

# 01 | RECORDING AND ANALYSIS OF FIELD DATA

John Pouncett – Gary Lock

## Closest Facility Analysis: Integration of Geophysical and Test-Pitting Data from the South Cadbury Environs Project

*Abstract:* Together with a paper presented at the 2006 CAA conference, this paper describes an innovative GIS-based methodology for dating systems of ditches identified by geophysics around the hillfort of South Cadbury, England. It uses Closest Facility Analysis (a form of Network Analysis) to distribute spot-dates around the system from start points established by excavation. An important part of the work is the manipulation of the logic inherent within the software to incorporate the required archaeological logic of stratigraphic relationships and intercutting features. The resulting methodology is shown to be robust by testing its results against those from simulated datasets. Overall, the approach is iterative and highlights problems and conflicts within the dating which need user input to resolve.

### Background

This paper is the culmination of a programme of research carried out as part of an AHRC funded project at the Universities of Bristol and Oxford under the auspices of the South Cadbury Environs Project<sup>1</sup>. It extends and tests the robustness of a methodology for the integration of geophysical and test-pitting data developed during the course of the project and presented at CAA2006, Fargo (LOCK / POUNCETT IN PRESS). South Cadbury is a seven hectare multi-vallate hillfort in Somerset, England, perhaps best known for its possible connections with King Arthur (ALCOCK 1972, 1995). Excavations undertaken at the site between 1966 and 1970 – both through the ramparts and within the interior of the hillfort – revealed a complex sequence of activity from the Later Bronze Age until the 5<sup>th</sup> and 6<sup>th</sup> centuries AD. Recent analysis has identified three main phases of activity: Early Cadbury (1,000 BC to 300 BC) – the first ramparts; Middle Cadbury (300 BC to AD 40/50) – the main occupation of the hillfort and Late Cadbury (AD 40/50 to AD 400) – possible (re)use as Roman barracks (BARRETT / FREEMANN / WOODWARD

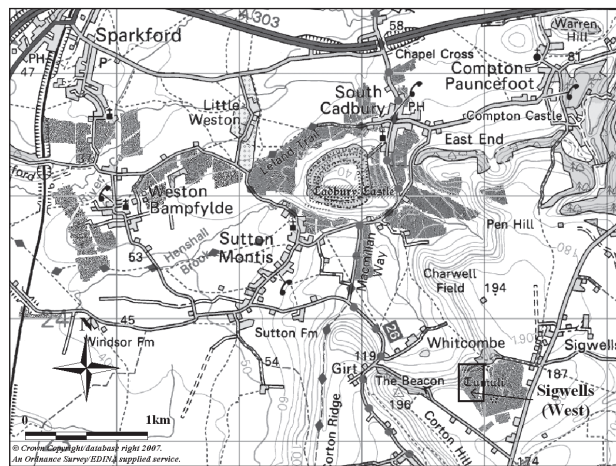


Fig. 1. The location of South Cadbury showing the areas of geophysical survey around the hillfort and the Sigwells (West) case-study area.

2000). The South Cadbury Environs Project aims to place the hillfort within a broader landscape context.

An extensive complex of magnetic anomalies thought to correspond to archaeological features has been identified within an area of 64 km<sup>2</sup> centred on the hillfort through geophysical survey (Fig. 1). Spot dates based on ceramic phasing have been obtained

<sup>1</sup> <http://web.arch.ox.ac.uk/~scep/home.php>

for features corresponding to individual anomalies through test-pitting. A GIS-based methodology has been developed to facilitate the integration of these datasets as part of the ongoing interpretation of the evolution of the archaeological landscape. This methodology was developed using a subset of the data from the South Cadbury Environs Project – part of a complex of probable ditches and enclosures from a large plot of land known as Sigwells for which provisional phasing had been established as an earlier part of the project (TABOR / JOHNSON 2000). Geophysical anomalies and test-pitting data from Sigwells (West) were encoded as a vector-based (Geodatabase) network dataset in ArcGIS 9.1 using the ESRI transportation network data model (LOCK / POUNCETT IN PRESS; see also below for more details).

### Closest Facility Analysis

Dates were distributed around the network dataset using the Closest Facility Analysis (CFA) routine in Network Analyst. The algorithm used in this routine is based on the minimisation of cost where costs are evaluated from attributes assigned to edge elements or junction elements with a network dataset. The length of edge elements is used by default. A

closest facility solution is obtained by calculating the lowest cumulative cost between two points on the network, referred to collectively as network locations. Three types of network location are recognised in Closest Facility Analysis – incidents, facilities and barriers – and more than one instance of each type of network location may be used. The basic principles of CFA are illustrated with reference to a simple network dataset (Fig. 2) comprised of three edge elements (labelled 1, 2 and 3), a junction (not labelled) and two network locations (Locations 1 and 2). Two possible closest facility solutions exist – Route 1 (up and across) and Route 2 (across and up) although only one solution will be obtained for each pair of incidents and facilities; in this instance, a single solution for Locations 1 and 2. The routine returns the first solution identified by the algorithm rather than all of the possible solutions.

