Measuring Chronological Uncertainty in Intensive Survey Finds. A Case Study from Antikythera, Greece.

Andrew Bevan, James Conolly, Christian Hennig, Alan Johnston, Alessandro Quercia, Lindsay Spencer, Joanita Vroom

Preprint of a paper published in 2012 in *Archaeometry* 54.

Abstract

This paper considers how to make the most out of the rather imprecise chronological knowledge that we often have about the past. We focus here on the relative dating of artefacts during archaeological fieldwork, with particular emphasis on new ways to express and analyse chronological uncertainty. A probabilistic method for assigning artefacts to particular chronological periods is advocated and implemented for a large pottery dataset from an intensive survey of the Greek island of Antikythera. We also highlight several statistical methods for exploring how uncertainty is shared amongst different periods in this dataset and how these observed associations can prompt more sensitive interpretations of landscape-scale patterns. The concluding discussion re-emphasises why these issues are relevant to wider methodological debates in archaeological field practice.

1. Introduction

Archaeologists have long focused on understanding patterns of spatial and temporal variation. Indeed, such patterns underpin almost any interpretation we might ever wish to make about the past. Intensive surface survey ('fieldwalking', landscape survey) is an important means by which robust evidence about whole landscapes of past human activity can be collected with at least some degree of accompanying temporal control (Banning 2002; Alcock and Cherry 2004). However, one major challenge for archaeologists has always been that we date the things we uncover via a bricolage of methods that are unusual and/or rather imprecise stand-ins for the direct measurement of time (Crema in press). Hence we often make only very uncertain assessments about whether a phenomenon has really occurred within a specific time frame, for how long it lasts, where it sits in a sequence of other events, or by how much it is separated in time from other phenomena. Intensive survey is a branch of archaeological data collection that is particularly prone to such chronological uncertainties because it typically works with: (a) relative dates based on the style of surface artefacts, (b) poorly preserved surface finds, and c) unstratified assemblages. Even so, there is by now an established theoretical literature on time and landscape (Rossignol and Wandsnider 1992; Bailey 2008), as well as a growing sense of the opportunities for more integrated space-time analysis in archaeology (Lock and Harris 2002; Bocquet-Appel *et al.* 2009; Collard *et al.* 2010; Crema *et al.* 2010), in step with a wider agenda across many areas of academic research (Peuquet 2002; Galton 2004; Diggle 2006; Cressie and Wikle 2011). Given this platform, it is surprising that temporal uncertainty has not figured more obviously in the formal discussion of landscape survey data, and that those few good treatments that do exist have been exploratory re-workings of existing datasets (e.g. Fentress *et al.* 2004; Johnson 2004). This paper seeks to address chronological uncertainty in survey datasets from the outset, by introducing a novel approach to the relative dating of surface finds. It begins by exploring both the strengths and weakness of current approaches to dating survey artefacts, at the same time as introducing a case study of surface pottery recovered by an intensive survey of the Greek island of Antikythera. It then introduces several statistical methods that can be used to understand the ways in which chronological uncertainties associated with different datable periods in a regional sequence are not independent of one another but shared in important ways. The penultimate section (4.3) then suggests ways in which this might be modelled spatially as a guide to further field investigation.

2. Problem Definition and Case Study

A traditional way to store information about the date of archaeological finds is to record, for each archaeological site, feature or artefact, a crisply-defined category in a database, such as ``Hellenistic" (if we adopt a Mediterranean periodisation as an example). Sometimes these categories exist within a hierarchy of more and less precise temporal definitions (e.g. Moore 2008) and sometimes they do not. Occasionally, more subtle recording methods might also include categories that express a degree of indecision and or a weight of opinion, such as a potsherd that is “possibly Hellenistic?” or “probably Classical, possibly Hellenistic” or “Hellenistic or Late Roman”. Note that in the latter hypothetical case, the two periods involved are not even adjacent in time (i.e. Hellenistic or Late Roman but not the Early and Middle Roman phases in-between). With abraded sherds from unstratified deposits, the degree of nominal-scale nuancing that might therefore be necessary to convey a full range of specialist knowledge (if we are thinking positively) or uncertainty (if we are being pessimistic) rapidly leads to very large sets of different database labels. Ultimately, the visual presentation of such variability becomes tricky, as does any subsequent quantitative treatment.

The alternative is to express archaeological dates in a probabilistic manner. For example, the presentation of absolute radiocarbon dates with standard errors, or better yet, as probability distributions is already common practice (e.g. Buck *et al.* 1996), and the resulting uncertainty can sometimes then be retained for spatial analysis (Bocquet-Appel *et al.* 2009; Collard *et al.* 2010; Green 2011; Grove 2011). However, the fuzzy expression of dates derived from traditional forms of relative chronology (e.g. via artefact typologies: Lyman and O’Brien 2006) is far less straightforward. One option for existing datasets with relative dates is to chop them up into small time blocks, of say 50-100 years each, and remodel the existing period categories as a set of probabilities that, for example, an artefact or archaeological feature falls into each of the new time-blocks (with probabilities in all time blocks summing to 1, and ignoring for the moment the problem of long-enduring features such as whole archaeological sites). This has sometimes been referred to as ‘aoristic’ analysis (Ratcliffe 2000; Johnson 2004) and is undeniably a useful way of exploring the different temporal scenarios that uncertain dating leaves open (e.g. Crema *et al.* 2010). However, so far it is only pre-existing archaeological data that has been reworked in such examples, with the obvious drawback that someone has already assigned the feature or artefact in question to a specific categorical phase and with that, some of the fuzziness surrounding their initial decision-making has already been thrown away. For example, Scott Ortman and colleagues (2007) model occupation period probabilities for Pueblo societies in the American Southwest using Bayesian methods, but have to base their approach on ceramic data that has already been assigned to chronological phases without any formal measure of uncertainty.

It is therefore attractive to consider assigning a probabilistic date from the outset (or several contending possibilities by different specialists, see below). In theory, one way this could be done is by suggesting a central date and a confidence interval around it, in close mimicry of the presentational style of absolute dates. However: (a) there is no easy way to construct such an interval, and (b) it implies a symmetrical distribution around the central date which takes no account of situations where one end of the chronological range is far more likely than the other, or where there is effectively a multi-modal distribution of non-successive possibilities (as in the case above of an artefact thought to be "either Hellenistic or Late Roman, but not the Early-Middle Roman phase in-between"). A second way to handle such uncertainty is to work with fairly small time blocks expressed in absolute date ranges in the manner of aoristic analysis above (e.g. AD 50-100, AD 100-150 , etc.). While this has one major advantage in that it can define timespans of equal length (that are hence more easily compared), it ultimately just trades off a well-established, culturally and historically meaningful set of timespans (site phases, regional periodisations) for a new arbitrary one. More importantly perhaps, it constitutes a fairly substantial assault on the way that artefact specialists traditionally operate and mixes in, at a dangerously early stage of the process, a second uncertainty about the exact relationship between relative and absolute dates.

