Journal of Archaeological Science: Reports

Re-evaluating archaeomagnetic dates of the vitrified hillforts of Scotland

Neil Suttie, Catherine M. Batt

A re-analysis of archaeomagnetic data from seven vitrified hillforts in Scotland, sampled in the 1980s, shows excellent agreement with recent radiocarbon dates. In the past thirty years our knowledge of the secular variation of the geomagnetic field has greatly improved, especially in the 1st millennium BC, allowing earlier archaeomagnetic data to be reconsidered. We evaluate the likelihood of the data with respect to a state-of-the-art field geomagnetic model and find close coherence between the observed directions and the model for the closing centuries of the first millennium BC. A new Bayesian method of calibration gives the most likely number of separate events required to produce a series of magnetic directions. We then show that the burning of three of the four oblong forts most likely took place around the same time, and our estimate for the date of this is indistinguishable from recent radiocarbon dates from another fort of similar type.

1. Introduction

The vitrified hillforts of Northern Scotland have inspired debate among archaeologists since the 18th C (Ralston, 2006: 148), with various theories being proposed for the purpose and method of their vitrification. Their place in the chronology of the Scottish Iron Age also remains uncertain: early attempts to use radiocarbon by Mackie (1969) gave date ranges that, when calibrated using modern methods, are considered too large to be useful (Alexander, 2002), though are consistent with activity in the first millennium BC. More recently excavations at Dunideer yielded two radiocarbon dates from burnt wood of 390–190 cal BC and 370–160 cal BC (Cook et al., 2010). The wood, it was argued, was unlikely to have been more than a few years old at the time of its combustion, although the date should strictly be interpreted as a *terminus post quem* for the firing of the fort. Thermoluminescence (Sanderson et al., 1985) and archaeomagnetism Gentles (1989, 1993) have been used to date the actual firing of the structures but the results were inconclusive at the time. The thermoluminescence dates for six of the hill-forts span some 2500 years, which many archaeologists find unconvincing for sites so similar in form (Ralston, 2006: 150). On the other hand, the archaeomagnetic dates for vitrified oblong forts were more tightly clustered around the end of the first millennium BC, in keeping with many archaeologists' expectations for this type of structure (e.g. Ralston 206: 151, Armit, 1997:108).

Gentles (1989) sampled seven vitrified forts for archaeomagnetic dating, which compares the direction of remanent magnetisation held by the heated rocks with a reference curve depicting the change of the Earth's magnetic field through time. Six of the structures (Knock Farril, Craig Phadrig, Finavon, Tap O'Noth, Langwell and Dun Skeig) yielded useful magnetic data with high enough precision to constrain the date of firing. The locations of the six sites are shown in Fig. 1. Apart from Dun Skeig, the five structures gave broadly consistent directions, and Gentles concluded that the oblong forts at Knock Farril, Craig Phadrig, Finavon and Tap O'Noth had been burned roughly contemporaneously, with the dun type forts Langwell and Dun Skeig being burned later. Using the secular variation curve available at the time (Clark et al., 1988), five of the structures (apart from Dun Skeig) were dated to between 200 BCE and 150 CE. In some cases, the precision reported was very high, with 95% confidence limits spanning as little as 90 years in three cases. At the time, ages were assigned by simple comparison of the data with a reference curve, and the precision reported is probably over-optimistic although the archaeomagnetic data is itself of good quality. Importantly, by distributing samples across the structures, Gentles (1989) was able to show consistency of the magnetic directions from within each site and was hence able to conclude that no largescale differential movement had occurred and that the directions he acquired should represent the actual geomagnetic field at the time of firing.

In recent years, our understanding of the evolution of the geomagnetic field has greatly increased with a number of both regional (Batt et al., 2017;, Pavón-Carrasco et al., 2009) and global models now available (Korte et al., 2011; Nilsson et al., 2014; Hellio and Gillet, 2018; Licht et al., 2013), which may change the interpretation of the data discussed above. Archaeomagnetic assay necessarily involves the removal of material from the structure to be sampled, and it is clearly