If movement about the network is unconstrained, the closest facility solution obtained is dependent upon the order in which the edges were created (Fig. 2a), and is, therefore, dictated by the logic of the used algorithm and not the logic of the underlying network dataset. This has important implications for the application of CFA to the distribution of dates from test-pitting around a network of geophysical anomalies. The default behaviour of the algorithm, however, can be over-ridden. A particular

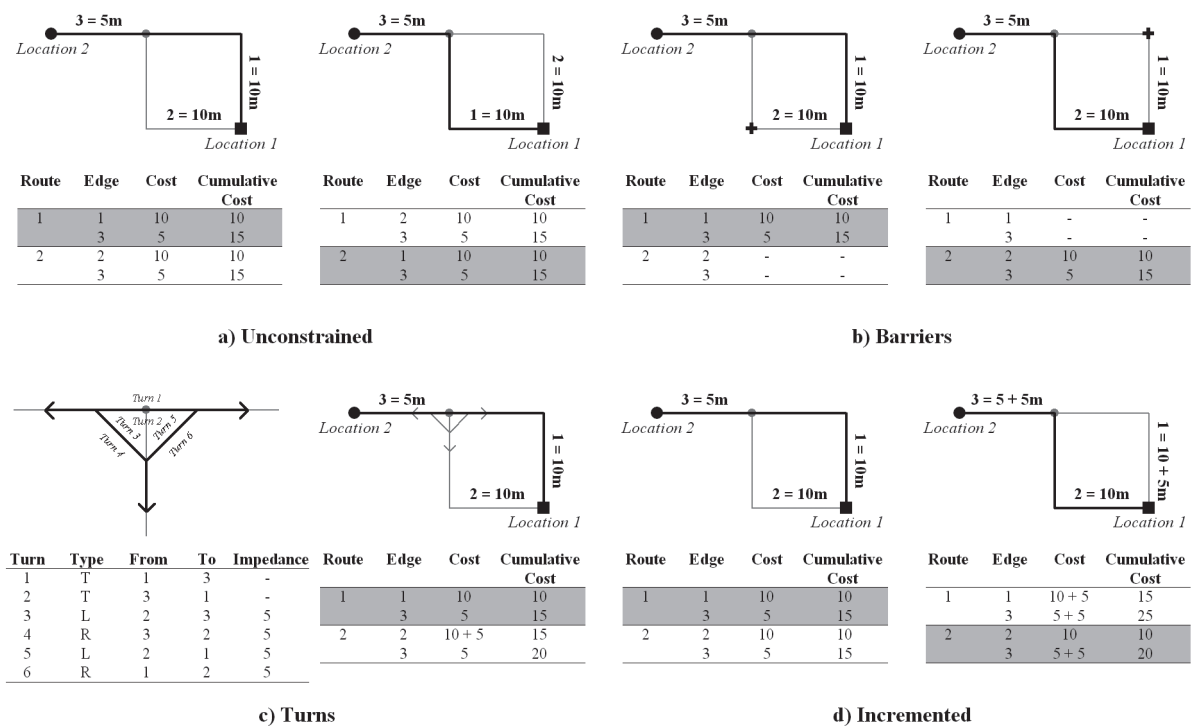


Fig. 2. Network logic: the basic principles of Closest Facility Analysis. The solution is shown shaded in the tables.

closest facility solution can be obtained by creating either a third network location to act as a barrier to prevent the use of an edge element in the solution (Fig. 2b), or by creating a turn feature class with impedances assigned to every possible turn at each junction element in the network dataset (Fig. 2c). Different impedances can be assigned to straight transitions and left or right turns to reflect the additional cost expended in moving between contiguous edge elements. Alternatively, both closest facility solutions could be obtained by incrementing the cost for each edge element used in the initial solution and running the routine again (Fig. 2d).