A third option, that we have prioritised in our own work, is to retain those relative dating categories already familiar to specialists on a research project (e.g. a sequence of well-known regional periods built largely from artefact typologies) and assign a value to each that expresses our confidence that the feature in question belongs to that period. In this manner, we seek to express, retain and exploit patterns of temporal uncertainty from the point at which such artefacts are first encountered in the field, through the process of laboratory study and on to final publication. With this agenda in mind, the Antikythera Survey Project (www.ucl.ac.uk/asp) adopted such an approach to artefact dating since its inception in 2005. Antikythera is a small (20.8 sq.km) Mediterranean island located in the midst of some important shipping lanes, between the Peloponnese and Crete and between the Adriatic and Aegean seas. It has had a history of human exploitation stretching back some 7,000 years and including discrete episodes of extensive settlement punctuated by periods of near or total abandonment. The surface survey of this island proceeded in two stages: in the first stage, the entire island was walked by five-person teams spaced 15m apart and collecting finds every 10m along their line. In the second stage, particularly interesting or unclear parts of the landscape were reinvestigated on a 10x10m grid. In both cases, artefacts were collected permanently and systematically. In this paper, we focus entirely on the pottery finds, and restrict our analysis to the assemblage of some 10,065 'feature' sherds that are fragments of rims, bases, handles or those with paint, glaze or other decoration. Taking inspiration from, and seeking to be consistent with, archaeological surveys in other parts of the south-west Aegean (especially on neighbouring Kythera), we adopted the periodisation shown in Table 1 for thinking about episodes of human activity and material culture on Antikythera.

Chronological Phase	**Absolute Dates (approx.)**
Middle to Late Neolithic (MN–LN)	6000–4500 BC
Final Neolithic to Early Bronze 1 (FN–EB1)	4500–2700 BC
Early Bronze 2 (EB2)	2700–2200 BC
Late Prepalatial (LPrepal)	2200–1900 BC
First Palace (FPal)	1900–1700 BC
Second Palace (SPal)	1700–1450 BC
Third Palace (TPal)	1450–1200 BC
Post-Palatial to Protogeometric (PPalPg)	1200–900 BC
Geometric (Geo)	900-600 BC
Archaic (Arch)	600-500 BC
Classical (Class)	500-325 BC
Hellenistic (Hell)	325 BC-AD 0
Early Roman (ERom)	AD 0-200
Middle Roman (MRom)	AD 200-350
Late Roman (LRom)	AD 350-650
Early Byzantine (EByz)	AD 650-900
Middle Byzantine (MByz)	AD 900-1200
Early Venetian (EVen)	AD 1200-1400
Middle Venetian (MVen)	AD 1400-1600
Late Venetian (LVen)	AD 1600-1800
Recent (Recent)	AD 1800-present

Table 1. Chronological periods used for dating purposes by the Antikythera Survey Project.

Data Treatment and Initial Discussion

Although finds were often given an initial period date in the field, or later on the same day as they were going through preliminary processing, the hard work of dating them properly was conducted some time later by a set of ceramic specialists, each with their own particular expertise in certain chronological periods of Aegean and eastern Mediterranean archaeology. More precisely, we first set up a series of chronological periods that were in common use amongst Aegean surveys and excavations. Then each potsherd was recorded individually, and each specialist suggested a probability that the sherd in question might belong to one or more of the periods with which they were most familiar (recording a separate comment in those rare cases where a sherd could be dated more precisely than available period categories). Purely for the comfort of those using this method, we also decided to refer to it as a 'percentage confidence' out of 100, rather than a probability out of 1. In the purely hypothetical case mentioned above, of a potsherd dated awkwardly as "either Hellenistic or Late Roman, but not the Early or Middle Roman phases in-between", the resulting record might conceivably lead to assigned values of 70% Hellenistic, 0% Early Roman, 0% Middle Roman, and 30% Late Roman. We later use the notation P_{ij}

for these percentages, where the first index refers to the sherd (i) and the second one refers to the period (j).

We will come back to the interesting issue of how best to elicit such probabilistic estimates from artefact specialists in the discussion at the end (but for an excellent overview of this issue beyond archaeology, see O'Hagan *et al.* 2006), as there are clearly opportunities to explore how expert judgements might vary in different circumstances, or develop with increasing study. In practice, we adopted the relatively simple approach of letting all artefacts specialists assert proportional ownership of artefacts that they thought might belong to their periods of expertise, in the order in which they came to them in the laboratory. In cases, where there was initial disagreement, a consensus view was eventually reached, and a residual 'Other' category allowed for some proportion of the overall probability to be left unattributed. Hence, the final percentages reflect the view of six pottery specialists each with an overlapping but different period focus, but all of them studying the entire permanent collection. We do not propose to consider here whether these are best thought of as prior probabilities in a Bayesian sense (Buck *et al.* 1996) or would be better framed within some other belief-based scheme (e.g. Shafer 1990), but informally, we found it helpful to conceive of this as a friendly wager among colleagues about the artefact's date.

