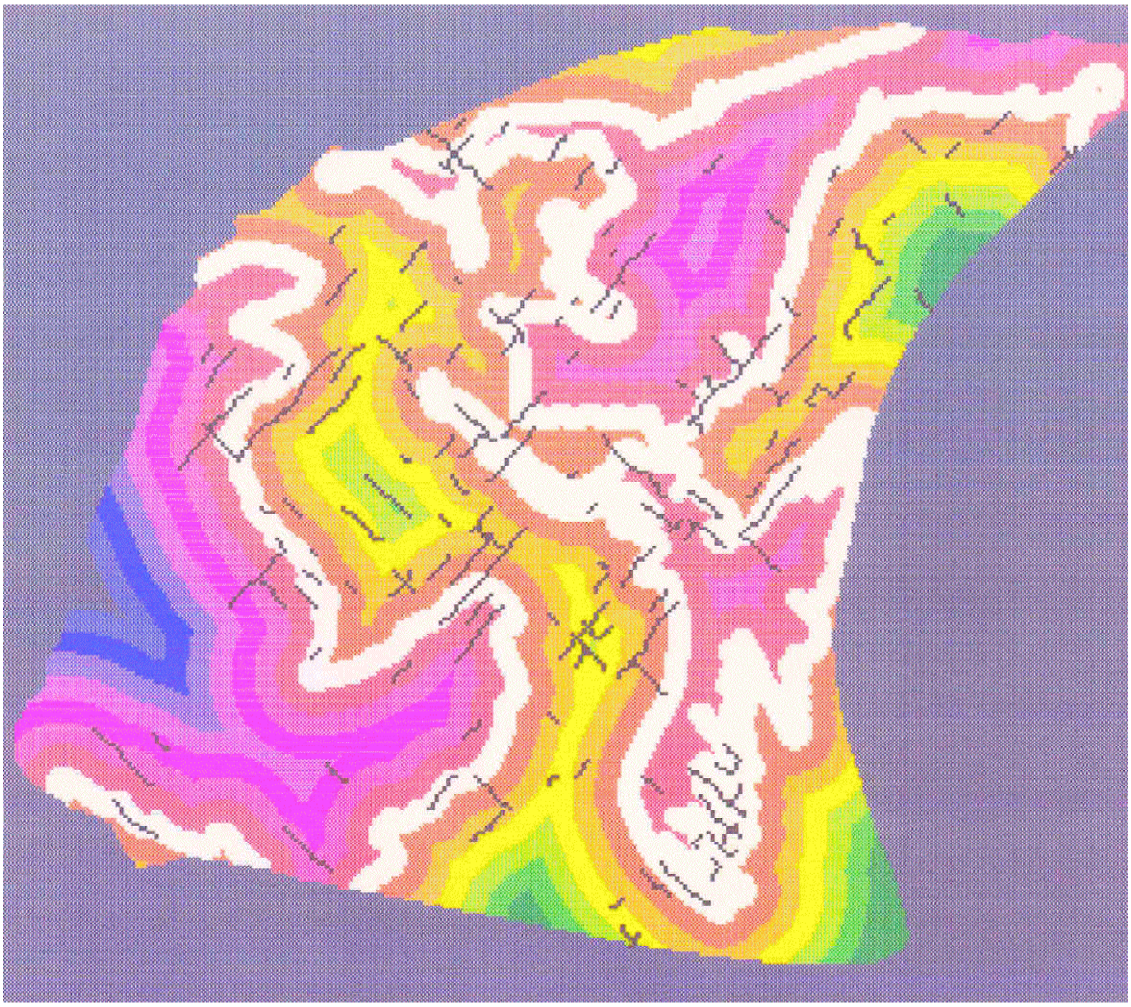


Figure 4.14 Distances from edge of Romney Marsh new soils.

We can also examine this relationship numerically using the 'histo' function. For all cells in the image a histogram was produced of the numbers of cells in each interval of 100m. The lowest limit was set to 88m, so that the edge of a band would coincide approximately with the New and Old Marsh boundary.

For the traces, the 'overlay' function was used again to multiply cell values from TRACE1 with DISTANCE cell values, giving an image with cells containing distance values if they were in a trace, and zero otherwise. A histogram of this image was then produced, using the same intervals as before. This allows us to express the number of trace cells as a proportion the total number of cells in each 100 metre band, thus giving us a measure of the density of traces in relation to distance from the New Marsh boundary.

4.2.2.2 Spreadsheet procedures

The two sets of histogram values were entered into a MS Works spreadsheet on a Macintosh plus. Trace densities in each 100m band were noticeably lower in the New Marsh, compared to the Old, and high on the boundary.

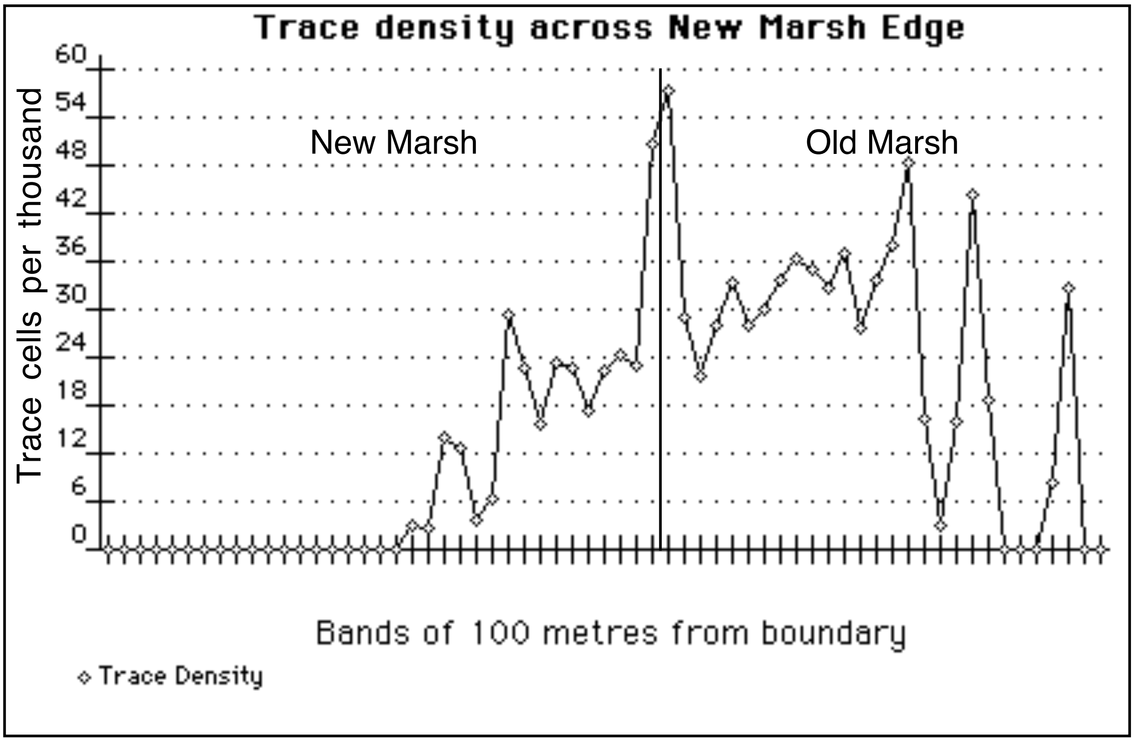


Figure 4.15 Relationship of trace density to distance from New/Old Marsh boundary.