### *Sigwells (West)*

An extensive complex of geophysical anomalies thought to correspond to ditches and enclosures has been identified in a large plot of land known as Sigwells, approximately 1.75 km to the south-west of the hillfort (Figs. 1 and 3). This complex has been an important focus for test-pitting and excavation. Provi-

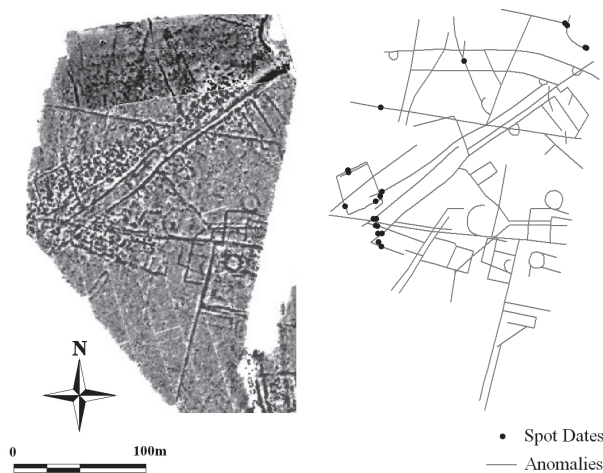


Fig. 3. The geophysical data and corresponding network dataset for Sigwells (West).

sional phasing has been established using traditional non-GIS methods (TABOR / JOHNSON 2000), and a portion of this dataset, Sigwells (West), is used here to develop and test this innovative GIS-based methodology. The provisional phasing identified six discrete phases of activity (Systems 1 to 6), from the Early Bronze Age until the late Romano-British period, with two underlying trends – one running WNW to ESE (System 1) and the other NE to SW (System 2). Continued

excavation, however, has resulted in a revised chronology (Phases 1 to 8):

1. Neolithic (NEO) – no equivalence
2. Early Bronze Age (EBA)  $\approx$  System 4
3. Middle Bronze Age (MBA)  $\approx$  System 4
4. Middle Iron Age 1 (MIA 1)  $\approx$  System 2
5. Middle Iron Age 2 (MIA 2)  $\approx$  System 2
6. Late Iron Age (LIA)  $\approx$  System 3
7. Early Romano-British (ERB)  
– no equivalence
8. Middle Romano-British (MRB)  $\approx$  System 1

### The Network Dataset

The methodology is explained elsewhere (LOCK / POUNCETT IN PRESS) and summarised here. Geophysical data was digitised as a line feature dataset, with individual polylines created between the start and end points of each geophysical anomaly or intersections with other anomalies. Stratigraphic relationships between anomalies corresponding to excavated features were encoded as z-values and the corresponding elevation fields were used to establish connectivity and create edge elements and junction elements for the resultant network dataset (Fig. 3). A separate turn feature class was created to encode the physical relationships between geophysical anomalies that intersected at junction elements. Impedances were assigned to reflect the likelihood that any two intersecting edge elements were related to one another and the network dataset was rebuilt.

Spot dates from test-pitting and excavation were digitised as a point feature dataset, with points corresponding to the centroids of sections excavated through features or deposits corresponding to geophysical anomalies. Two datasets were created for each phase of activity within the revised chronology – one for spot dates belonging to that phase and another for spot dates corresponding to later phases. A third point feature dataset, common to all of the phases of activity, was also created. This dataset contained point features corresponding to the start and end points of each of the geophysical anomalies and together, these feature datasets – incidents, facilities and barriers respectively – constituted the network locations used for the purposes of Closest Facility Analysis. The number of later spot dates (barriers) decreases with time, and movement around the network becomes less constrained.

### An Initial Solution

Separate closest facility layers were created for each phase of activity. In each instance, a preliminary solution was obtained, with costs calculated on the basis of the length of edge elements rounded to the nearest integer. Where a phase was associated with more than one spot date, the solution comprised individual routes between each of the spot dates (incidents) and every start or end point of a geophysical anomaly (facility) that was not blocked by a barrier. Facilities were ranked according to their proximity to each of the incidents. Separate solutions were subsequently obtained for each of the incidents and facilities in rank order, each time incrementing the costs of edge elements used in individual solutions and rebuilding the network dataset. Incrementation increases the probability that all possible routes are included in the final solution.

Several different methods have been identified for integrating the solutions for each of the phases of activity (LOCK / POUNCETT IN PRESS) and the method used here is based upon the dominant phase by standardised count. Geoprocessing tools are used to obtain counts of the number of times an edge element is used in each of the incremented solutions which are standardised by the maximum number of possible routes (i.e. the product of the number of incidents and the number of facilities) and converted to percentages. The phase of activity with the highest standardised count is identified and the corresponding dates are assigned to the edge elements. Dates for individual edge elements are subsequently aggregated to identify the dominant phase for each geophysical anomaly. Conflicts are detected where two or more phases of activity are associated with the same number of edge elements (below).