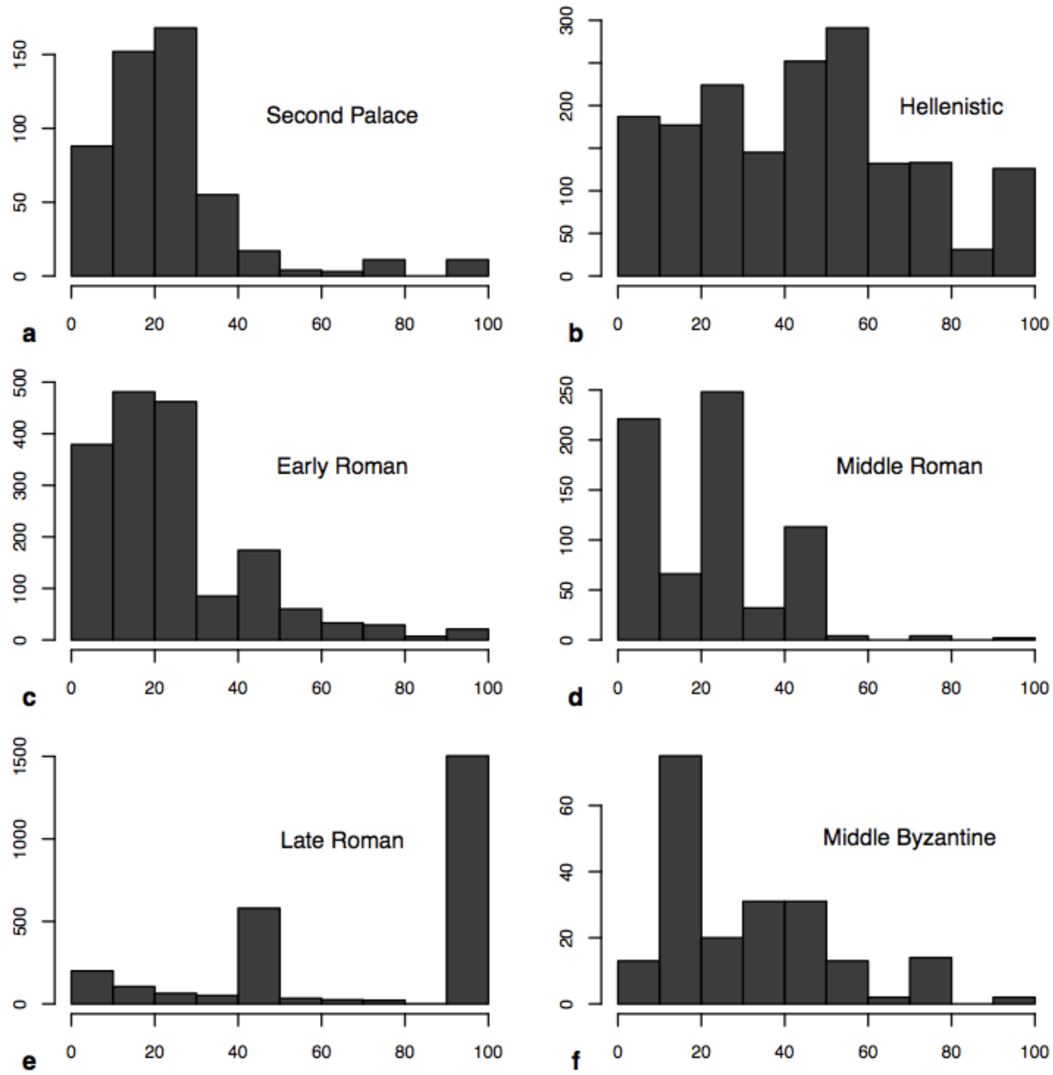

Figure 1. Histograms of the percentage confidences assigned across the whole Antikythera assemblage for six different periods: (a) Second Palace (c.1700-1450 BC), (b) Hellenistic (c.325-100 BC), (c) Early Roman (c.100 BC-100 AD), (d) Middle Roman (c.100-350 AD), (e) Late Roman (c.350-650 AD) and (f) Middle Byzantine (c.1000-1200 AD). Bin ranges of percentage confidences on the x-axis are all 10% wide and only sherds with at least some suggestion of belonging to a given period are included in each plot (i.e. those with >0% for that period).

Once a whole assemblage has been treated in this way, what useful things can be done with the resulting data that could not be so easily done otherwise? A first simple approach is to plot the frequencies of different assigned confidences by period. For example, in all the analyses below, we select a range of interesting periods to consider out of the full set shown in table 1. Figure 1, for example, selects six of these periods, and among these, the Late Roman, is clearly one for which the ceramic evidence is particularly recognisable. For our purposes below, we will refer to this situation as one in which a chronological period exhibits high 'diagnosticity'. In practical terms, the crucial factor here is the production in the Late Roman period of particularly distinctive slipped tablewares and groove-decorated amphorae that inflate the degree to which this phase is visible in the surface archaeological record, relative to other periods (and see also Pettegrew 2007; Quercia *et al.* 2011). The degree of settlement or other human activity on the island in this phase may still have been unusually high (indeed, we suspect that it was), but our attempts to assess whether this is so are confounded by the varying clarity of our evidence.

In contrast to the Late Roman, several other periods suffer from generally low levels of diagnosticity. So, for example, the skew of the Second Palace histogram in figure 1a is very much towards low percentage confidences suggesting that there are only a few potsherds that are ever unequivocally assigned to this date, but quite a few for which there remains a lower level suspicion that they might be from this period. For the Middle Roman (figure 1d), there is also a clear 'sawtooth' pattern to the histogram which suggests only a very limited number of types of ceramics are generating consistent ascribed percentages (e.g. sherds dated 50% Middle Roman 50% Late Roman, or 30% Early Roman 40% Middle Roman 40% Late Roman). For the Hellenistic instead we have a slightly more uniform distribution suggesting a whole range of different degrees to which this period can be pinned down. In the case of all of these histograms, the raw count (on the y-axis) of sherds for which there is even the slightest suspicion that they might be dated to this period is a useful indicator of the scale of overall evidence.

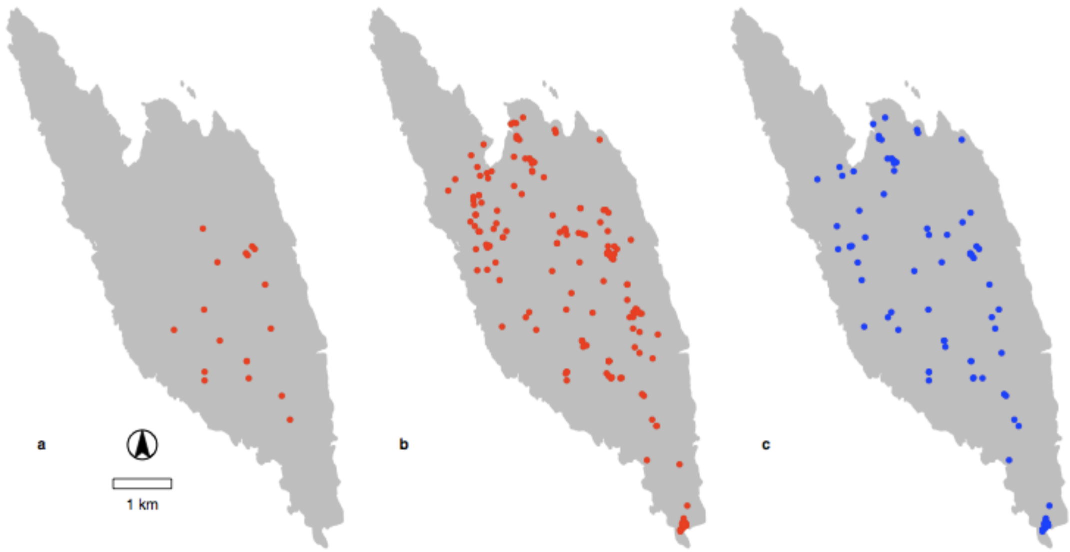

Figure 2. The spatial impact of temporal uncertainty: distribution maps of all sherds that are (a) ≥70% percentage confidence of being Middle Byzantine (c.1000-1200 AD), (b) ≥20% Middle Byzantine, and (c) ≥70% Middle Byzantine and Early Venetian (c.1000-1400 AD).

Such plots therefore constitute an important first stage by which we can interrogate patterns of chronological uncertainty in the artefact record. While our interest here is in their relevance to intensive survey and landscape archaeology, we see no reason why they should not also be relevant for understanding excavated assemblages. In any case, how best do we visualise such uncertainty on a map? The simplest way is to explore the distribution for a particular period by defining different percentage cut-offs for what constitutes a likely find of that date. Hence, figure 2a-b shows the distribution of Middle Byzantine potsherds where we have respectively taken ≥70% and ≥20% as the arbitrary thresholds. It was already clear from figure 2f that the Middle Byzantine is another period with comparatively low diagnosticity, where certain glazed finewares are highly recognisable, as are an amphora type or two, but much of the local coarseware assemblage remains harder to define in the present state of our knowledge (see Vroom 2005: 66-105). When we look at the spatial distributions of such finds at these two different thresholds, it becomes clear that our interpretations of likely land use and settlement will differ substantially depending on how much uncertainty we are willing to accept. In the most pessimistic case (figure 2a), we have a distribution of finds that highlights some very specific upland landscapes of hard limestone in the interior of the island. If in contrast, we take a much more optimistic view (figure 2b), then the northern harbour areas, an inlet in the far south of the island and a series of other landscapes become candidates for more substantial activity in this period. Some of this wider activity clearly does date to the Medieval period, rather than just being diagnostic 'noise', because if we consider the summed percentages of two consecutive periods such as Middle Byzantine and Early Venetian (c.1000-1400 AD), the wider pattern persists (figure 2c).

