A Vector-based Approach to the Integration of Geophysical and Test-pitting Data: Phasing the South Cadbury Environs Project

John Pouncett and Gary Lock

Institute of Archaeology
University of Oxford
United Kingdom
john.pouncett@arch.ox.ac.uk

Abstract

This paper describes a new methodology to date a complex series of ditches and enclosures discovered by geophysical survey around the hillfort of South Cadbury, England. Using Network Analysis, Closest Facility Analysis to be precise, spot dates from small-scale excavations are transferred around the ditch system to establish a phasing. The technicalities of Network Analysis are described together with the iterative procedure gone through to match the logic of the software with the established logic of archaeological phasing. A robust methodology is produced and the initial results described.

1 Background

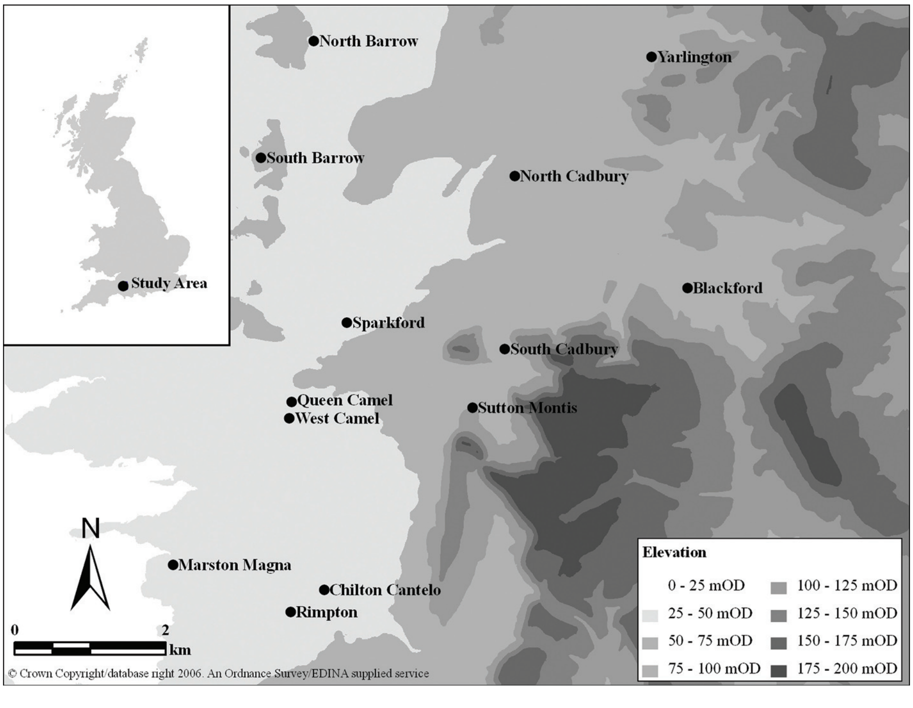
South Cadbury is a medium-sized multi-vallate hillfort enclosing seven hectares situated on an isolated hilltop overlooking the rolling countryside of central Somerset, England, as seen in Figure 1. It has been the focus of various fieldwork campaigns and their resulting publications over many decades, although initially it was known primarily due to the post-Roman activity there and possible connections with King Arthur (Alcock 1972, 1995).

Alcock's extensive excavations within the hillfort between the years 1966 and 1970, a radical departure from the traditional focus on ramparts, revealed a wealth of evidence for activity in the 5th and 6th centuries AD but also for occupation during the later prehistoric and Romano-British periods. These have been more recently published by Barrett et al. (2000) so that the evidence for the full chronological range of activity on the hilltop is now available. Based mainly on ceramic evidence, the more recent work proposes three chronological phases for the site:

- Early Cadbury, circa 1000-300 BC, occupation from the Late Bronze Age, the first phase ramparts;
- Middle Cadbury, 300 BC-AD 40/50, the main hillfort occupation, extensive outer ramparts and the enhanced south-western entrance;
- Late Cadbury, AD 40/50-400, reduced activity, a possible Roman attack, possible Roman barracks.

In an attempt to situate the hillfort and the excavations within a landscape context, the South Cadbury Environs Project (SCEP, http://web.arch.ox.ac.uk/~scep/home.php) was initially funded by the Leverhulme Trust and more recently, 2004 to 2008, by the Arts and Humanities Research Council (AHRC) through a joint project based at Bristol and Oxford universities. The methodology employed uses a range of techniques within an area of 64 sq. km centered on the hillfort so that large-scale geophysical survey (Figure 2) is ground-proofed by targeted test-pitting enhanced through occasional larger-scale excavation (Tabor and Johnson 2000).

This extensive fieldwork has produced a wealth of detailed information in the form of complex patterns of ditches and enclosures from the geophysics and pottery from test pits and excavations. A major focus of SCEP has been the integration of



© Crown Copyright/database right 2006. An Ordnance Survey/EDINA supplied service

Figure 1. The location of South Cadbury, Somerset, England.

these two sets of data to develop a detailed chronology from the pottery that can be used to phase the features from the geophysics and thus develop an understanding of landscape evolution. A novel methodology based on the use of GIS-based (Geographic Information Systems) Network Analysis (NA) was a part of the AHRC funding and its provisional results are reported here.

1.1 The Sigwells Data-set

The large field to the south of the hillfort, known as Sigwells (Figure 2), has produced particularly good results from geophysical survey (Figure 3) and has consequently been one important focus for test pitting and excavation.

This data-set has already formed the basis for a preliminary phasing based on traditional non-GIS methods (Tabor and Johnson 2000), and a portion of the data, Sigwells (West), is used here as a comparison for the NA techniques. The preliminary phasing resulted in six systems (phases) that were identified on the basis of the geophysics and excavated data, as follows:

System 1 – AD 200-AD 400
System 2 – 100 BC-AD 200
System 3 – 100 BC-AD 100
System 4 – 2nd Millennium BC
System 5 – AD 300-AD 400
System 6 – AD 300-AD

This shows two major trends within the ditched features seen in Figure 4, one running WNW to ESE and belonging to System 1 (Middle Romano-British) and the other NE to SW dating to System 2 (Late Iron Age and Early Romano-British).