### Testing the Methodology

The integrated network-based solution was considered to be a good fit with the archaeological data (Fig. 4). Both of the trends highlighted in the provisional phasing of the Sigwells (West) dataset were identified. On closer examination, however, several key differences were noted between the provisional phasing and the network-based solution, most notably the undue emphasis placed on the earliest phase of activity (Phase 1 – Neolithic) in the latter. A series of tests were subsequently carried out in order to test the accuracy (internal consistency) or



Fig. 4. The provisional phasing (left) and network-based solution (right) for Sigwells (West).

robustness of the methodology. The basic premise of these tests was to repeat the methodology using a sample of points from the initial network-based solution and compare the dates suggested by the solutions for random datasets with those from the initial solution.

### Sample Datasets

Five random samples of points (Samples 1 to 5) were generated within SPSS 14.0.2. Points, corresponding to the centroids of individual edge elements, were distributed randomly along geophysical anomalies dated to each phase of activity in the initial network-based solution with the number of spot dates per phase kept constant. Separate point feature classes were generated for the corresponding network locations (incidents and barriers) for each of the sample datasets. Whilst the spot dates from test-pitting and excavation were typically clustered, the random samples of points were dis-

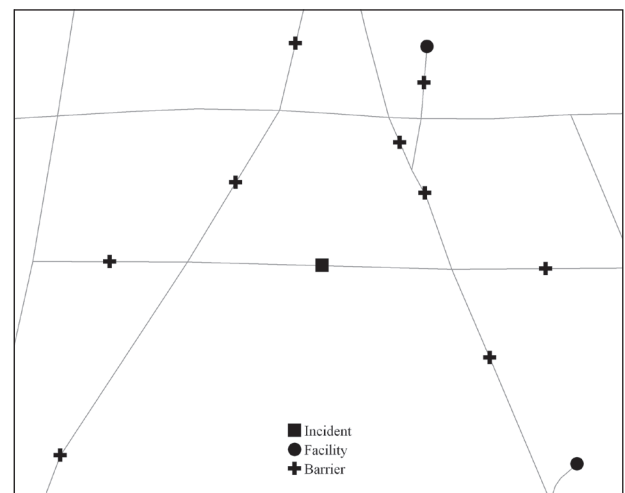


Fig. 5. Systemic error generated in Sample 2.



tributed more evenly around the network. A systemic error, with no possible solution, was generated in one of the datasets (Sample 2) where all of the adjacent edge elements for one of the incidents (Phase 1) were blocked by barriers (Fig. 5). Solutions for each of the sample datasets were obtained and geophysical anomalies corresponding with the spot dates from test-pitting and excavation were used to compare these solutions with the initial solution for the network dataset.

### Comparison of Dates

The dates suggested in the network-based solutions for each of the sample datasets correlated well with those of the initial network-based solution (Fig. 6). Three different outcomes were identified: the solutions for the sample datasets identified the same date as the initial solution (direct hits); a conflict was detected and the solutions for the sample datasets identified the date from the initial solution as one of the possibilities (indirect hits); or, the solutions for the sample dates identified a different date to the initial solution (misses).

| Anomaly | Sample 1 | Sample 2 | Sample 3 | Sample 4 | Sample 5 | Initial Solution |
|---------|----------|----------|----------|----------|----------|------------------|
| 1       | ●        | -        | ●        | ●        | ●        | ●                |
| 2       | ○        | ●        | ○        | ○        | ●        | ●                |
| 45      | -        | -        | -        | -        | -        | -                |
| 46      | -        | -        | -        | -        | -        | -                |
| 119     | ●        | ●        | ●        | ●        | ●        | ●                |
| 120     | ●        | ●        | ○        | ●        | ●        | -                |
| 122     | ○        | ○        | ○        | ○        | ○        | ●                |
| 123     | -        | ●        | ●        | ●        | -        | ●                |
| 124     | -        | -        | -        | -        | ○        | -                |
| 138     | ○        | ○        | ○        | ○        | ○        | ●                |
| 148     | -        | -        | -        | -        | -        | -                |

● Direct Hit      ○ Indirect Hit      - Miss

Fig. 6. Concordance of the dates from the network-based solutions for the sample datasets and the spot dates from test-pitting and excavation (latest phase of activity only).

The majority of the hits, whether direct or indirect, require no further explanation. Two of the indirect hits (Anomalies 122 and 138), however, corresponded to excavated features that had been re-cut and were consequently associated with more than one spot date. In both instances, one of the dates identified in the solutions for the sample datasets matched the date from the initial solution, which in turn corresponded to the most recent spot date i.e. the latest phase of activity with which the geophysical anomaly was potentially associated. One of the

misses (Anomaly 1, Sample 2) corresponded to the systemic error identified above.