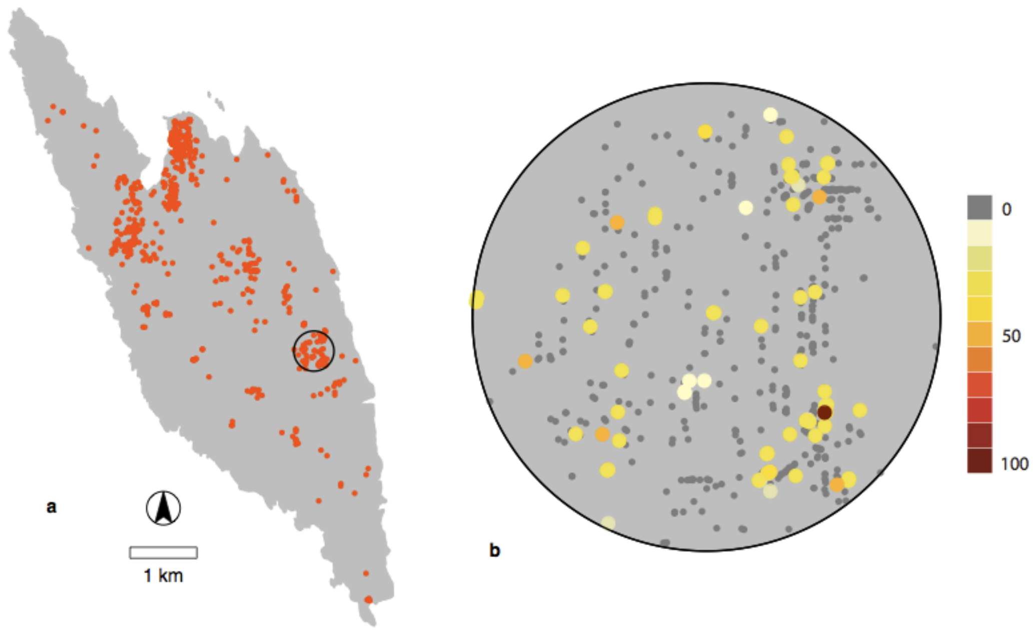

Figure 3. Visualising diagnostic confidence: (a) all possible candidate sherds for a Middle Roman date (i.e. >0% confidence), and (b) close-up of the circled area (diameter 600m) with percentage confidences of Middle Roman date shown as graduated colours overlain on all other sherds (in grey)

One possible way to read this behaviour is that the upland was colonised slightly earlier and that activity then consolidated in other parts of the landscape in the century or two that followed. Even so, what the above analysis should make clear is that, in the presence of substantial uncertainty about dating, there is no single solution – the requisite evidence is simply not there or not yet recognised. By exploring different options all we can do is come up with a plausible set of scenarios that prompt further, more refined investigation. Simulating the sherd distribution many times based on the percentage confidences and considering variation in results is one possible tactic (i.e. a Monte Carlo approach: Crema *et al.* 2010). In this way, a potsherd with a 20% confidence of being Middle Roman will, on average, be included in one of every five simulated distribution maps for this period. Viewing maps of different simulation runs side-by-side thereby communicates both where the patterns are robust and where they are not, whilst any form of spatial analysis can be re-run on each separate simulation to assess the degree to which the results do or do not vary. A further related way to visualise the uncertainty associated with a specific period is, instead of plotting by a particular cut-off confidence or many simulations, to assign graduated colour levels to the percentage confidences (i.e. a colour 'ramp'). Figure 3b does this for a small sub-region of the overall distribution and makes clear that sherds confidently attributed to the Middle Roman period are very rare on Antikythera (as we have already seen in figure 1d), and that, within this example sub-region at least, there is no obvious sub-patterning to the more confidently dated finds.

4. Shared Temporal Uncertainty

Of the examples so far discussed, the one that considered the combination of Middle Byzantine and Early Venetian periods (figure 2c) implicitly suggested that these two consecutive periods share a degree of chronological uncertainty between them. Put simply, we could anticipate that many sherds with >0 and <100% chance of being Middle Byzantine would also have a chance of being Early Venetian. This stands to reason, for while the chronological periods with which we choose to work will certainly match some broad cultural trends if we have chosen them wisely, certain artefact styles will undoubtedly extend over a time range that overlaps with two or more of these periods. More generally, this anticipated sharing of uncertainty by adjacent periods fits into a broader theoretical proposition: that things dating close together in time are more likely to be similar (e.g. in style) than those found further apart (a specific instance of a general property known as temporal autocorrelation). However, due to the nature of relative dating methods, which are based on the definition of groups with consistent decorative, morphological and compositional characteristics (i.e. the creation of a typology), it is also quite possible to have non-consecutive periods whose uncertainty is shared: hence in the Bronze Age, coarseware vessels in red micaceous fabrics are a reasonably common import to Antikythera from the neighbouring island of Kythera in both the mid 3rd millennium BC (Early Bronze 2) and the latter half of the 2nd millennium BC (Second to Third Palace periods) without any sign of such a fabric in-between (Pentedeka *et al.* 2010: fabric class 'MIC').

Such sharing of uncertainty is a major problem, as it means that the presence of material from one period risks suggesting the presence of another period, even if the latter is not really there. One obvious thing to try to do therefore is to measure this shared uncertainty and thereby at least understand the size of the problem in each case. Unfortunately, the most common measures of correlation (e.g. a Pearson's or Spearman's coefficient) are not useful here because they produce misleading results. For example, take two periods, j and k, whose percentage confidences are typically correlated with one another such that when there is a percentage confidence P_{ij} of period j for one sherd i, there is also a percentage confidence P_{ik} of period k. In some instances this will be measured as a positive correlation, but as the sum total of the two periods nears 100% a negative correlation will set in because $P_{ij} \approx 100 - P_{ik}$. The situation becomes even more awkward when we start dealing with more than two periods. An alternative which gets around this issue is to use one of the two methods discussed below: a first simply considers the degree to which the assigned probabilities in one period can be explained by any other period, and a second which treats the shared probabilities exhibited by each pair of periods separately.

4.1 Overall Uncertainty

To be as clear as possible below, we will make use of the toy dataset shown in table 2, alongside our consideration of the larger and more complex dataset represented by the real survey finds. In this table, five artefacts have been assigned to five different chronological periods with hypothetical percentage confidences (P_1 ... P_5). These values imply, for example, that artefact 5 is highly diagnostic and can be attributed to a single period (P_3) with complete confidence, whilst artefact 4 is not very diagnostic and can only be assigned an equal chance of belonging to any of the five periods. The other sherds fall, diagnostically-speaking, somewhere in-between.

Artefact	**P_1**	**P_2**	**P_3**	**P_4**	**P_5**
1	0	40	50	0	10
2	80	10	10	0	0
3	0	0	0	50	50
4	20	20	20	20	20
5	0	0	100	0	0

Table 2. A hypothetical dataset of five artefacts variously attributed to five different time periods.