These figures *(figure 4.15)* are difficult to interpret because of the wide range of areas in each 100m band. When the area in each band becomes relatively small, as in the case of the extreme values, the density figures are unreliable.

For this reason the data were examined on the basis of bands of equal area, rather than equal distance from the new marsh edge. A band area of 800 cells (of 75x75m) was chosen as a convenient round number which would produce a number of bands approaching the maximum number of 80 points allowed in an MS Works series chart.

Columns for corresponding cumulative marsh area and trace area were created from the original histogram values, and from these, by linear interpolation, cumulative trace area figures were calculated at intervals of 800 units of Marsh area. It was noted that the cumulative trace area figure nearest to 18382 (the number of cells in the New Marsh area) was 18400, the 23rd point in the series. This point is thus the one nearest to the boundary *(figure 4.16)*.

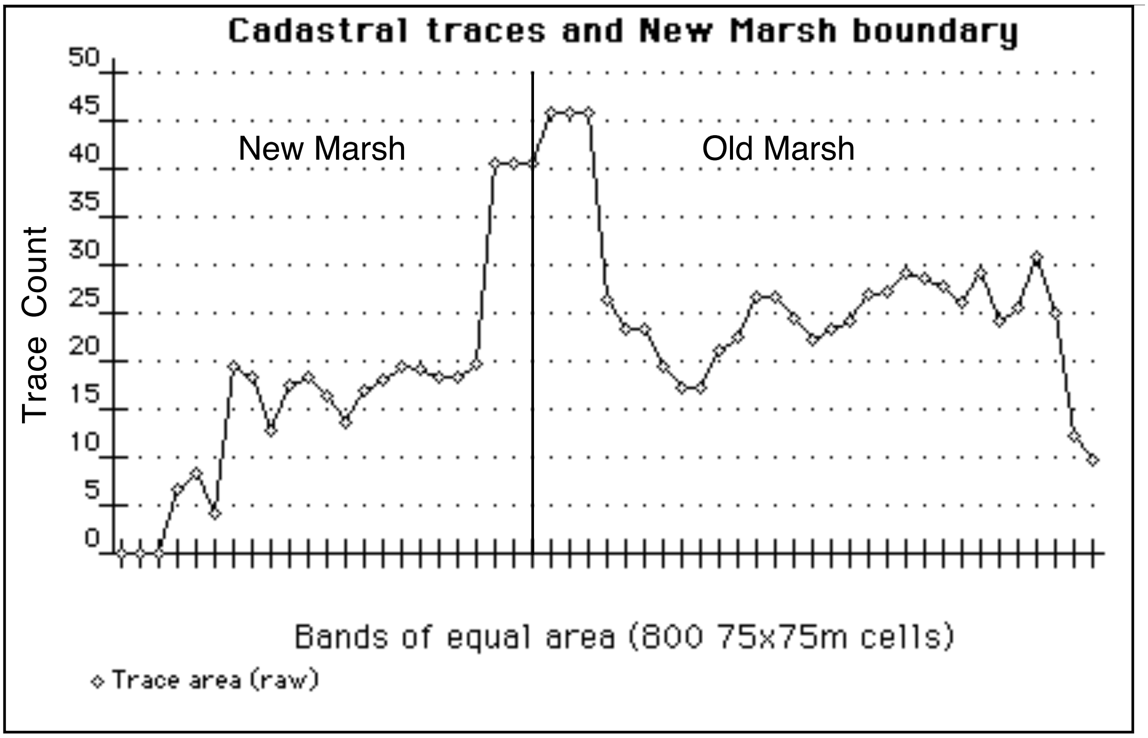


Figure 4.16 Variation in density of cadastral traces in bands of equal area with distance from New/Old Marsh boundary.

This chart shows the density of traces is low in the new marsh, higher in the old marsh and very high on the boundary between the two.

However, before rushing to interpret further, we need to recognise that these results are uncalibrated. Perhaps this distribution of possible cadastral traces merely reflects a general distribution of topographic features. Possibly all traces have a highest density near the edge of the new marsh, with lower values elsewhere, particularly in new marsh areas, and the procedure for extracting traces at 355m has merely sampled this. We need to obtain a measure of the density of all topographic traces with respect to the Old/New Marsh boundary.

It would be arduous to digitise all the features in the Marsh from the 1:50,000 map[83]. Accordingly a sampling strategy was adopted. Squares were selected at random, using an MS Works speadsheet to generate their national grid coordinates, until the set of squares selected included 10 which were completely in the Marsh. All the map traces lying within these squares and within the Marsh area[84] were digitised as if they might have been cadastral traces[85]. The outlines of the sample squares were also digitised.

The sample squares and the digitised traces within them were converted to IDRISI images with value 1 in the squares or on the traces, and zero elsewhere. The 'overlay' function was then used to multiply their cell values with the values of corresponding cells in the DISTANCE image, thus producing two images of distance values one for a sample of the Marsh area and the other for all the topographic features within it, except railways and power lines. Histograms were produced for these images using the same intervals and limits as previously.

---

[83] Based on the time taken to digitise the 5% sample, which was approximately 45 minutes, it would have taken about 2 days to digitise the whole.

[84] That is, within the areas of Marsh soils defined by Green.

[85] Roads were therefore digitised with single line, as were footpaths and watercouses running within 0.5mm of each other on the map.

The spreadsheet was modified to accept either the data for the cadastral traces or the data for the sample of all traces and still produce comparable output. The aim was to compare the two distributions of traces in bands of equal area at a given distance from the New Marsh edge.

A constant band area of 800 units of Marsh area cannot be used for both cases, because the sets of data cover different areas. It was thus decided to use a band area of 2% of the total Marsh area in each case, giving 50 bands for both sets of data. As before, linear interpolation from the count of traces per 100m band gives us cumulative trace values for these intervals and hence trace values per band.