Three repeated misses were identified (Anomalies 45, 46 and 148) in the solutions for the sample datasets. In each instance, the date identified by the initial solution did not correlate with the spot date from test-pitting and excavation. The standardised counts for the edge elements that constituted part of Anomalies 45 and 46 were all low, suggesting little confidence in the dates assigned on the basis of the network-based solutions. The dates suggested for these anomalies (Phases 1 and 5 respectively) reflect a bias towards the earliest phase of activity or phases of activity with a single spot date inherent within the methodology. Standardised scores for the edge elements that corresponded to the misses for Anomaly 123 (Samples 1 and 5) were also comparatively low.

In contrast, the standardised scores for the edge elements that constituted part of Anomaly 148 were much higher, suggesting a greater degree of confidence in the dates identified on the basis of the network-based solutions. The date obtained by the solutions for each of the sample datasets (Phase 8) agrees with that suggested in the provisional phasing (TABOR / JOHNSON 2000). As such, the methodology would appear to have identified an erroneous date and the spot date from test-pitting and excavation (Phase 2) was thought to correspond to residual Early Bronze Age pottery incorporated into the fill of the feature. The standardised scores for the edge elements that corresponded to the misses for Anomaly 124 (Samples 1 to 4) were also high, possibly indicating the continued use or re-use of the feature during the Early Romano-British period (Phase 7).

### Conclusions

The testing against results from random datasets has shown convincingly that the CFA-based methodology developed here produces believable results and is robust. The movement of known spot-dates from excavation around the geophysical network has been demonstrated although this did involve considerable manipulation of the ESRI logic inherent within the software to incorporate the required archaeological logic of stratigraphic relationships and intercutting features. The methodology has been shown to be internally consistent by working backwards from CFA solutions to known dating characteristics of the excavated data.

The methodology does not, however, offer a single ‘push-button’ solution but is an iterative process that requires user-input to resolve dating conflicts. It identifies chronological trends within the dataset but also, and in some ways more importantly, it highlights potential errors in dating that need to be checked. This is illustrated above by the discussion of ‘anomalies’ which are shown to be unreliable dates, or to have more than one possible date, for a variety of reasons, for example the well known archaeological problem of residuality. Through an archaeological consideration of the highlighted dating conflicts, any resulting re-classification can be re-analysed and a best fit solution arrived at.

Even so there are limitations within the methodology which need to be acknowledged and accepted. A solution will only include the latest (most recent) phase of activity for any feature, thus ignoring any earlier possible dates for that feature. There is also a bias towards the initial (earliest) phase and any phases that only have a single spot date. As a result of this, it is clear that the methodology is best suited to a large-scale dataset with a widespread spatial distribution of spot dates which includes multiple spot dates for each phase of activity.

This work was intended to develop a methodology which is applicable to the larger area of geophysics completed within the South Cadbury Environs Project, and at the same time may be of wider interest and application. This has been achieved and the next stage is the larger analysis, the results of which will be reported in the final Project monograph.

## References

ALCOCK 1972

L. ALCOCK, ‘By South Cadbury, is that Camelot...’ Excavations at Cadbury Castle 1966-1970 (London 1972).

ALCOCK 1995

L. ALCOCK, Cadbury Castle, Somerset: The Early Medieval Archaeology (Cardiff 1995).

BARRETT / FREEMAN / WOODWARD 2000

J. C. BARRETT / P. FREEMAN / A. WOODWARD, Cadbury Castle, Somerset: The Later Prehistoric and Early Historic Archaeology. English Heritage (London 2000).

LOCK / POUNCETT IN PRESS

G. LOCK / J. POUNCETT, A Vector-based Approach to the Integration of Geophysical and Test-pitting Data: Phasing the South Cadbury Environs Project. In: J. CLARK (ED.), Digital Discovery: Exploring New Frontiers in Human Heritage. Proceedings of the Computer Applications and Quantitative Methods Conference CAA 2006 (Fargo 2006).

TABOR / JOHNSON 2000

R. TABOR / P. JOHNSON, Sigwells, Somerset, England: Regional Application and Interpretation of Geophysical Data. *Antiquity* 74, 2000, 319–325.

*John Pouncett*

*Gary Lock*

*University of Oxford,*

*Institute of Archaeology*

*36 Beaumont Street*

*Oxford OX1 2PG, United Kingdom*

*johnpouncett@ase-limited.co.uk*