The first proposed method is relatively straightforward and calculates the overall proportion of uncertainty in any one period that can be explained by the percentages in *any* other period, as follows:

$$U_j = \sum_{i=1}^{n} min\left(P_{ij}, max_{k \neq j}\left(P_{ik}\right)\right) \Big/ \sum_{i=1}^{n} P_{ij}. \qquad (1)$$

Rephrasing this in plain language, if we wish to calculate the overall uncertainty associated with period j (U_j), we would go through each artefact (i) in the sample (from 1 to n), take the percentage confidence (P_{ij}) assigned to that artefact for period j and compare it to the maximum percentage confidence associated with any other period for that artefact (P_{ik}). Then take the minimum of these two values. Do the same for all artefacts in the sample and sum all of these minimum values. U_j is then this sum divided by the sum of all percentage confidences for period j. The resulting index ranges from 0 to 1 and a low value suggests that the period concerned is diagnostically independent, for the most part, from other periods, whilst a high value suggests that the ascribed percentages in that period are strongly associated with one or more other periods. For example in the hypothetical example shown in table 1, the overall uncertainty U_j for period 2 would be:

(40+10+0+20+0) / (40+10+0+20+0) = 1

and for period 3:

(40+10+0+20+0) / (50+10+0+20+100) = 0.389.

It is primarily artefact 5 that leads to these different results in this case. If we apply this method to the real survey dataset of potsherds from Antikythera, we get the summary statistics shown in table 3. What is abundantly clear from this is that certain periods have next to no potsherds whose percentage attribution is not somehow explained by the presence of another more dominant period.

Period	**U_j**
Second Palace	0.823
Hellenistic	0.357
Early Roman	0.673
Middle Roman	0.961
Late Roman	0.091
Middle Byzantine	0.753

Table 3. Overall uncertainty values for six periods from the Antikythera survey dataset.

4.2 Pairwise Uncertainty

The second method is a bivariate technique. It is possible to express the temporal uncertainty shared by two distinct chronological periods or timespans, *j* and *k*, over *n* artefacts in a dataset, via the following ratio:

$$U_{jk} = \sum_{i=1}^{n} min\left(P_{ij}, P_{ik}\right) \times 2 \Big/ \left(P_{ij} + P_{ik}\right). \qquad (2)$$

For any two periods, this calculates the overlap in the two attributed percentages for each artefact, sums these overlaps, and then divides by the sum of the maximum possible overlap (which is the average of the two percentages). The result is a matrix that describes the average shared temporal uncertainty between each pair of periods, as a number between 0 (no shared percentages) and 1 (the period percentages are in all cases identical). If we again consider the hypothetical artefacts shown in table 2, and focus on only the first artefact for a moment, then the shared percentage is 40 between P_2 and P_3, 10 between P_5 and both P_2 and P_3, and 0 for all other pairs of periods. If we now consider the pairwise shared uncertainty of periods 2 and 3 across all five artefacts we would get the following value for U_{jk} using equation 2:

(40+10+0+20+0) * 2 / (90+20+0+40+100) = 140/250 = 0.56.

The full matrix of pairwise measures of shared temporal uncertainty would then look like those in table 4. Adopting this approach allows us to explore the global relationships between pairs of periods in useful ways. For example, figure 4 plots the pairwise values for six actual periods used on Antikythera that were first considered in figure 1.

	P_1	P_2	P_3	P_4	P_5
P_1	-				
P_2	0.353	-			
P_3	0.214	0.560	-		
P_4	0.235	0.286	0.160	-	
P_5	0.222	0.400	0.231	0.933	-

Table 4. Pairwise uncertainty values for the hypothetical dataset shown in table 2.

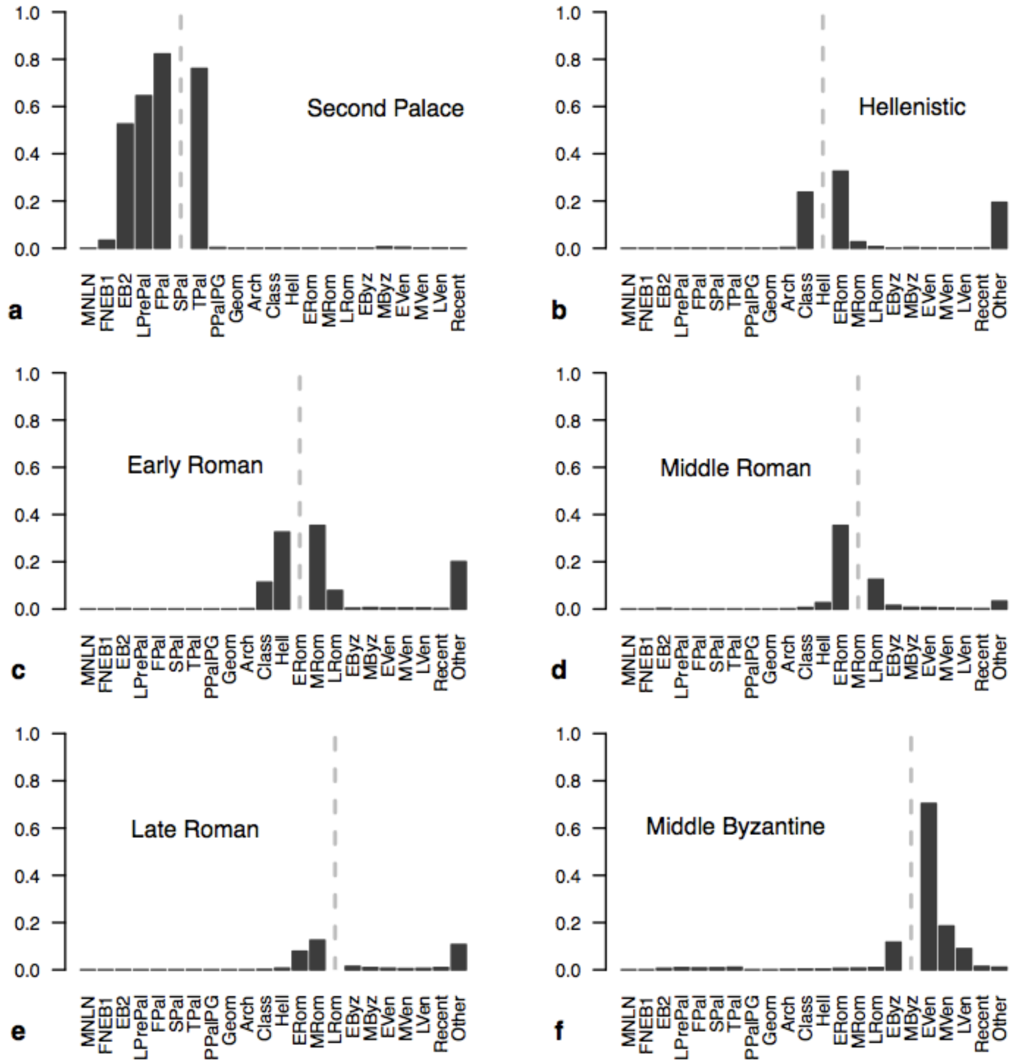

Figure 4. Comparison of the pairwise shared temporal uncertainty for six different periods: (a) Second Palace (c.1700-1450 BC), (b) Hellenistic (c.325-100 BC), (c) Early Roman (c.100BC-100AD), (d) Middle Roman (c.100-350 AD), (e) Late Roman (c.350-650 AD) and (f) Middle Byzantine (c.1000-1200 AD).