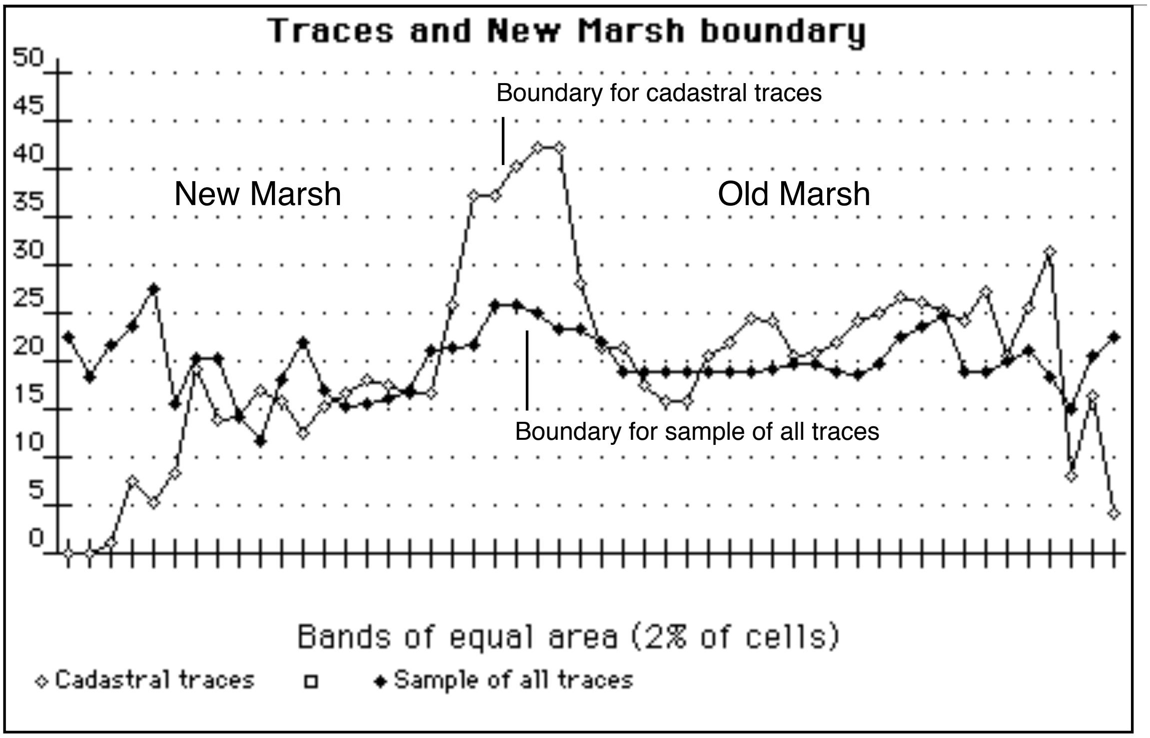


Figure 4.17 Distribution of cadastral and sample traces, both in bands of 2% of total area covered in each case.

The spreadsheet was run for both sets of data and the two results plotted on one chart *(figure 4.17)*. The values for trace cells in each 2% band were scaled so that they would total 1000 in both cases, thus giving a mean value per band of 20. It can be seen that the cadastral traces depart grossly from this mean, with 100% variation both ways. The count of trace cells is more than twice as high just inside the Old Marsh and zero or very low for the furthest fifth of the New Marsh. On the other hand, with the exception of five values, the sample traces do not vary from the mean by more than 25%. The only possibly significant departure from a constant value with random fluctuations may be the group of five elevated values which peak just inside the New Marsh.

Thus the cadastral data and the sample data appear to have a different distribution when compared on the basis of 2% bands of the area covered in each case.

We may also compare the figures for the mean trace count in New and Old Marsh *(figure 4.18)*. Here again there is a contrast between the values for the sample, which differ by no more than 1% from the mean, and the cadastral trace values, which differ by as much as 27%.[86]

	New Marsh	Old Marsh
Cadastral Traces	15.4	23.4
Sample Traces	19.8	20.1

Figure 4.18 New and old marsh - mean trace densities.

However, before attempting further comparison of these two sets of data, we note that their boundary points are not between the same intervals. The reason for this is that the sample represents only 5.3% of the total Marsh area so there is no guarantee that the proportions of Old to New soils in the sample will be the same as that for the Marsh as a whole. Hence, if we are to compare the distributions of traces on the basis of the same bands of area it is necessary to adjust the figures in the sample. The calculation

$$\frac{\text{Sample Trace Area for band x Total Marsh Area for band}}{\text{Sample Marsh Area for band}}$$

---

[86] Although this seems to be an obvious difference, the author would not like to comment on its statistical significance.

gives us a figure for the number of trace cells which would have been in that band, if the sample densities applied to the whole Marsh. In cases where no cells in a particular 100m band were sampled, we set the trace count value to zero.

The adjusted sample values can then be interpolated as before and charted with the cadastral trace values *(figure 4.19)*.

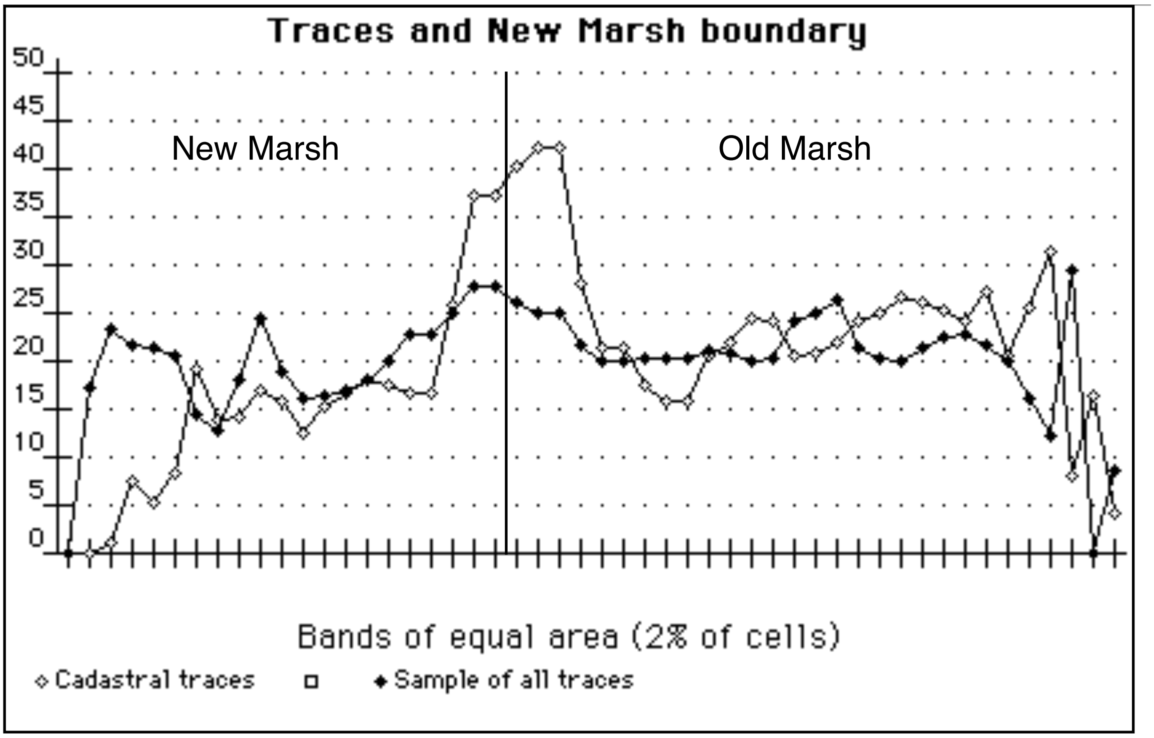


Figure 4.19 Comparison of distribution of cadastral traces and adjusted sample of all traces.