Ceramic phasing based on continuing excavation since the publication of the preliminary phasing, however, has produced the currently accepted eight-phase chronology as follows, with equivalences to the preliminary phasing (Tabor pers. comm.):

Phase 1 - Neolithic (NEO)
(no equivalence)
Phase 2 - Early Bronze Age (EBA)
(equivalent to System 4)
Phase 3 - Middle Bronze Age (MBA)
(equivalent to System 4)
Phase 4 - Middle Iron Age 1 (MIA 1)
(equivalent to System 2)
Phase 5 - Middle Iron Age 2 (MIA 2)
(equivalent to System 2)
Phase 6 - Late Iron Age (LIA)
(equivalent to System 3)
Phase 7 - Early Romano-British (ERB)
(no equivalence)
Phase 8 - Middle Romano-British (MRB)
(equivalent to System 1

1.2 The Network Dataset

The network dataset created to represent the archaeology of Sigwells, and as required by ArcGIS, consisted of vector

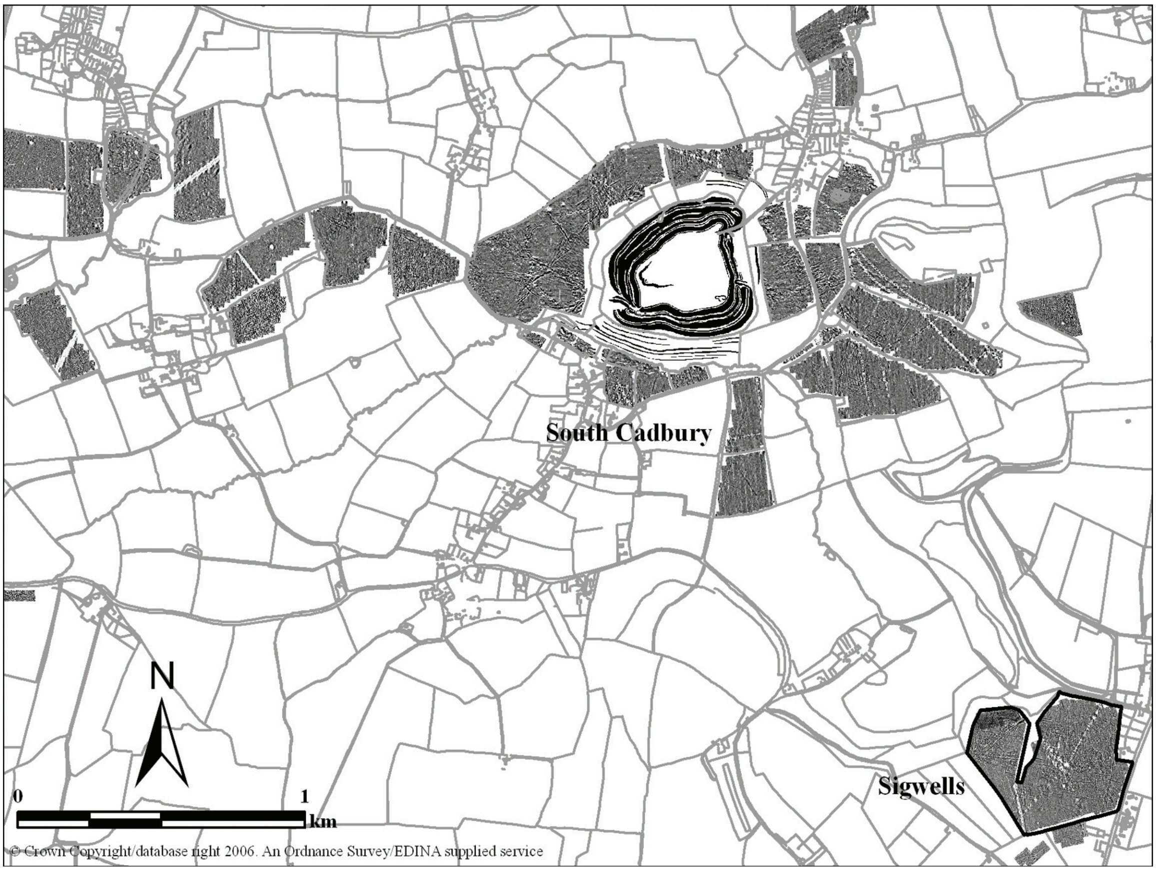


Figure 2. Areas of geophysics around South Cadbury showing the location of Sigwells.

data (points and lines) and encoded turn features. This was created by heads-up-digitizing of the geophysical anomalies as a series of lines or polylines (edges), with nodes (junctions) corresponding to the start and end points of each anomaly or an intersection between anomalies as shown in Figure 5.

A node is required where two edges cross or intersect (planarity), regardless of their connectivity (i.e., the relationships between the corresponding anomalies). The actual connectivity relationships were subsequently assigned to the start and end points of edges as z-values. In effect, this encodes the stratigraphic relationships identified through excavation (Table 1).

Table 1. Attribute data encoding stratigraphic relationships between anomalies.