As we anticipated above, the temporal uncertainty for any one period is most strongly associated with those immediately before or after it in time, but the degree to which this is so, the number of adjacent periods involved, and the degree to which the pattern is symmetrical varies. Hence, the uncertainty with which Second Palace period finds are identified is strongly associated with uncertainties in the dating of just one subsequent period (TPal or Third Palace) and several preceding periods (figure 2a). In contrast the uncertainty associated with the Middle Roman involves preceding and succeeding periods to similar degrees (figure 2d). We can also see limited ways in which the uncertainty of the Middle Byzantine is slightly associated with later prehistoric periods (figure 2f; EB2, LPrePal, FPal, SPal and TPal) in which there existed some coarsewares that are similar to Medieval ones and hence certain poorly preserved examples might be attributed to both. The overall size of the pairwise uncertainty statistic is also relevant: as a working rule of thumb, we suspect that U_{jk} values of >0.5 are often an indication that, at least for certain analyses, it would be more sensible to lump the two periods rather than attempt to infer anything about them individually.

4.3 Local Uncertainty

The above pairwise method highlights the fact that there are very specific linkages between the chronological uncertainty present in one period and that in another. However, the exact relationship between any two periods will exhibit both a global trend that might for example be summarised by a single summary statistic as above, and a locally varying pattern. The latter localised patterning is also of great interest because it might be mapped to suggest sub-regions of Antikythera where the standard uncertainties between periods break down and where other interesting local patterns may be occurring, such as sites for periods that would otherwise be difficult to identify. We can for example compute the density (or more accurately for our purposes here, the 'intensity') of finds of one period across the island. So for example, first, taking all sherds with >0% chance of being Middle Roman, this would produce an intensity of 33.2 observed sherds per sq.km. Second and more subtly, we could do the same thing again, but this time weight each point according to its percentage confidence of belonging to this period (i.e. a sherd attributed with 100% confidence is worth five times that with 20%). Third and finally, we can express differences in overall intensity among periods by calculating the ratio of one period's intensity divided by that from one or more others. This ratio is a useful global summary of the relative chances of finding sherds of one date versus of those of others. If, for example, we consider the ratio of Middle Roman to Late Roman, we get a value of 0.09 indicating that Middle Roman sherds are far less common than Late Roman ones (or at least less commonly identified).

The advantage of this approach is that a very similar ratio (of the percentage-weighted intensity of one period versus that of one or more others) can also be calculated for each of a series of local neighbourhoods across the island (more precisely for Gaussian kernels bandwidths of σ = 50m, the latter being a good compromise between the precision of artefact positioning and the size of many typical sherd scatters). This approach fits well within an emerging geographical consensus about the need to move backwards and forwards between general characterisations of a spatial pattern and closer inspection of where there might be heterogeneous local variations (e.g. Lloyd 2007; Bevan and Conolly 2009). It is also similar to the kinds of 'relative risk' surface increasingly common in other disciplines such as epidemiology where there is an interest in localised comparisons between the observed cases of something (e.g. of flu virus) and a changing control population (e.g. those observed medically at a certain hospital, which is not the same as those at risk of flu), except that in this case it is summed percentage confidences rather than pure presence/absence values that are being assessed (see Kelsall and Diggle 1995; Hazelton and Davies 2009, and for an archaeological application: Bevan in press).

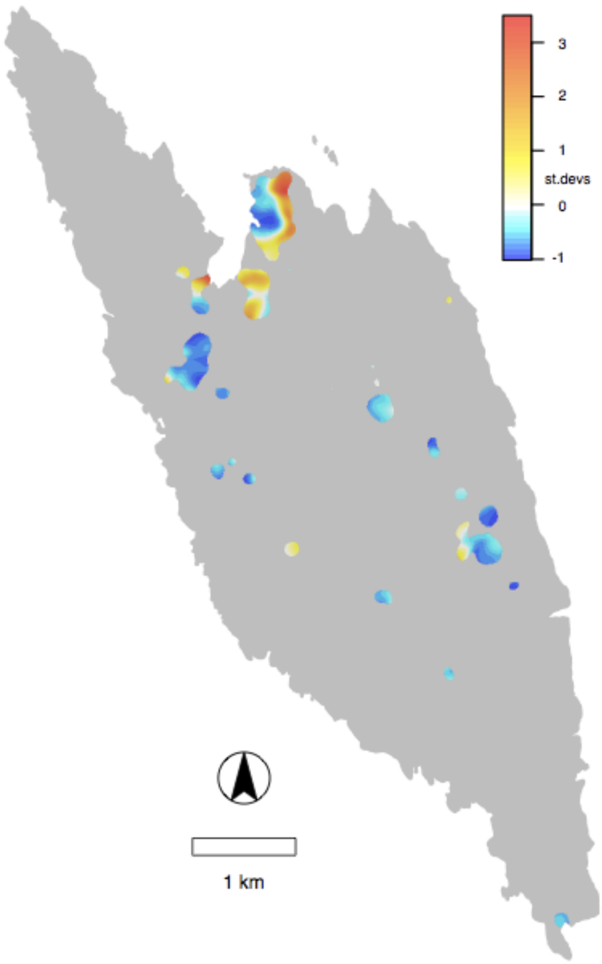

Figure 5. The relative chance of Middle Roman dates compared to Early and Late Roman ones. The colour ramp expresses standard deviations away from the mean, global odds of finding pottery with percentages ascribed to Middle Roman. Hence red colours suggest areas with a higher than average association with Middle Roman. The kernel used here is a 2D Gaussian function with σ = 50m.

Figure 5 results from applying the technique suggested above to compare the ratio of Middle Roman weighted intensities to the two periods on either side of it (Early Roman and Late Roman, these latter two being the periods that are shown in figure 2d to be the dominant sources of shared uncertainty with Middle Roman) across a series of local kernels. It expresses these local variations in the form of standard deviations away from the global (mean) ratio and the surface has also been truncated so that only areas with above a certain number of candidate Middle Roman sherds are shown. The resulting map suggests that only a fairly limited number of sub-regions of the island exhibit denser patterns of Middle Roman percentages than we might expect on average (given the number of surrounding Early and Late Roman percentages). Most of these denser patterns are around the northern harbours, around the fringes of the Hellenistic, Early Roman and Late Roman settlements but not on them. This dovetails well with our more informal impression that the island saw comparatively little permanent settlement in the Middle Roman, but was nonetheless a fairly regular stopping place at this time for passing ships and seasonal visitors who used the northern harbour zone by preference. We would stress, however, that this surface should not be seen as a straightforward map of Middle Roman activity, but rather as a guide to where further field investigation might focus.