The problem with this result is the presence of the two zero values in the sample distribution. Inspection of the spreadsheet reveals that this is due to the fact that no Marsh area was sampled at those distances. Thus the values represent an 'unknown' rather than a true zero. It thus seems appropriate to discard the leftmost and the

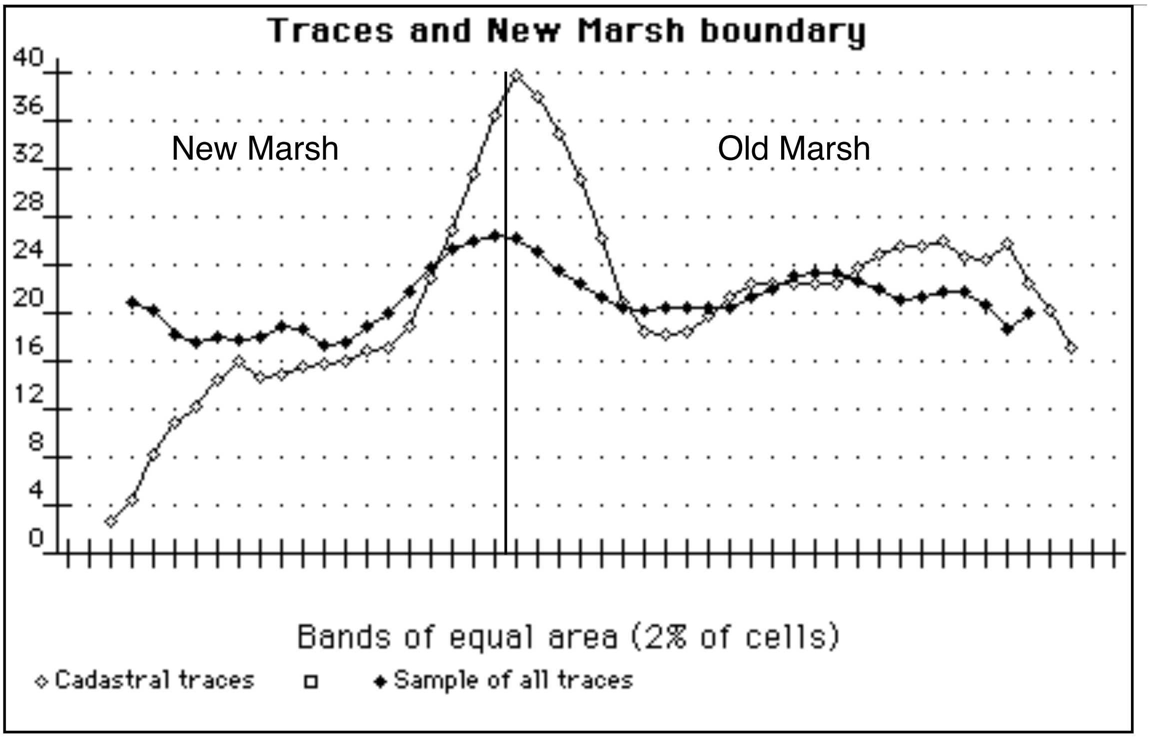


Figure 4.20 *Cadastral traces and adjusted sample of all traces (smoothed by 5 point moving average).*

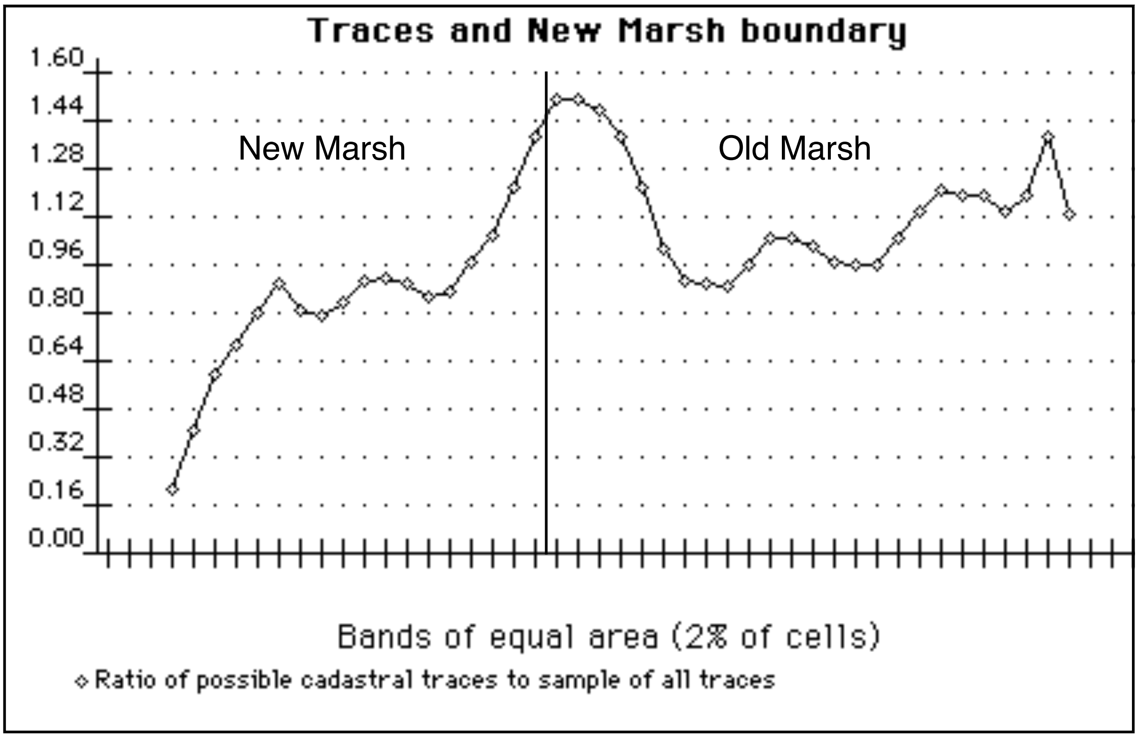


Figure 4.21 *Ratio of cadastral to sample trace counts shown in Figure 4. 20.*

two rightmost figures, since it would be misleading to retain them. This was done, and the figures smoothed by a 5 point moving average *(figure 4.20).*

The two trace counts in each column were then expressed as a ratio *(figure 4.21)* and the ratio values smoothed by a three point moving average to produce a simplified form of the original data *(figure 4.22)* which represents, in all probability, the most outstanding features of the distribution of cadastral traces when compared to the background of all topographic traces in the Marsh.