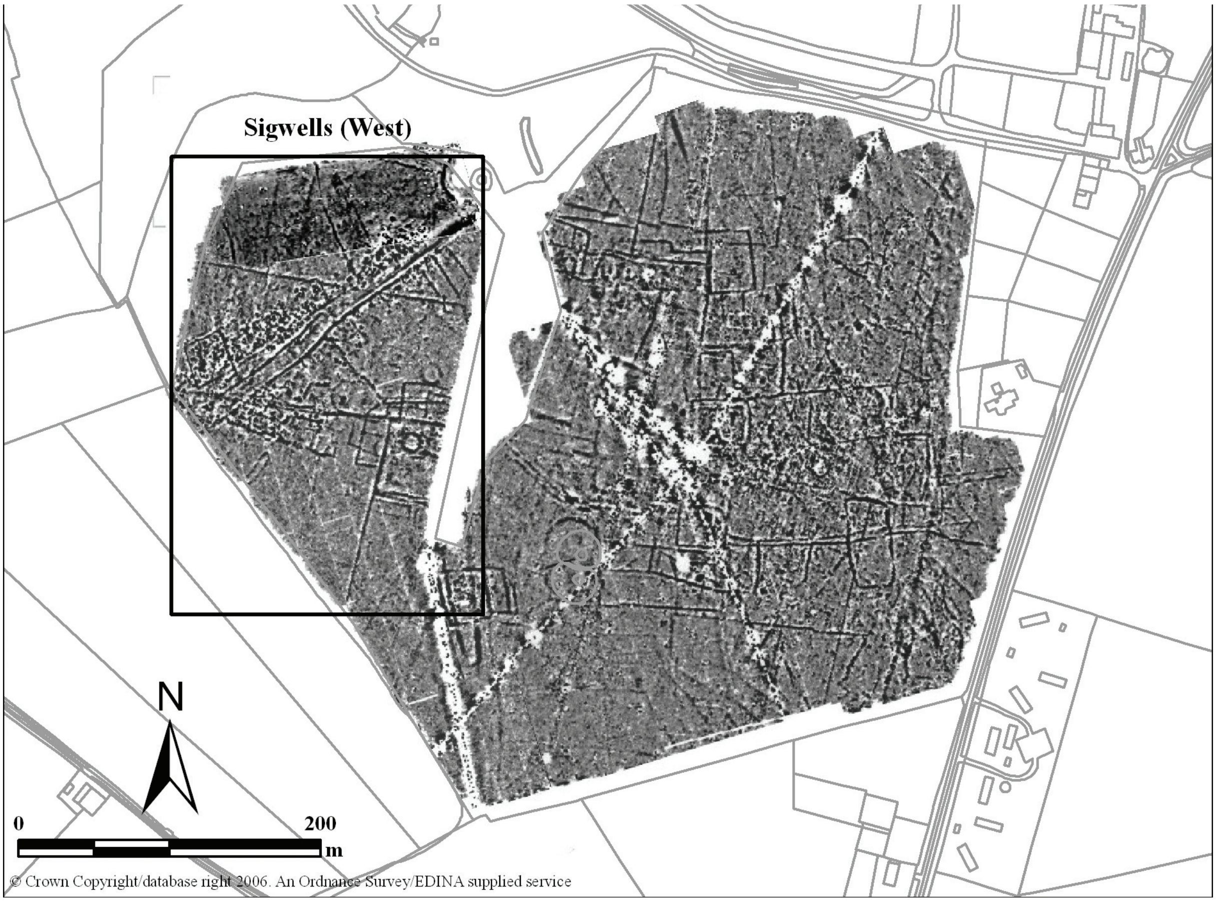
Anomaly	Edge	From Junction	To Junction	Z1	Z2
1	1	781	807	0	0
1	2	807	838	0	0
4	29	798	807	0	0
119	129	853	838	0	1
119	130	838	841	1	0

The resulting dataset was then cleaned by correlating the geophysical anomalies with excavated evidence where possible (Figure 6) involving the joining of gaps, for example entrances through ditches, and the linking of interrupted anomalies.

Impedance values were then assigned to edges to reflect the ease of movement along that length of the network. In network solutions that seek to find the shortest route between two points, as here, the length of an arc must be used as an impedance. At this point the network is built with its elements created, connectivity established and impedances defined.

It is important to state here that "movement" around the network is not actual physical movement but the extrapolation of spot-dates from test pits around the non-excavated sections of ditch as discovered by geophysics. We are interested in establishing how far a known date can be applied to connecting ditches (i.e., can move) until a point is reached whereby the date is in conflict with another being extrapolated from another start point or a previously assigned later date. Using this very specific adaptation of NA, the aim is to phase the whole complex of ditches.

At each node of the network where movement through it involves a choice of turn directions, these need to be defined by creating turn feature classes. Turns were encoded as a series of polylines, each linking a maximum of two edges (Edge 1 and Edge 2). A polyline, with a start node on Edge 1, intermediate vertices on Edges 1 and 2, and an end node on Edge 2, is created for every possible turn for each junction. Edges ranged between 1 m and 40 m in length, with the majority between 10 m and 25 m. The number of turns defined for a junction is directly proportional to the number of intersecting edges:



© Crown Copyright/database right 2006. An Ordnance Survey/EDINA supplied service

Figure 3. The geophysics of Sigwells.