5. Discussion

This paper has sought to place the relative dating of archaeological artefacts on a firmer footing, by developing methods that quantify the uncertainty with which artefact specialists assign their material to particular chronological periods. We think that these methods offer great potential for exploring long-established practices of artefact categorisation, indicating areas where the existing state of our knowledge is sufficient to address the questions in which we are interested, and other areas where further work is necessary. Ultimately however, it is worth reiterating that no amount of statistical manipulation can wholly make up for major gaps in our chronological knowledge, With this in mind, we would like to conclude by raising three topics where we think there are continuing research opportunities in this regard: (a) observer variability and the role of permanent artefact collections in fieldwork, (b) physical reinvestigation strategies, and (c) the importance of regional profiling.

5.1 Observer Variability

We can start considering the first of these topics by asking how variable the chronological judgements of ceramic specialists are when either conducted several times (intra-observer error) or when several specialists in the same period are involved (inter-observer error)? One way to exploit further some of the advantages of a probabilistic approach to dating artefacts is to encourage multiple, blind (and therefore potentially conflicting) dates of each artefact, by different specialists, such that areas of agreement and disagreement can be highlighted. It is clear for example that when a specialist assigns a sherd to a particular period with a 70% confidence that this is an over-precise estimate and, the next time they might suggest 60% or 80% (for the Antikythera assemblage considered above, most specialists preferred rounding off their estimates to multiples of 5 or 10%). While the process would no doubt be a personally traumatic one in some cases, we suspect that multiple, blind dating of the same finds by the same specialist would allow a more concrete assessment of this variability. Furthermore, although we emphasised consensus via discussion as a way of resolving conflicts in dating among specialists who studied the Antikythera material (and also allowed a limited number of artefacts to remain in an ‘Other’ category' if they were not picked up by anyone), multiple blind comparisons between two specialists in the same period would be a very useful way in future to highlight problem areas.

More importantly, this kind of statistical introspection is not simply methodological tinkering, but is directly relevant to wider debates associated with the collection and storage, or not, of permanent artefact collections from excavation and survey. For example, our own position is that permanent, systematic collection during intensive surveys is crucial as it makes an analysis scientifically repeatable and anticipates the gradual accumulation of extra chronological clarity as laboratory study progresses (or upon much later re-study). This contrast with a growing assertion, in Mediterranean archaeology at least (e.g. Gregory 2004; Tartaron *et al.* 2006), that the most effective trade-off of time, effort and storage capacity is to encourage surveys to make only very minimal collections and/or perform their entire dating of artefacts in the field. Our own view is that the latter approach may occasionally be acceptable if there are no other options (e.g. if imposed by operating permits), but should in no way be considered ideal. However, given that dating in the field is an action that cannot be easily revisited, the burden of proof shifts to researchers with permanent collections to demonstrate that these do indeed offer substantial added value. Using a probabilistic method of dating finds, over the course of multiple artefact study seasons, would in theory make it possible to quantify any incremental diagnostic gains.

5.2 Physical Reinvestigation

The second area where we feel further work is necessary involves the strategies for restudy, resurvey or further excavation that we might adopt to improve existing levels of diagnostic certainty about an assemblage. For some, the answer to what we should prioritise is simple: new excavations and/or close re-study of well-excavated deposits are information-rich and likely to offer stratified relationships (in formal terms a set of physical `topologies' that can be used as a chronological proxy) that will resolve many kinds of chronological debate. While largely in agreement with this common sense approach we would also argue that it may be insufficient if those contexts are simply treated in a traditional way: what is often necessary is closer attention to seemingly less interesting parts of the assemblage (e.g. the coarsewares, the undecorated pieces etc.) with greater petrographic attention to small variations in fabric for example, as these are often the most useful indicators in less advantageous contexts. Furthermore, in the absence of any opportunity for opening up or restudying a stratified sample, how should we approach the physical reinvestigation of surface material on a site (e.g. as part of a multi-stage survey)? At present, we suspect that the close reinvestigation of all isolated scatters with definite diagnostics from a problem period is a useful approach, even if the surrounding material is very sparse, as this at least raises the possibility of finding an area that is not a temporal palimpsest and can be considered in its entirety as a meaningful contemporary assemblage. Increasingly, we might also consider opportunities for the direct absolute dating of artefacts even if these methods are still very much experimental (e.g. for one promising approach, see Wilson *et al.* 2009). However, there is also need for further method-building here, and any interpretative decisions about the character of poorly understood timespans must always take into account a much wider variety of soft information about the locale in question including the negative evidence offered by the *absence* of classes of artefact we might otherwise expect to find. However, the approach outlined above, does force us to think explicitly about the diagnostic profile of particular periods, the ways in which individual pairs of periods might share a pattern of temporal uncertainty and the means by which we might isolate local areas of the landscape that deserve further attention in any given period as more likely than average to have seen period specific activity.

5.3 Regional Profiling

The third and final area in which we feel there is plenty of scope for further work is larger scale regional profiling. While we have argued that probabilistic methods, such as the ones outlined above, work in useful ways within single research projects, a further strength is the fact that they enable comparison across projects in a useful way (even in the presence of inter- and intra-observer error). For example, if a much higher shared temporal uncertainty is observed between two periods by a single excavation or survey than can be demonstrated in other projects in the same geographic area, then this may well constitute a good argument for the real local absence of one of the periods. In other words, ideally what we would like is a parent population of uncertainties for the wider regional ceramic sequence (e.g. across the whole south-west Aegean or in some cases the Mediterranean) that would then allow us to compare what we might expect in terms of finds with what we actually have. It is this measure of regional expectation that is at present very difficult to quantify, and indeed rarely even considered qualitatively (but see Fentress *et al.* 2004).

In any event, the above three topics open out a whole range of interesting questions that cannot be treated in proper detail here, but which are all given added traction by adopting this kind of probabilistic, belief-based approach and by considering the degree to which the uncertainty associated with one period is linked to the uncertainty associated with another.

Acknowledgements

The Antikythera Survey Project has been a formal collaboration between the Hellenic Ministry of Culture and the Canadian Institute in Greece. It was implemented by the authors (Bevan and Conolly) in cooperation with Aris Tsaravopoulos (Greek Archaeological Service, 26th EPKA), and we would like to thank Aris in particular for his encouragement and assistance at many stages. In addition we are grateful to the Greek Ministry of Culture and to the Canadian Institute in Greece under whose aegis this work was conducted. Our principle external funding agencies over the duration of the project have been the Social Science and Humanities Research Council of Canada, the UK Arts and Humanities Research Council and the Institute for Aegean Prehistory. Several of us have also been involved in survey work on the neighbouring island of Kythera, with the Kythera Island Project (www.ucl.ac.uk/kip), and we are grateful to Cyprian Broodbank and Evangelia Kiriatzi (KIP co-directors) for advice and support, including much useful discussion on the topics addressed here. Thanks also to Enrico Crema who offered useful comments on optimisation of the program code used in the above analysis and on a draft version of this paper. The comments of the anonymous reviewers were also extremely useful. ASP databases and photographs can be downloaded at www.ucl.ac.uk/asp and/or www.tuarc.trentu.ca/asp, are archived with the UK Archaeology Data Service (http://dx.doi.org/10.5284/1012484) and can also be obtained from Bevan or Conolly on request.