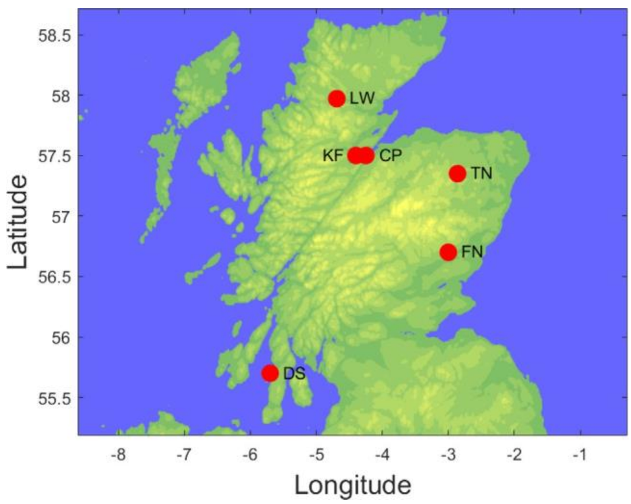

Fig. 1. Map showing the locations of the sites discussed: Knock Farril (KF), Craig Phadrig (CP), Finavon(FN) ,Tap O'Noth (TN), Langwell (LW) and Dun Skeig (DS).

undesirable to resample such important cultural sites as these hillforts. It is therefore important to evaluate the existing data and reinterpret it in light of recent advances in the field. The purpose of this paper is to use a state-of-the art field model and modern Bayesian methods to re-analyse the data from Gentles (1989) to both provide a new chronology for the vitrification events and to test the hypothesis of contemporaneity of events between sites.

2. Preliminaries

2.1. Models

In this section we will describe the secular variation model that will be used in the analysis. As the data from Gentles (1989) is already in the database that is usually used to construct field models (along with the dates derived by TL which themselves seem unrealistic on archaeological grounds), there is a danger that prior assumptions may influence our results. For instance, all archaeomagnetic data from around the world was used to construct the ARCH3k1 model (Korte et al., 2009) which itself was used to constrain the ARCH-UK.1 model (Batt et al., 2017), so the data that we wish to analyse already influences both of these models. However, Korte et al. (2009) also selected a subset of the available archaeomagnetic data with good chronological controls to construct the model ARCH3k1_cst. This model is not influenced by the data we are analysing here, as it did not have sufficient chronological control to be used, and we believe it is a good representation of the Earth's magnetic field for the location. A drawback is that it is not provided with an error envelope. In what follows we will assume that ARCH3k1_cst is an accurate representation of the field and acknowledge that to some extent our conclusions will be conditional on that assumption.

2.2. Data

The data used is the same data as Gentles (1989) used to calculate the site means. We have not reanalysed any outlier rejection, but note that there were very few sites where a stable magnetic direction was not used, although at Tap O'Noth two directions were thought to be the result of movement or collapse of a section of wall and were discarded. Directions were relocated to a single location (55°N 3°W) using the method of Noel and Batt (1990), although this correction made minimal difference. The complete data set is given in the supplementary material, and the individual site means with their 63% confidence ellipses ($\alpha_{63} = 4/7\ \alpha_{95}$) and the model prediction for 55°N 3°W are shown in Fig. 2. Dun Skeig (DS) appears somewhat later than the other five sites Knock Farril (KF n = 12), Craig Phadrig (CP n = 8), Finavon (FN n = 12), Tap O'Noth (TN n = 11) and Langwell (LW n = 35), where n is the number of individual magnetic directions at each site. We expect the α_{63} circle of confidence to include the mean direction in at least four out of the six sites and note that the model cuts the error circle in 5 cases. TN is a slightly further from the model but is still consistent within its 95% uncertainty bounds. Given that there is reasonable consistency between model and data, which was not used to construct the model, we now ask which possible chronology is most consistent with the observed data. To do this we employ a novel Bayesian method of hypothesis comparison.

3. Analysis

There are two questions which must be addressed: how many chronologically separate events can be identified and what were their dates? In saying chronologically separate, we mean that events took place with enough time lapsed between them for their dates to be distinguished by both the data precision and the change in the geomagnetic field, a point that will be considered further in the discussion (Section 4.2). We have assumed the model to be accurate so the probability of a date for the j^{th} site can be expressed using Bayes' theorem.