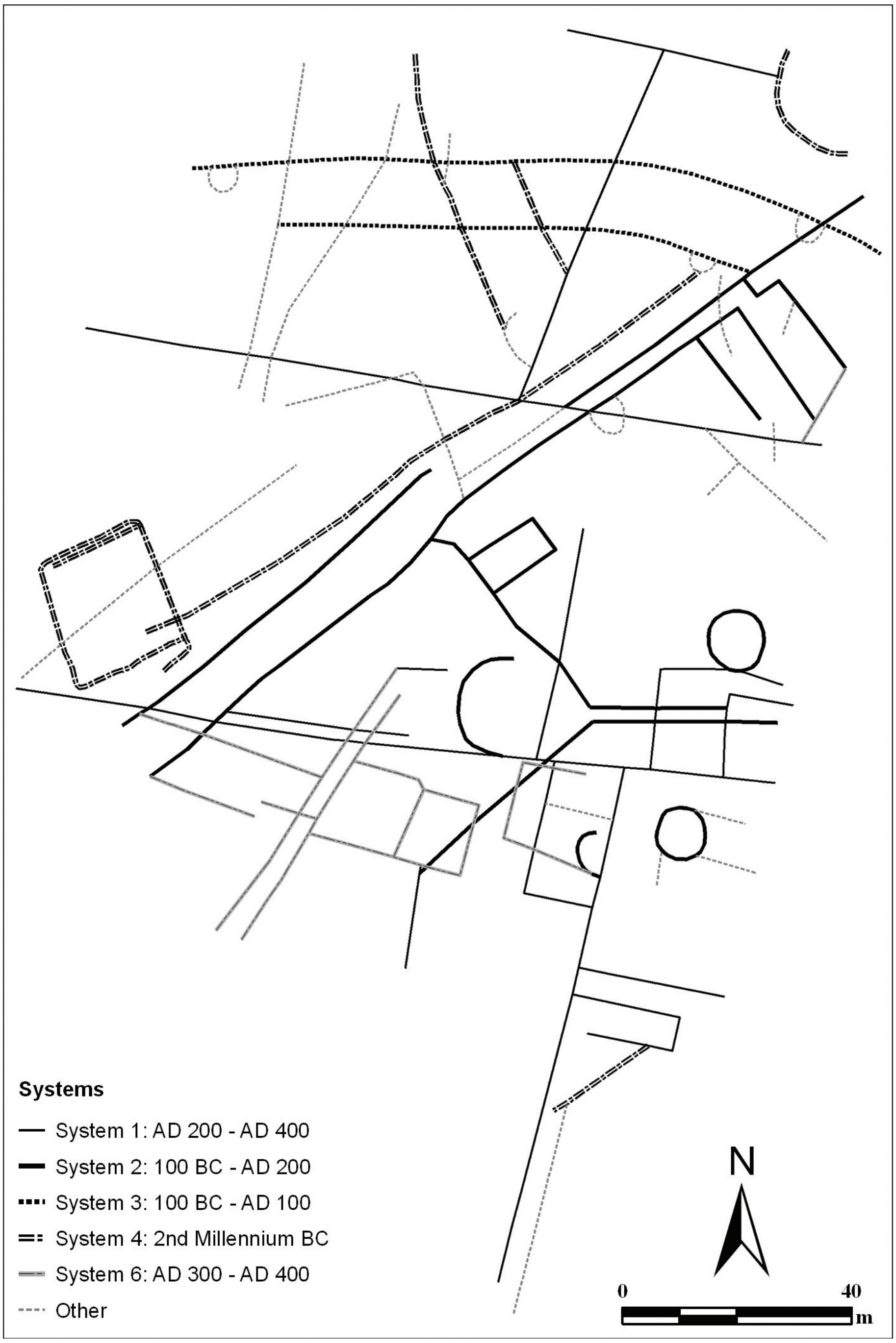


Figure 4. The preliminary phasing of the features within Sigwells (West).

$$T = n(n - 1) \quad (1)$$

where: T = number of turns
n = number of edges

An impedance can then be assigned to every turn to reflect the ease of movement from one edge to another. The actual physical relationships were encoded by assigning impedances to reflect the probability that edges were contemporary with one another. This is based on the assumption that features that are parallel to or perpendicular to one another are more likely to be contemporary than those which intersect at an acute or obtuse angle. This is supported empirically and gives rise to the two directional trends as shown in Figure 4. An impedance was assigned to each turn in the range of 0 m to 25 m (Table 2); once all impedances have been assigned the network is rebuilt.

Table 2. Impedances assigned to model physical relationships between anomalies.

Angle	Impedance
± 000-010º	0m
± 010-080º	10m
± 080-100º	5m
± 100-135º	15m
± 135-170º	20m
± 170-180º	25m

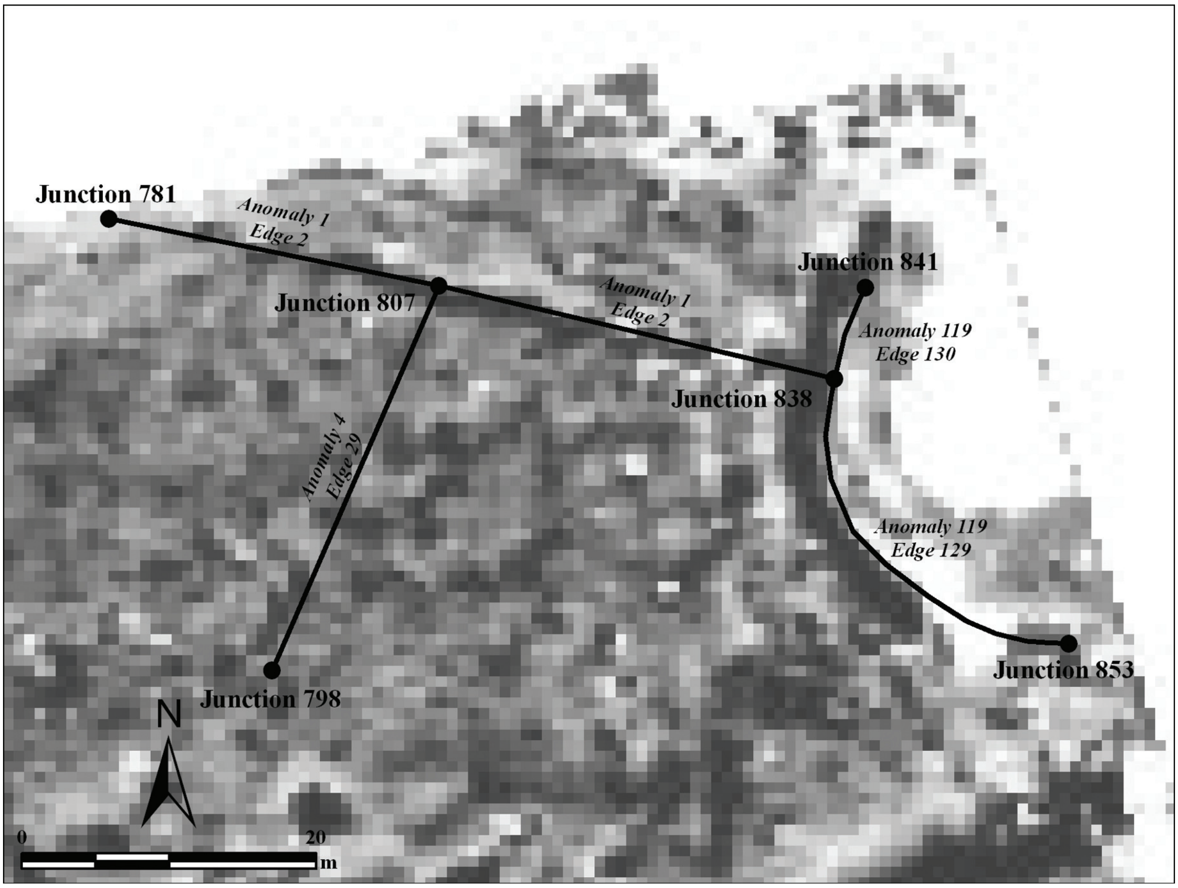


Figure 5. Producing vector data from the geophysical survey data.