References

Alcock, S. E. and Cherry, J. F. (eds.) 2004, *Side-by-Side Survey. Comparative Regional Studies in the Mediterranean World*, Oxbow, Oxford.

Bailey, G. 2008, Time perspectivism: Origins and consequences, in *Time in Archaeology: Time Perspectvism Revisited* (eds. S. Holdaway and L. Wandsnider), 13–30. The University of Utah Press, Salt Lake City.

Banning, E. B. 2002, *Archaeological Survey*, Kluwer, New York.

Bevan, A. and Conolly, J., 2009, Modelling spatial heterogeneity and non-stationarity in artifact-rich landscapes, *Journal of Archaeological Science* 36(4), 956–964.

Bevan, A. H., in press, Spatial methods for analysing large-scale artefact inventories, *Antiquity* 86.

Bocquet-Appel, J.-P., Naji, S., Vander Linden, M., and Kozlowski, J. K., 2009, Detection of diffusion and contact zones of early farming in Europe from the space-time distribution of ^{14}C dates, *Journal of Archaeological Science* 36, 807–820.

Buck, C. E., Cavanagh, W., and Litton, C. D. 1996, *Bayesian Approach to Interpreting Archaeological Data*, John Wiley, Chichester.

Collard, M., Edinborough, K., Shennan, S. and Thomas, M. G., 2010, Radiocarbon evidence indicates that migrants introduced farming to Britain, *Journal of Archaeological Science* 37, 866–870.

Crema, E. R., in press, Modelling temporal uncertainty in archaeological analysis. *Journal of Archaeological Method and Theory*.

Crema, E. R., A. Bevan, and M. W. Lake (2010). A probabilistic framework for assessing spatio-temporal point patterns in the archaeological record. *Journal of Archaeological Science* 37(5), 1118–1130.

Cressie, N. and Wikle C., 2011, *Statistics for Spatio-Temporal Data*, Wiley, New York.

Diggle, P. J. 2006, Spatio-temporal Point Processes: Methods and Applications, in *Statistical Methods for Spatio-Temporal Systems* (eds. B Finkenstadt, L. Held, and V. Isham), 1–45, Chapman and Hall/CRC, London.

Fentress, E., Fontana, S., Hitchner, R. B., and Perkins, P. 2004, Accounting for ARS: Fineware and sites in Sicily and Africa, in *Side-by-Side Survey. Comparative Regional Studies in the Mediterranean World* (eds. S. E. Alcock and J. F. Cherry), 147–162, Oxbow, Oxford.

Galton, A. 2004, Fields and objects in space, time and space-time. *Spatial Cognition and Computation* 4(1), 39–68.

Green, C., 2011, *Winding Dali's Clock The Construction of a Fuzzy Temporal-GIS for Archaeology*, Archaeopress, Oxford.

Gregory, T. 2004, Less is better: The quality of ceramic evidence from archaeological survey and practical proposals for low-impact survey in a mediterranean context, in *Mediterranean Archaeological Landscapes: Current Issues* (eds. E. F. Athanassopoulos and L. Wandsnider): 15– 36, University of Pennsylvania Museum, Philadelphia.

Grove, M. 2011, A spatio-temporal kernel method for mapping changes in prehistoric land-use patterns, *Archaeometry* 53(5), 1012–1030.

Hazelton, M. L. and Davies, T. M. 2009, Inference based on kernel estimates of the relative risk function in geographical epidemiology, *Biometrical Journal* 51(1), 98–109.

Johnson, I. 2004, Mapping artefact distributions through time using aoristic analysis, in *Proceedings Computer Applications in Archaeology Conference, Vienna, April 2003* (ed. W. Borner), Tempus Reparatum, Oxford

Kelsall, J. E. and Diggle, P. J. 1995, Non-parametric estimation of spatial variation in relative risk, *Statistics in Medicine* 14, 2335–2343.

Lloyd, C. D., 2007, *Local Models for Spatial Analysis*, CRC/Taylor and Francis, Boca Raton.

Lock, G. and Harris, D., 2002, Analysing change through time within a cultural landscape: conceptual and functional limitations of a GIS approach, in *The Development of Urbanism from a Global Perspective* (ed. P. Sinclair), 1–13, Uppsala Universitet, Uppsala.

Lyman, R. L. and O'Brien, M. J., 2006, M*easuring Time with Artifacts: A History of Methods in American Archaeology*, University of Nebraska Press, Lincoln.

Moore, R. S. (2008). A decade later: The chronotype system revisited, in *Archaeology and History in Roman, Medieval and Post-Medieval Greece: Studies on Method and Meaning in Honor of Timothy E. Gregory* (eds. W. Caraher, L. J. Hall, and R. S. Moore), 137–151, Ashgate, Aldershot.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., and Rakow, T., 2006, *Uncertain Judgements: Eliciting Experts' Probabilities*, Wiley, Chichester.

Ortman, S. G., Varien, M. D. and Gripp, T. L., 2007, Empirical Bayesian methods for archaeological survey data: an application from the Mesa Verde region. *American Antiquity* 72(2), 241–272.

Pentedeka, A., Kiriatzi, E., Spencer, L., Bevan, A. and Conolly, J. 2010, From fabrics to island connections: macroscopic and microscopic approaches to the prehistoric pottery of Antikythera, *Annual of the British School at Athens* 105, 1–81.

Pettegrew, D. K., 2007, The busy countryside of Late Roman Corinth: Interpreting ceramic data produced by regional archaeological surveys, *Hesperia* 76, 743–784.

Peuquet, D. J. 2002, *Representations of Space and Time*, Guilford, New York.

Quercia, A., Johnston, A., Bevan, A., Conolly, J. and Tsaravopoulos, A. 2011, Roman pottery from an intensive survey of Antikythera, Greece, *Annual of the British School at Athens* 106, 47-98.

Ratcliffe, J. H., 2000, Aoristic analysis: the spatial interpretation of unspecific temporal events, *International Journal of Geographical Information Science* 14(7), 669–679.

Rossignol, J. and Wandsnider, L., 1992, *Space, Time, and Archaeological Landscapes*, Plenum Press, New York.

Shafer, G. 1990, Perspectives on the theory and practice of belief functions, *International Journal of Approximate Reasoning* 4, 323–362.

Tartaron, T. F., Gregory, T. E., Pullen, D. J., Noller, J. S., Rothaus, R. M., Rife, J. L., Tzortzopoulou-Gregory, L., Schon, R., Caraher, W. R., Pettegrew, D. K. and Nakassis, D. 2006, The Eastern Korinthia Archaeological Survey: integrated methods for a dynamic landscape, *Hesperia* 75(4), 453– 523.

Vroom, J., 2005, *Byzantine to Modern Pottery in the Aegean: An Introduc- tion and Field Guide,* Parnassus Press, Utrecht.

Wilson, M. A., Carter, M. A., Hall, C., Hoff, W. D., Ince, C., Savage, S. D., McKay, B. and Betts, I. M., 2009, Dating fired-clay ceramics using long-term power law rehydroxylation kinetics, *Proceedings of the British Academy* A, 1–9.