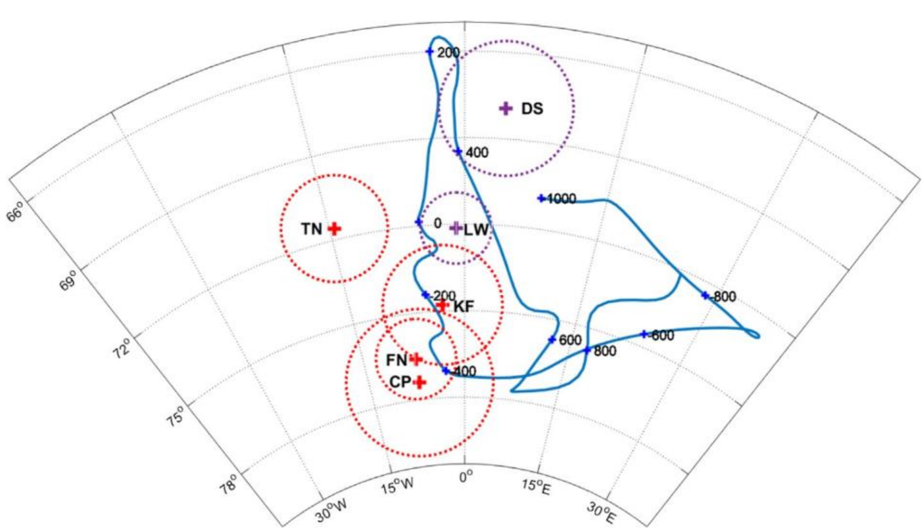

Fig. 2. Archaeomagnetic directions for the six sites along with the field model ARCH3k_cst. Circles are 63% cones of confidence of the mean (crosses). The oblong type forts Knock Farril (KF), Craig Phadrig (CP), Finavon(FN) and Tap O'Noth (TN) are shown in red, with the dun type forts of Langwell (LW) and Dun Skeig (DS) in purple.

$$P(a|D_j, M) = \frac{P(D_j|a, M)P(a|M)}{\sum_a P(D|a, M)P(a|M)} \quad (1)$$

Equation (1) says the posterior probability (P) of age (a) given data (D_j) and model (M) is the product of the likelihood of the data given the age multiplied by the prior probability of the age divided by the "marginal probability". The denominator is expressed as a summation to emphasise that the calculations are carried out on a discrete grid, evaluated at the 10 year intervals that the model coefficients are supplied at. Unlike the likelihood, prior and posterior, in Bayesian analysis, there is no universally agreed name for this term, with "prior predictive", "marginal likelihood" and "evidence" all being used (Skilling, 2006). We will use the term "evidence" (Z) with

$$Z_j \equiv \sum_a P(D_j|a, M)P(a|M) \quad (2)$$

Note that Z is simply the average value of the likelihood over the age range, if the prior probability of the age is uniform.

We assign a Fisher distribution of unknown dispersion k to each set of data and marginalise the unknown dispersion k:

$$P(D_j|a, M) = \int \prod_i \frac{k}{2\pi} \exp k(\mathbf{v}_i \cdot \mathbf{u}_a - 1)P(k)dk \quad (3)$$

In Eq. (2), $\mathbf{u}_a$ is the unit direction vector for the model at time a and $\mathbf{v}_i$ is the unit direction vector for the i^{th} datum in D_j and P(k) is the prior distribution for the dispersion parameter k, for which we use a uniform distribution between 10 and 500. Similarly we assign a uniform prior for the age between limits of 1000 BCE and 1000 CE. Fig. 3 shows the posterior probability for each of the individual sites. There is a clear overlap of the age ranges for five of the sites, with Dun Skeig having a later age, 95% of the posterior lying between 70 and 420 CE.

A key archaeological question is whether the remaining five sites should be explained as several different events with appreciable time between them, or by fewer periods of burning. Suppose we have m sites. We want to know if we should assign one common age for the burning at all m sites (hypothesis A), or whether we should assign m different ages (hypothesis B). We compare the posterior probability of each hypothesis using Bayes theorem:

$$\frac{P(A|D, M)}{P(B|D, M)} = \frac{P(D|A, M)P(A|M)}{P(D|B, M)P(B|M)} \quad (4)$$