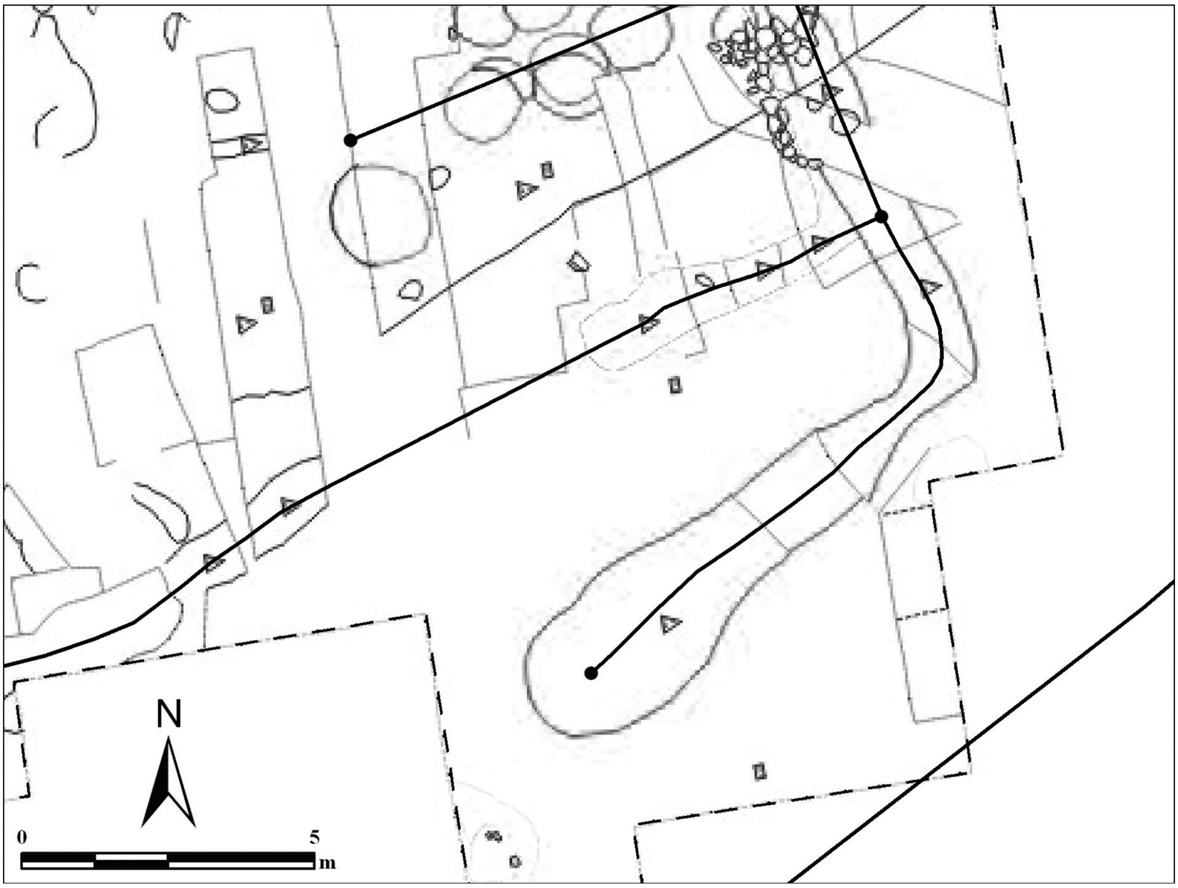


Figure 6. Correlating geophysical anomalies with excavated data.

The next, and final, step in constructing the network dataset is the recording of certain network locations. This comprised point data for the centroids of archaeological features or deposits that had been dated as a result of trial trenching or test pitting. Two separate point feature classes were created for each ceramic phase, one for archaeological features and deposits that were contemporary with one another and the other for those that corresponded to later phases. An additional feature class was created for nodes corresponding to the start and end points of each geophysical anomaly.

2 Developing the Methodology

Within NA there are various types of analysis that can be used to model flow through the network (Conolly and Lake 2006; Fischer 2004). The challenge for any archaeological application, however, is adapting one of these to suit one's needs, and here we use Closest Facility Analysis. This establishes the shortest route to every "facility" from each "incident," avoiding known "barriers" on the way, with the result being the so-called network solution. Whereas usual applications of this are shortest routes from emergency service stations to various sorts of incidents, we use it here to move known spot dates around the network of ditches in an attempt to see if the date can be extrapolated to lengths of ditch of unknown date. The "facilities" are the start/end points of the network and of the flow, "incidents" are points on the network where spot dates are known from excavation, and "barriers" are where the flow stops if a later spot date is encountered. This last point is based on accepted archaeological logic that a feature is dated by the latest material contained within it. It follows, therefore, that the later the spot date the fewer barriers it is likely to encounter as it moves around the network (i.e., the fewer later dates), so movement around the network is less constrained for later periods.

This is operationalized through a composite layer comprised of four individual layers, each of which represents a separate network feature:

1. facilities feature layer – the network locations used as facilities, i.e., network locations (system junctions) corresponding to the start or end points of geophysical anomalies (end points);
2. incidents feature layer – the network locations used as incidents, i.e., network locations (point features) relating to a particular ceramic phase (start points);
3. barriers feature layer – the network locations used as barriers, i.e., network locations (point features) corresponding to later ceramic phases;
4. routes feature layer – closest facility network solutions, i.e., the shortest route between each of the incidents and facilities.