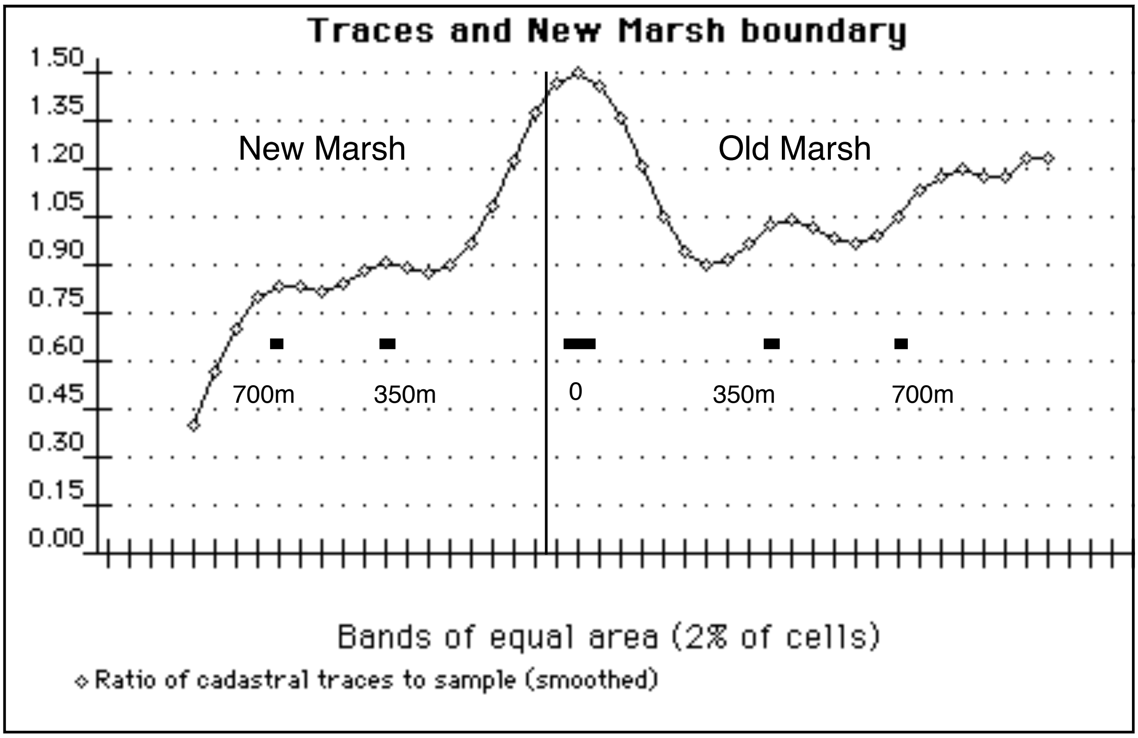


Figure 4.22 Romney Marsh - cadastral trace density compared to background density of all traces.

This idealised density distribution shows a sharp rise from very low values at the points furthest from the boundary in the New Marsh, followed by a steady rise (with fluctuations) from about 80% to 120% of the mean value on going from the New Marsh into the Old Marsh. This is interrupted by a peak of 150% of the mean value just inside the boundary.

4.2.2.3.Interpretation of results

The final result of this analysis *(figure 4.22)* seems to show a number of features which can be explained according to the following hypothesis: that the partial inundation, possibly on more than one occasion, of a centuriated cadastre has led to the selective preservation of its major features, which are *limites* at 355m. The most striking feature of the distribution is the high density of possible cadastral traces corresponding to the well-documented division between Old and New Marsh. The position of this peak is shown on the smoothed graph of the ratio of cadastral to sample traces *(figure 4.22)* at about 50-60m on the Old Marsh side of the Old/New Marsh Boundary. This small shift in the peak position from that shown in the previous figure, where it is on the boundary, is the result of smoothing an asymmetrical peak. In any case, the association with the boundary seems clear.

There also appear to be subsidiary peaks at distances of approximately 350m and 700m on each side of it. These distances are indicated by bars about 50m wide placed on the figure at the appropriate points[87]. The suggested explanation for these subsidiary peaks is that they represent those traces of *limites* spaced at 355m which are parallel to sections of the boundary defined by a *limes*.

If we ignore these fluctuations, and consider the general trend of the distribution, we can suggest that there is very little possible trace of the cadastre at a distance within the New Marsh greater than 1km from its boundary *(see also figure 4.15)*; this would represent areas inundated totally since the Roman period, in which

---

[87] The distance values from the New Marsh boundary for the upper limits of the 2% area bands were obtained by interpolation in a spreadsheet. They were used to select the numbers of the bands nearest to the figures shown. The figures selected from the spreadsheet, giving good approximations, were

Distance	-656	-291	55	415	748
Band Number	9	14	23	32	38

all organisation has been lost. Nearer the boundary, however, traces are apparent at about 85% of their mean value; this would represent areas in which the inundation was neither total nor constant, so cadastral structures were preserved physically or in human memory and could be reconstituted when possible. Finally, the higher level of traces in the Old Marsh, at about 105% of the mean value, is a result of the better preservation of cadastral structure on land that was flooded only a little, if at all.

Thus the numerical treatment of these data confirms the subjective impression of a major difference in density of traces between the Old and New soils, even when any possible variation in the density of all traces is taken into account. It also suggests that there is a real association between their boundary and the boundaries which would have been established by the Kent A cadastre.

However we must conclude this section on a cautionary note. It is possible that in the sample the density of traces in a kilometre square is underestimated for those kilometre squares which include a large numbers of traces. They are originally digitised as vectors, but are represented as lines of raster cells. Thus there is an upper bound on the apparent trace density, when all raster cells in a sample are trace cells. This would occur when no point on a trace is further than the raster grid distance from another trace.

Only for low trace densities will the ratio between trace density, expressed in two ways, as a vector length and as cell count, be approximately constant. At higher densities, approaching the upper bound of cell count, the relationship is non linear. This non-linearity is a function of the grid size.