If the prior probabilities for each hypothesis are judged to be equal, $P(A|M) = P(B|M)$. The numerator of the right hand side of (3) is the average likelihood of the combined data over the prior age, which is just Z (Eq. (2)). The denominator of the right hand side of (4) is the product of the individual average likelihoods for each site. For the j^{th} site we define p_{aj} as the likelihood of age a:

$$p_{aj} = P(D_{aj}|a, M)$$

The posterior odds in (3) become

$$\frac{P(A|D, M)}{P(B|D, M)} = \frac{P(A|M)\sum_a \prod_j p_{aj}}{\prod_j P(A|M)\sum_a p_{aj}}. \quad (5)$$

The odds ratio in (5) is the average likelihood of the combined data, divided by the product of the average likelihoods of the site data. If it is much larger than 1, we should prefer a single age for the activities to the hypothesis of m different ages, on the basis of the archaeomagnetic data.

To get an intuitive feel for the odds ratio, it is useful to give it in decibels (Jaynes, 2003) i.e. 10 times the logarithm to the base 10 of (Eq. (5)). Roughly, a hypothesis would be slightly preferred at 3–10 dB, more so at 10–20 dB, and greater than 20 dB would constitute strong evidence for one hypothesis relative to another. The convenience of this method is that being additive, one can compare any 2 hypotheses simply be considering their difference. This makes it possible to compare the hypotheses of any combinations of sites as belonging to the same period of activity. There are a total of 52 ways of grouping the five sites, so it is straightforward to enumerate every possibility. Some possibilities are, compared to assigning 5 different ages (0 dB):

One age for all 5 sites: −7 dB
(CP + FN)(KF)(LW)(TN): +8.7 dB
(CP + FN + KF)(LW)(TN): +14.7 dB
(CP + FN)(KF + LW + TN): +18.6 dB
(CP + FN + KF)(LW + TN): +22.3 dB

The highest probability is found for two separate ages, one accounting for Knock Farril, Craig Phadrig and Finavon and one accounting for Tap O'Noth and Langwell. The evidence is strongest for two discrete periods of firing, but there is only 3.7 dB preference for placing Langwell in the first rather than the second group i.e. it is

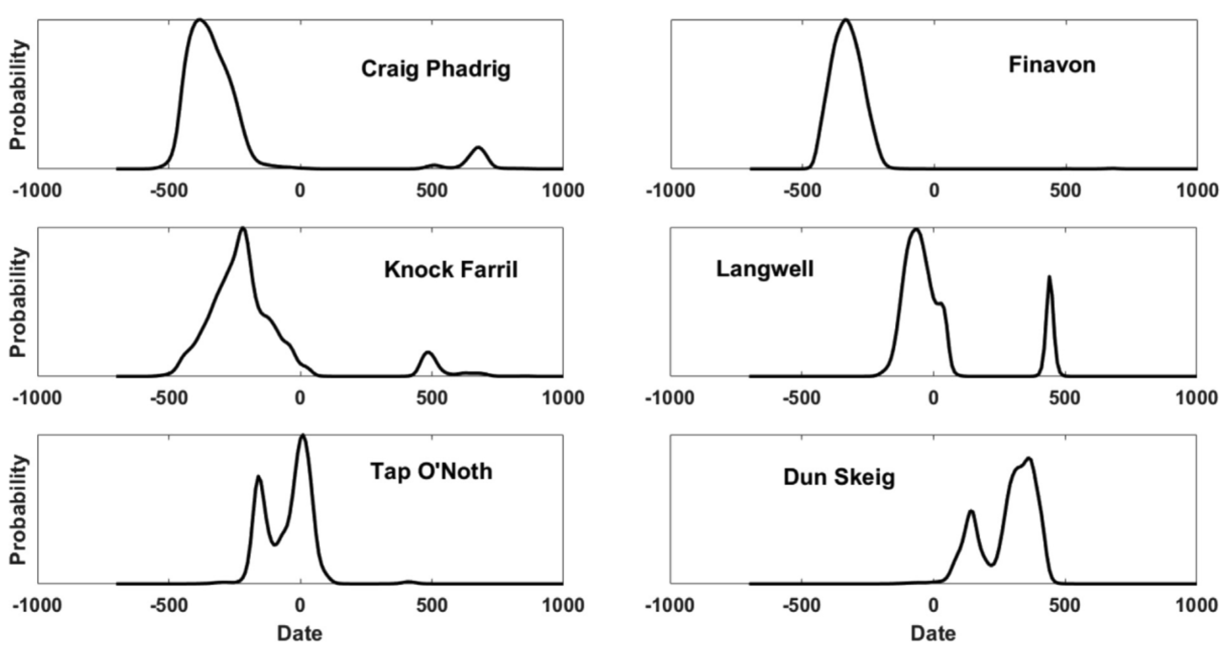

Fig. 3. Posterior probabilities for the age of each of the sites considered individually.