Various sorts of network solutions can be generated, which are described here together with aspects of network logistics:

1. individual solutions for each ceramic phase, starting with the most recent phase of activity (Phase 8: Middle Romano-British) (Figure 7);
2. solutions for more than one incident, i.e., where more than one archaeological feature or deposit can be assigned to a particular ceramic phase, will determine the shortest route to any given facility from each of the incidents;
3. the maximum number of routes in a network solution is directly proportional to the number of incidents:

$$R_{\max} = IxF \qquad \textbf{(2)}$$

where: R_{max} = number of facilities
I = number of incidents
F = number of facilities

4. no route is created between an incident and facility if all possible routes are blocked by barriers;
5. the number of edges without network locations that can be assigned to a particular ceramic phase remains constant;
6. the number of barriers, network locations corresponding to later ceramic phases, decreases and movement around the network becomes less constrained;
7. edges with network locations (point features) corresponding to earlier ceramic phases can be (re)used in network solutions for later phases.

3 Network solutions—Towards a Phasing

The methodology developed here was based on an iterative process of introducing refinements so that the Closest Facility network solutions moved towards matching the archaeological logic. The initial solutions were obtained using an unconstrained network without correction of planarity or assigning impedances to turns. For later solutions the stratigraphic and physical constraints imposed by the archaeology were introduced, once an effective methodology for combining solutions had been established.

3.1 Provisional Phasing

For the initial solution, individual ceramic phase solutions were overlain, one on top of another, starting with the most recent phase. The resulting phase diagram indicates the earliest ceramic phase for which each edge is used (Figure 8a). This is confusing, however, and can be simplified by aggregating the overlain solutions and assigning dates to individual anomalies on the basis of the dominant ceramic phase (Figure 8b). Whilst this was simpler, several problems were encountered. Firstly, aggregation of the solutions based on the dominant or modal ceramic phase generated a large number of conflicts, i.e., anomalies that could be assigned to more than one ceramic phase. Secondly, the aggregated solution places undue emphasis on the earliest ceramic phase (Phase 1: Neolithic).

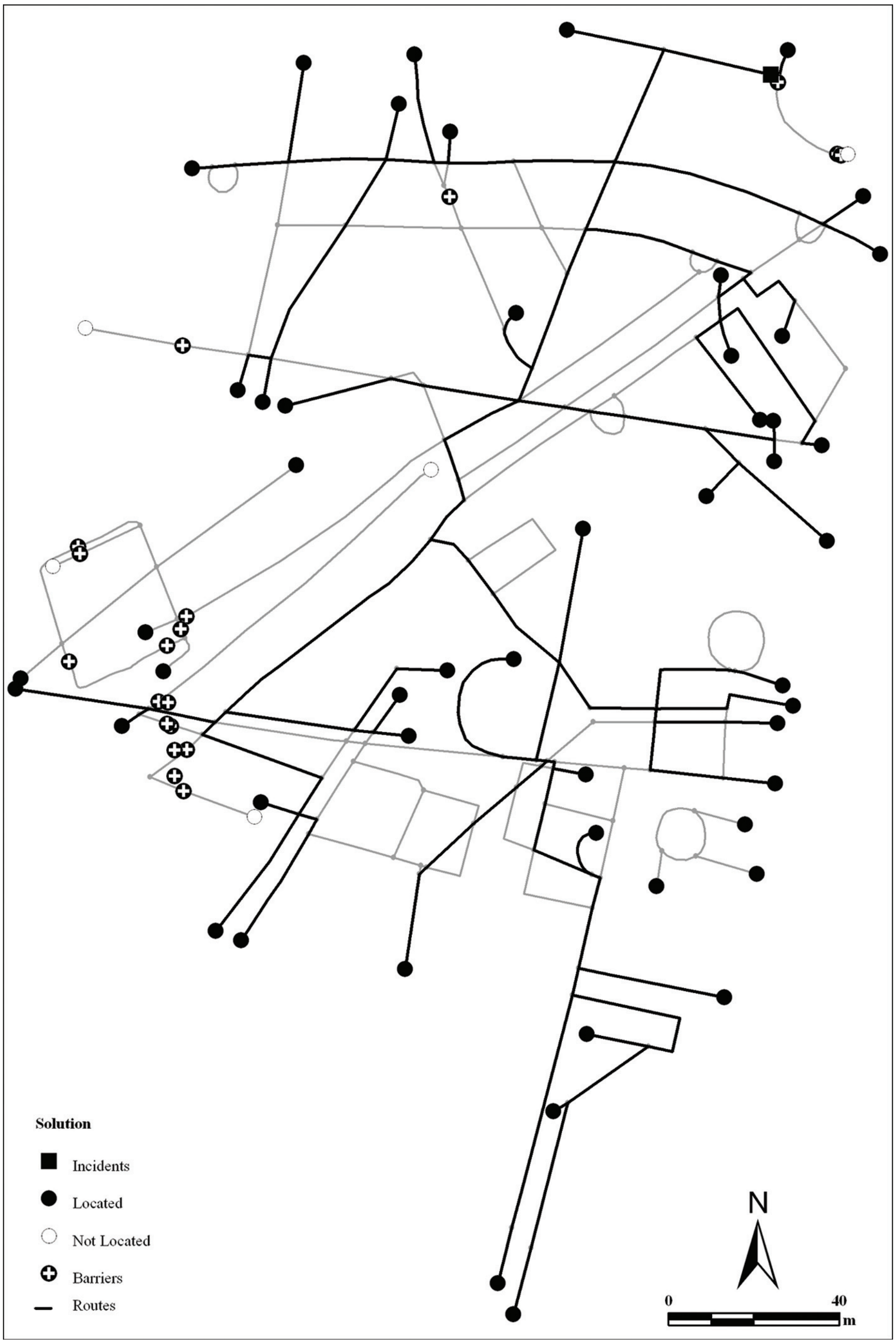


Figure 7. Network solution for Phase 1 (NEOLITHIC).