In this example, visual inspection of the raster representation of the traces in the most dense kilometre square does not show it to have approached the upper bound. Also, if we halve the side of the raster cell, convert the vector sample traces to an image at this new cell size and produce a chart which compares the distribution of sample traces in bands of equal area for the two different raster

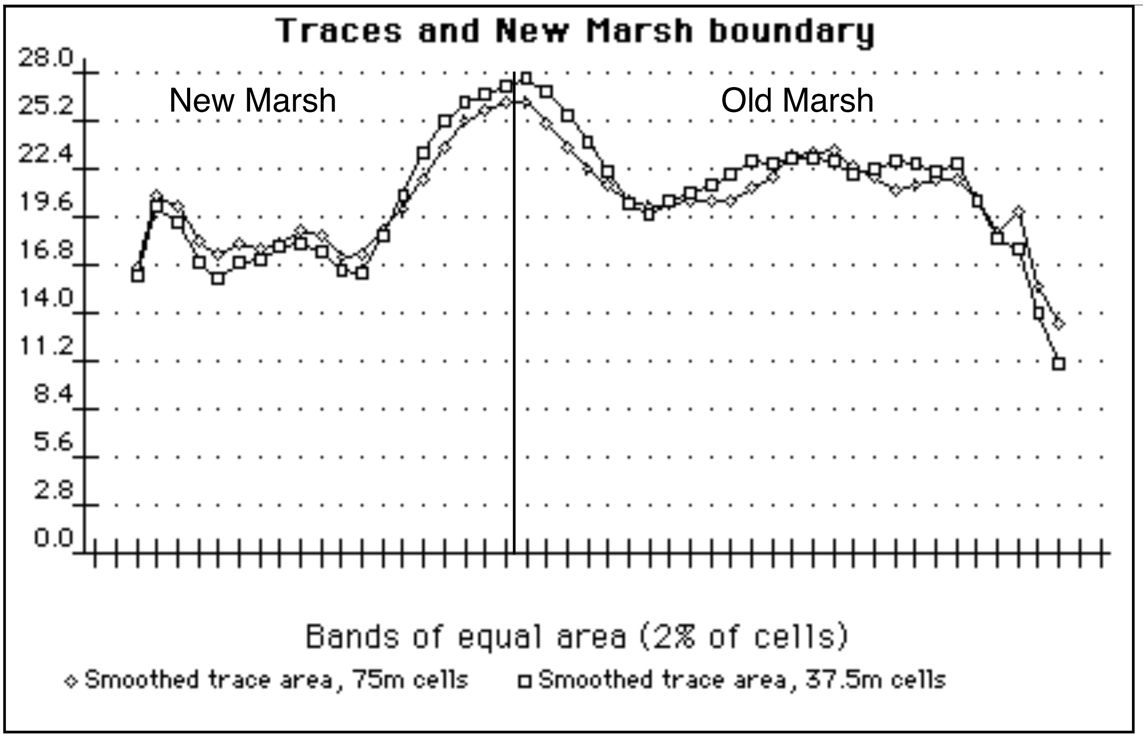


Figure 4.23 Comparison of distribution of raster cells in sample of traces for different raster cell sizes.

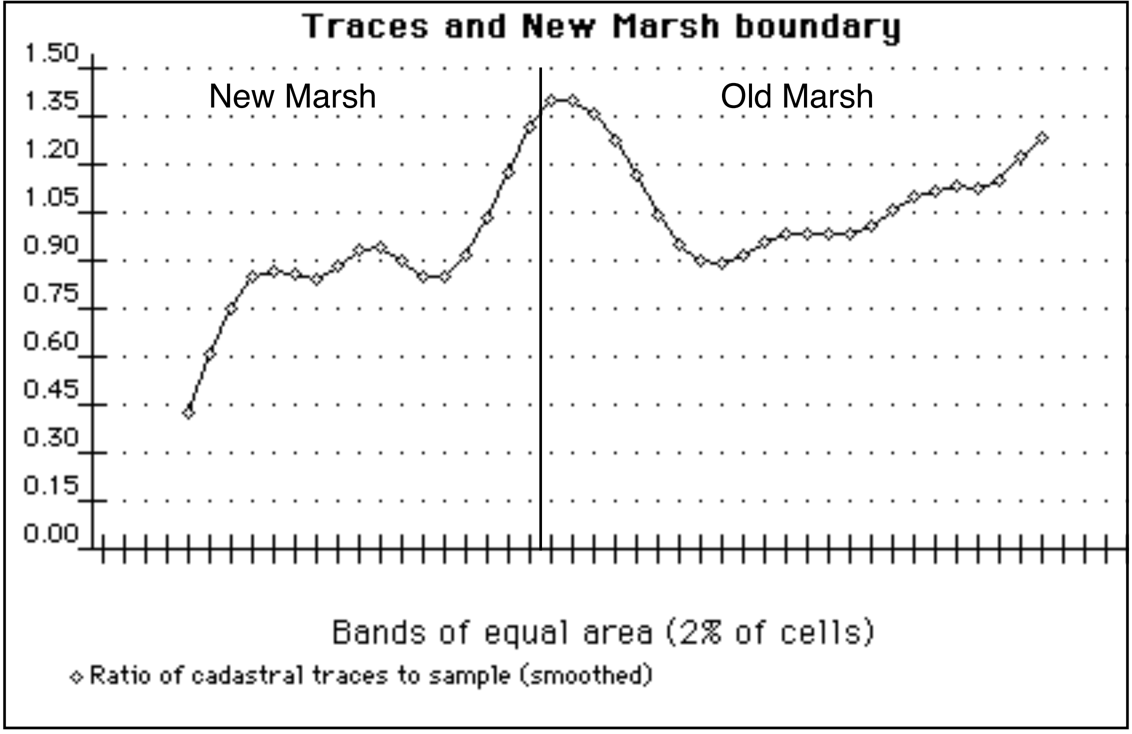


Figure 4.24 Ratio of cadastral traces to a sample of all traces for a raster grid size of 37.5m.

grids, we find that there is not a large difference between the two *(figure 4.23)*. As expected, the peak heights for the distribution using 37.5m cells are slightly higher, but this has little effect on the graph which compares the density of potential cadastral traces to the background of traces as a whole *(figure 4.24)*. It is clear that this conveys much the same information as the graph produced from a 75m raster cell image.

Thus the general conclusions remain unchanged. The distribution of potential cadastral traces is not typical of the distribution of traces in the Marsh as a whole, and has anomalously high values apparently associated with the boundary between Old and New Marsh. One possible explanation for this association is that it was these parts of the cadastral structure which became fossilised most rigidly by the construction of sea defences at a time when the New Marsh was under water, but the Old Marsh was not.

4.2.3 The location of sites with Roman building material in the South Norfolk 'A' cadastre

A strength of GIS software is the ability to combine information from different sources in an objective way. So, as part of the continuing investigation of the South Norfolk 'A' cadastre (Peterson 1988b), it appeared that some potentially interesting results might be obtained by combining data for some particular class of Romano-British (RB) site with data for modern soil classifications. As it turned out, this project could be seen to be impractical. This may have implications for the more general use of GIS software, particularly if it is going to be used with only those parts of sites and monuments record (SMR) data which are machine readable.

One source of data was the Soil Survey map of Norfolk at a scale of 1:100,000. The other was the Norfolk Archaeological Unit SMR. At the start of this particular project (1991) the latter were about to be converted from a system using Superfile[88] on a micro computer to one based on the National Archaeological Record (Hart and Leech 1989) using Oracle database software. It seemed that if the distribution of soil types were converted to a file which could be processed with GIS software it would be possible to answer questions about the distribution of Roman material in an objective fashion.

The significance of the term "objective" will be considered further below *(7.2.2)*, but provisionally we can take it to imply an approach which minimises the personal bias of the interpreter of the data, and which also convinces others that this bias has in fact been minimised.