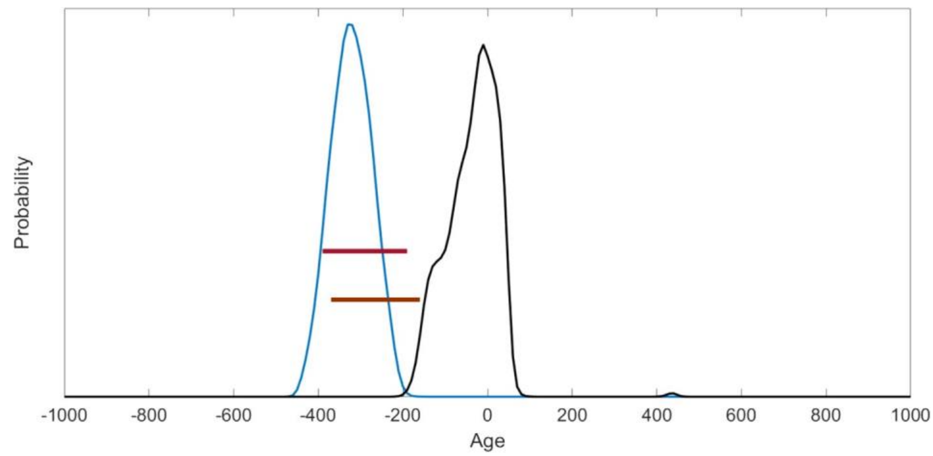

Fig. 4. Posterior probabilities for the preferred age model, assigning the sites to two groups. The first group (blue) contains Craig Phadrig, Knock Farril and Finavon, and the second group (black) contains Tap O'Noth and Langwell. Radiocarbon 95% confidence ranges for Dunideer (Cook et al., 2010) are shown as bars.

approximately 2 ½ times more likely to belong to the first group. Posterior age distributions for the groups of highest probability are shown in Fig. 4; the best fitting Gaussian approximations have mean and standard deviations of 320 BCE ± 75 years and 25 BC ± 80 years. The earlier dates are in close agreement with the radiocarbon dates from the typologically similar site at Dunideer, discussed in the introduction.

4. Discussion

4.1. Analytical considerations

Although no use is made of the sampling distribution of the mean, or its oft cited statistic, α_{95}, the distributions shown in Fig. 2 could similarly be derived using the distribution of the mean given by Fisher (1953). It is however the evidence function which plays a key role in the analysis presented here. The evidence (Eq. (2)) is the average of the likelihood over the age. The algorithm that is used to find which sites should be grouped together chronologically can be expressed in terms of this average. The evidence for a number of distinct events is the product of the average likelihoods for each site; the evidence for a single event is the average of the product of the likelihoods. Despite its simplicity the algorithm has not been used previously, although the question as to whether two archaeomagnetic directions are coeval commonly occurs. Often this question is addressed by means of a tail-test: a null hypothesis supposes that two mean directions are the same, and then the distribution of some t-type statistic is calculated. The method often used to test palaeomagnetism directions: the so-called reversal test (McFadden and McElhinny, 1990) involves some classification of the results according to a somewhat arbitrary code. A common feature of Bayesian analysis is that it seeks to compare competing hypothesis directly, rather than reject a null hypothesis without specifying a viable alternative.

Another important point is that the marginal probability of the data (i.e. Z in Eq. (2)) is a crucial part of the analysis whereas in general Bayesian calculation it is only considered to be only a normalising factor. As the prior age distribution appears explicitly in Z, the results depend partially on the size of the prior age assignment, which here covers 2000 years. If instead we only considered a prior age range of 1000 years, the odds favouring one date over two would drop by 3 dB, assuming the posterior was not significantly truncated. Although such a dependency might seem subjective, it is quite natural that what is considered to be a single phase of activity within a long time frame, may appear to show more nuances when looked at more closely. The width of the prior age distribution is analogous to the field of view of a microscope: a narrower field reveals more detail, but at the risk of missing important features.