3.2 Revised Phasing

An alternative approach was consequently adopted, counting the number of times an edge was used in the network solution for each ceramic phase and aggregating the resultant counts to identify the dominant phase for each anomaly (Figure 8c). Whilst this approach succeeded in redressing the balance towards the later Iron Age and Romano-British activity, the phasing of geophysical anomalies on the basis of raw counts was skewed towards the phases with the highest number of spot dates, i.e., Phase 6: Late Iron Age. To overcome this, the counts for each ceramic phase were standardized on the basis of the number of routes in the corresponding network solution:

$$N = 100x\frac{R}{A} \tag{3}$$

where: N = standardized count
R = number of routes in solution
A = number of times arc used

The standardized counts indicated the percentage of routes in which an edge participated for the solution of each phase. Percentages for edges forming part of the same anomalies were averaged and dates assigned on the basis of the dominant phase (Figure 8d). It can be seen that the range and relative frequency of the ceramic phases in the resultant phase diagram is broadly consistent with the preliminary

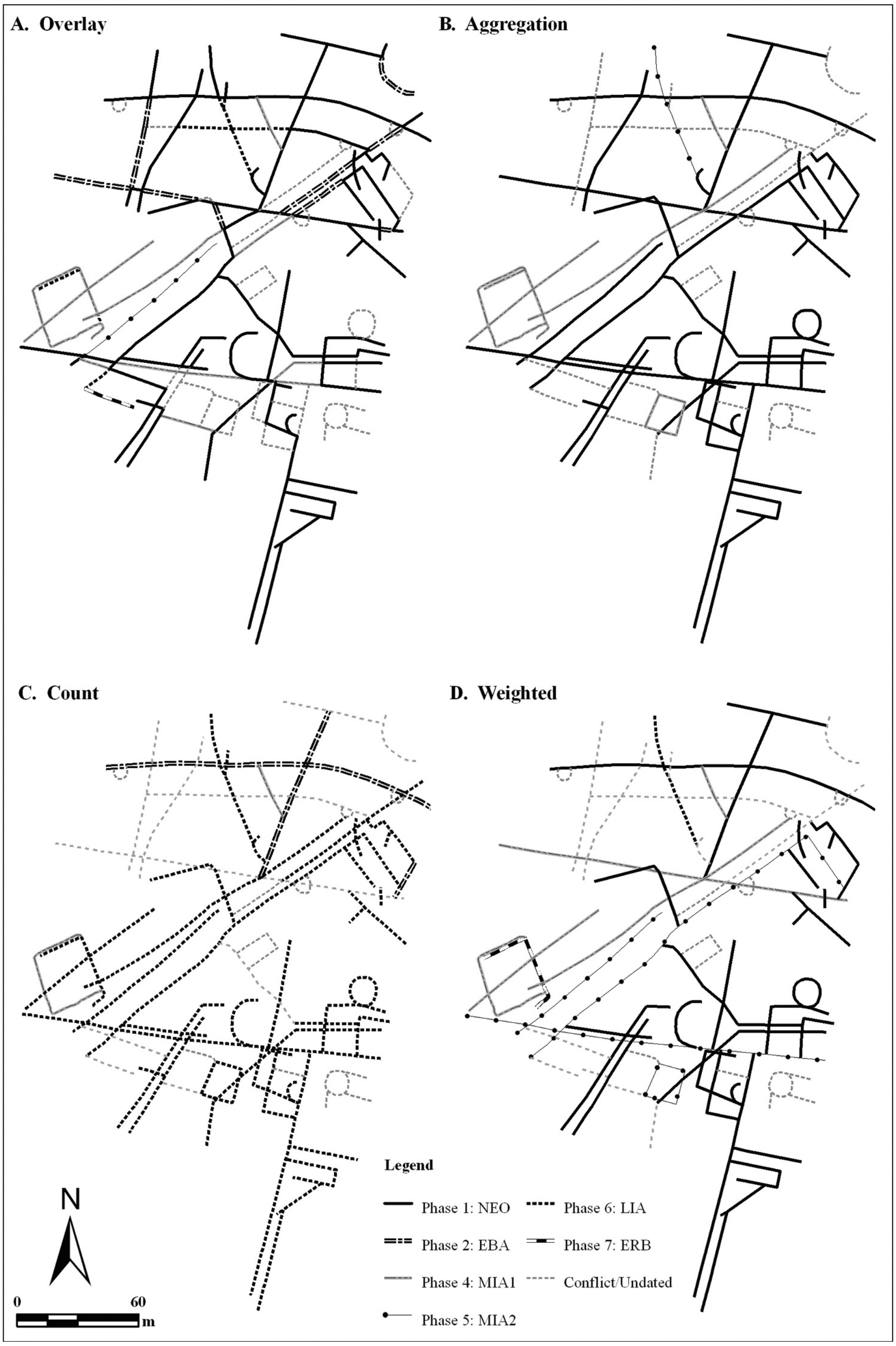


Figure 8. Methods of combining individual network solutions.

phasing suggested by Tabor and Johnson (2000), and we consider this method of combining Closest Facility network solutions for individual ceramic phases to be fairly robust.

3.3 Comparative Phasing

It is important to incorporate into this process any existing dating evidence, particularly relative dating in the form of stratigraphic and physical relationships between ditches. This was established once the methodology for combining solutions from individual ceramic phases had been refined. Stratigraphic relationships between ditches in the form of "cuts/cut by" have been established through excavation to provide localized relative sequences. Physical relationships are based on the accepted understanding that contemporary ditches tend to be parallel or perpendicular to each other to create systems of boundaries, enclosures, and trackways. This results in the angles between contemporary ditches being within predictable ranges of value.