If others are to be convinced, it will probably be for two reasons: that the data are collected by independent witnesses and that they are treated numerically. Both these features are present here; the soil map and the SMR are independent of each other, and they can be processed numerically by GIS and other software.

---

[88] Superfile has been a system widely used for SMR data (Chadburn 1989).

Accordingly the area of South Norfolk District lying to the east of a north-south line with x coordinate (easting) of 613000m[89] was digitised from the soil map. However, since a large file had been created with many more arcs than in the case of Romney Marsh, and any software to convert them to IDRISI polygons remained unavailable, they were not converted. It was decided to wait until the computerised SMR was ready for processing.

The author was able to extract several data sets from the Oracle-based SMR newly established in the Spring of 1992. One aim was to test alternative hypotheses about the choice of location of new sites in the Roman period. It was felt that the most definite of these would be sites with evidence of substantial Roman buildings, particularly those with mosaic floors and heating systems Can we see if are they related to the cadastre or to soil type or, possibly, to both?

The sites were selected from the database by a query which initially restricted the area searched to that covered by the digitised portion of the soil map. Then Roman sites were selected which had evidence of any of "tile", "roofing tile", "box tile", "tessera" and "building material"[90]. This produced a total of 80 records. However, in the main part of the cadastre as previously published *(figure 5.2)*, excluding the doubtful area on the east, there are a total of 30 such sites. None of them is on an Iron Age site.

---

[89] This is the western border of kilometre squares number 13 on map sheets TM and TG .

[90] This is not in fact a very sensible definition of the substantial Roman buildings which are being sought. As Andrew Rogerson has pointed out (1992 pers. comm.) it excludes quite reasonable Roman sites which yield no hint of building material and it may include sites where the tile or brick is used only as hearth base, in an oven or for some other secondary non-architectural function.

It was decided to exclude from consideration the 14 'urban' (and suburban) sites *(figure 4.25)* which cluster in and around *Venta*. They are not individually relevant to soil type, but to the importance of the town. Thus, in considering an association with soils and/or the cadastre we have only the 16 rural sites with building material, of which 8 have evidence of higher status luxuries in the form of heating systems and/or mosaic floors.

	Any building material	Tesserae or box tile
Urban	14	7
Rural	16	8
Total	30	15

Fig. 4.25 Numbers of RB sites with building material in South Norfolk 'A'.

This is a very small number for an area of about 250km^2, and much smaller than expected. Moreover, we cannot take it to be a fair reflection of the number of buildings present in the Roman period. A prolonged campaign of field walking in a nearby area (Davison 1990) reveals a completely different picture. Only 8km east of Brooke the SMR has data for 16 sites with building material in 9km^2. Eight of them have box tiles or tesserae, giving a density of such sites which is apparently 28 times that seen in the part of the cadastre which we are considering. This is such a large difference that it seems likely that it is due to the relative intensity of research, and one has to wonder if the cluster of sites around *Venta* may also be partly due to the interest aroused by the obvious presence of a Roman town.

The site data are thus very few in number, and are not guaranteed to be representative. Also, because this computational treatment of the SMR data does not include consideration of the secondary files held on paper, they have a more troublesome characteristic.

In three cases where the author had gained information about sites from sources other than the computerised SMR file, it appears that the database information may be misleading if taken in isolation from the secondary files. One site is 1km south west of Newton Flotman with RB building material. The SMR coordinate is that of the possible church at one end of a wood, but the details on paper make it clear there is a dispute over the find position of the RB tile.

Earlier information makes it possible that it was built into the church. A later claim is that it was found at the other end of the wood, near the river Tas, 150m away.

Another site is that at Saxlingham. The coordinates of this site on the SMR are equivalent to TM 2360 9770. However, the principal concentrations of material are at three points, 2350 9765, 2362 9760, and 2360 9750[91]. The 100m square surrounding the SMR coordinate does not come within 50m of any of these points, nor is it an approximation to the position of the centroid of the group.

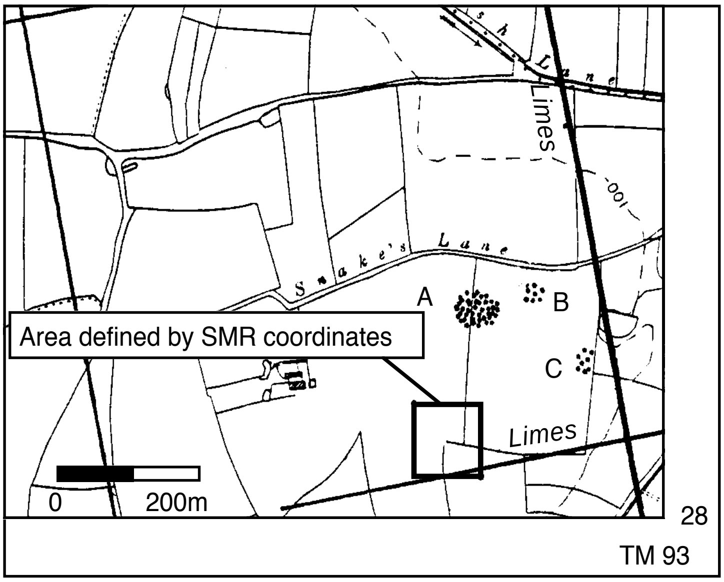


Figure 4.26 Topcroft. Romano-British site with building material.

Thirdly we have the Topcroft site *(figure 4.26)*. The area around this was walked by the author and Dr R. Hadman on 8 March 1992. In one small area, at A, a dense scatter of building material was found, with much box tile[92]. This confirmed the position of a substantial Roman building, as reported by Mr and Mrs Bell in 1984. Two other small scatters including undateable possible roof tile, B & C, were also seen. None of these three scatters is near the 100m square surrounding the SMR coordinate. This latter area was walked carefully, except for its south eastern quadrant which was inaccessible since it lay in the next field and no artifacts were found. In this case the position of the point specified in the SMR is not in the centre of a field, so does not seem to be designating an area which includes the RB material.

---

[91] Information from Mrs Mary Muir (1992 pers. comm.).

[92] The author picked up three palm-sized fragments in five minutes.

In these last two cases the secondary SMR files hold information which is essentially that given here (Rogerson 1992 pers. comm.) so a study of it would correct any misleading impression given by the computerised data. The problem is that this would inevitably compromise the objectivity of the exercise, since, in certain cases, decisions would have to be made to use coordinates other than those on the SMR database.

The situation is similar for some of the sites in the area to the east surveyed by field-walking. There are instances of a lack of correspondence between the square defined by the SMR coordinates and the location of finds of RB building material, as published (Davison 1990: fig 6).