The combined solution for the Closest Facility analysis identified both of the major trends within the ditched features identified on the basis of the preliminary phasing (Figure 9). Furthermore, the phasing of the rectangular enclosure at the western edge of the site indicated by the network analysis

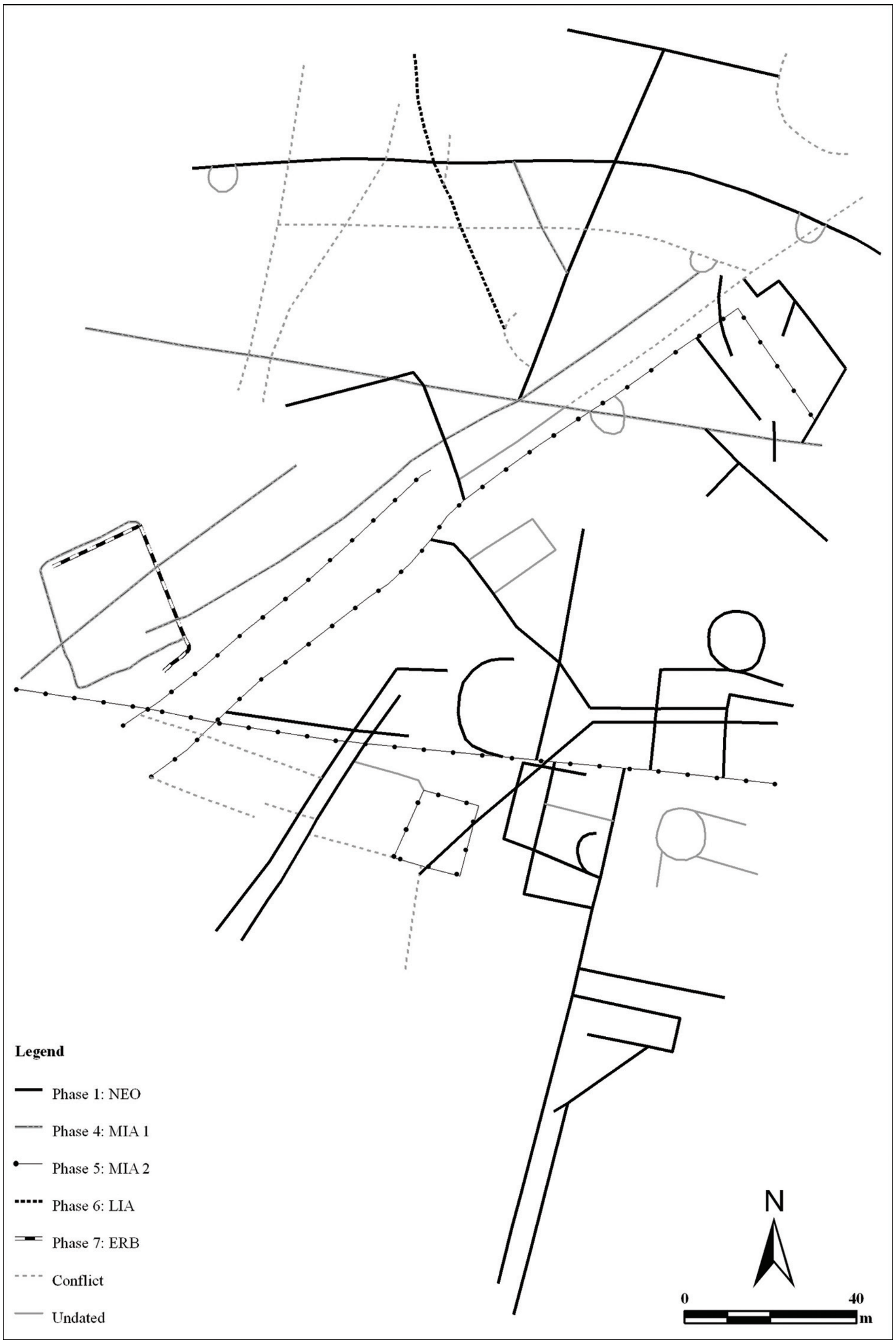


Figure 9. The network based phasing of the features within Sigwells (West).

is consistent with revised dating suggested by further excavation (Tabor pers. comm). Also, using this method only a small number of conflicts were identified, eliminating the need for manual intervention.

4 Conclusions

Although this work is as yet in its early stages it is possible to draw some conclusions. A major challenge lies in the area that we can call "logic matching," in this case matching the logic of formal NA as systematized in the commercial software to the archaeological analyses in which we are interested. This is not a new problem, especially in GIS applications in archaeology, which has seen a lot of work with, for example, visibility and movement that has required creative development by archaeologists to move the software on from its original intended applications and functionality. Despite this, we have shown that NA is useful as an iterative tool to highlight inconsistencies in the phasing of systems of ditches and in trying to interpolate spot dates.

Our methodology has developed incrementally as described above within an intuitive framework that integrates traditional dating logic with NA functionality. There is still some ways to go before we have a robust dating tool that can be applied to the whole of Sigwells and, eventually, to other areas of geophysical survey around South Cadbury. As a next stage we will experiment with assigning impedance to each arc and with methods of incrementing impedance after each incident and each phase solution.

References Cited

Alcock, L. 1972. *'By South Cadbury, is that Camelot…' Excavations at Cadbury Castle 1966-70.* London: Thames and Hudson.

Alcock, L. 1995. *Cadbury Castle, Somerset: The Early Medieval Archaeology*. Cardiff: University of Wales Press.

Barrett, J. C., Freeman, P., and Woodward, A. 2000. *Cadbury Castle, Somerset: The Later Prehistoric and Early Historic Archaeology*. London: English Heritage.

Conolly, J. and Lake, M. 2006. *Geographical Information Systems in Archaeology*. Cambridge: Cambridge University Press.

Fischer, M. 2004. GIS and network analysis. In, *Handbook of Transport Geography and Spatial Systems*. Hensher, D., Button, K., Haynes, K. and Stopher, P., eds., pp. 391-408. London: Elsevier.

Tabor, R. and Johnson, P. 2000. Sigwells, Somerset, England: regional application and interpretation of geophysical data. *Antiquity* 74:319-25.