Here *(figure 4.27)* the sites with building material in a 3x3km square immediately south of Loddon are shown in black. They were selected according to the tabulated account of finds (Davison 1990: 77-88), to include all those with tegulae/tiles, imbrices, bonding tiles, flue tiles or box tiles. Their positions were then traced from Davison's figure. The larger black areas represent concentrations of material and round dots indicate 1-10 finds. The author's interpretation of Davison's mapping conventions (1990: 12) is that these dots, all being away from the centre of the field in which they occur, represent finds in one particular area, rather than indicating a general scatter in the field.

The SMR coordinates were used to define squares surrounding them equally on all sides, to a certain precision. If the given coordinates were both rounded to 100m it was assumed that the precision was ±50m. If one of the coordinates was defined to 50m it was assumed that the precision was ±25m. These assumptions had to be made in the absence of any knowledge of the accuracy intended by those who input the data.

There are some incompatibilities between the two representations of the find distribution. SMR number 13009 (sites 17, 18) seems to have no RB building material listed in Davison's tabulated account of finds, and 17809 (site 31) has no RB material listed at all.

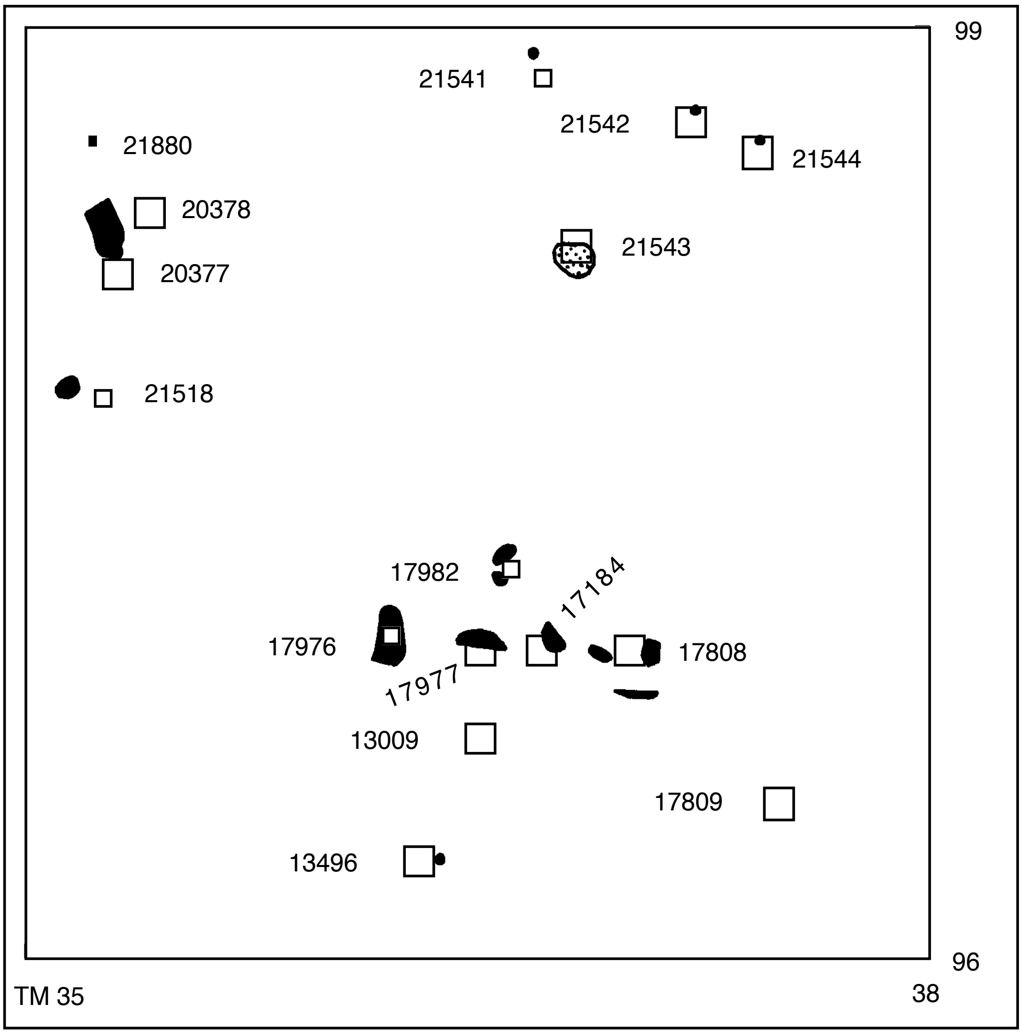


Figure 4.27 Area of 9km² south of Loddon: RB sites with building material compared to SMR coordinates.

Several SMR coordinate squares, SMR numbers 21518, 20377, 20378 and 21541, do not cover or intersect the areas of finds. This is because the coordinates of their centre are coordinates for the centre of the field, rather than the site.[93]

These facts make one wonder if the computerised SMR data are suitable for the hypothesis testing that was originally envisaged; for

---

[93] Information from Andrew Rogerson (1992 pers. comm.). Again, inspection of the secondary SMR file would allow the computerised data to be interpreted, including those for number 17809 which was mis-classified on data entry.

the area of the South Norfolk 'A' cadastre we find that its quantity and precision are limited. A comparison of sites with soil types, which would not require a close description of site location, would be of little value because there are so few sites in the class defined (or perhaps mis-defined) by the database search. A study of the relationship (if any) between sites and the cadastre requires a set of data whose grid coordinates are sufficiently precise to allow comparison with the cadastral grid of about 710m. Such a set of data could not be obtained unless the originally envisaged objective approach is compromised. In some cases it would be necessary to use the secondary SMR files, with personal knowledge and judgement, to modify the "objective" data provided on the computer files.

Perhaps there is a lesson to be learnt. The treatment of data cannot normally be approached mechanically and blindly. Although it may be the case that for large collections of data we will find that errors and distortions will cancel out, this is unlikely for small quantities. The data need to be checked, and this inevitably introduces a clearly recognisable element of subjectivity.

This conclusion raises the question of the validity of the approach to the Limburg data described above *(3.2.1)*. How unreliable are its results?

Even without any information on the methods used to collect and store Dutch site data, it can be suspected that it will share some of the characteristics of an English SMR. One therefore expects that locational information will tend to be generalised, and a look at typical coordinates *(figure 3.18)* shows that this is so.

However, the Limburg data make up a large set (491 items), and we have seen that they are associated with the cadastre in a way which would occur only one in 200 times on a chance basis. The generalised nature of the site coordinates would do nothing to increase the chance of seeing such an association. The opposite must be the case, since a reduction of the precision of site location would tend to make any genuine association less visible. Thus it is remarkable that the Limburg association shows through, despite the inherent vagueness of location of some sites and the restrictions imposed by having to specify a possibly unrepresentative point as their sole reference.

This suggests that, in an ideal world, systems may be developed which hold the area, defined by boundaries or even raster cells, as part of the site data, as envisaged by, among others, Nick Ryan (1992: 4). This would be expensive, since it would involve modification to the National Archaeological Record itself. However, such developments are too speculative to be discussed further here.