4.2. Interpretation

In the absence of decadal scale dating, as might be afforded by dendrochronology or tephrachronology, the precision of scientific dates for the late Iron Age are unlikely to improve greatly upon the ± 100 years at 95% confidence reported by Cook et al. (2010) for radiocarbon from Dunideer. Archaeomagnetic dates might be possible with a similar precision, although the preferred model presented here has slightly larger uncertainties. Given that the modelled field moves by around 3–4 degrees per century, a sampling precision less than this would be a prerequisite to improve the precision of the archaeomagnetic dates. Ignoring Tap o' Noth, the evidence from archaeomagnetism and the recent radiocarbon from Dunideer combine to suggest that the oblong forts were burnt is a single phase of activity, sometime in the 3rd or 4th century BCE. To avoid the sort of confusion regarding scientific dates highlighted by Alexander (2002), it is worthwhile to attach some caveats to this conclusion.

Firstly, it should be realised that errors assigned to a scientific date are derived from data uncertainties; such aleatoric errors cannot account for insufficiencies of the model used and there must always be a chance that future work will lead to a revision of dates, in much the way that this paper seeks to do, and as has been seen in the radiocarbon community over recent years. Secondly, we should clarify what we mean by a single phase of activity. Formally, the Bayesian calculation compared the probabilities of the activities being synchronous on a 10 year grid, and hence the conditional probability of two activities taking place 10 years apart contributes to the marginal probability of their being distinct. This rather unrealistic precision was used for computational ease, and we do not suggest that the method can distinguish between events that happened over the course of a century or so. The 1 sigma error on the archaeomagnetic date ranges is probably a fair indication of the sort of timescale that the method is sensitive to. What we call a single phase of activity might possibly have occurred over a few generations. Thirdly, the later date for Tap o'Noth should be treated cautiously for a number of reasons. Grouping it with other oblong forts would not change the age range for the group by a great amount, making it slightly later and, interestingly, even more consistent with the radiocarbon from Dunideer. The evidence for a separate date from the other three oblong forts can be quantified as 10 dB using eq.5, which while significant, would be rendered meaningless by a single radiocarbon date, for instance. A later date for the burning activity at Tap o' Noth should probably be regarded as an interesting hypothesis that could only be verified by further excavation, as called for by Hunter (2007).

5. Conclusions

A reanalysis of the archaeomagnetic data from Scottish vitrified hillforts first presented by Gentles (1989) suggests that the oblong hillforts were mainly burned in the 4th–3rd centuries BC. Using the field model ARCH3k_cst (Korte et al., 2009) as a reference, the sequence of events found to be most consistent with the archaeomagnetic data is that of two main episodes of burning. The first event saw the firing of the forts at Finavon, Knock Farril and Craig Phadrig with 95% of the credible interval lying between 430 and 230 BCE. The second phase of activity suggested includes the fires at Tap O'Noth and Langwell, with a 95% credible interval between 190 BCE and 50 AD. Other scenarios can certainly not be ruled out on the strength of the archaeomagnetic data: the possibility that Tap O' Noth belongs to the first group should be considered, which would yield a 95% credible interval of 380 BCE to 190 BCE for the burning of the four oblong forts, which is in excellent agreement with the radiocarbon dates of 390–190 cal BC and 370–160 cal BC from the typologically similar site of Dunideer (Cook et al., 2010). Two features of this study should be considered in relation to other sites. Firstly, the use of an independent field model, if avail- able, allows the revision of old archaeomagnetic dates and should be considered where appropriate. Secondly, the Bayesian method of grouping age distributions could further constrain the ages of sites where sequences of age distributions, derived by any method, are available.

6. Author statement

NS envisioned and carried out all analyses.
CMB provided support and archaeological background.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Alexander, D., 2002. An oblong fort at Finavon, Angus: an example of the over-reliance of the appliance of science. In: Ballin Smith, B., Banks, I. (Eds.), In the Shadow of the Brochs. Tempus, Stroud, pp. 45–54.

Armit, I., 1997. Celtic Scotland. Batsford, London.

Batt, C.M., Brown, M.C., Clelland, S.-J., Korte, M., Linford, P., Outram, Z., 2017. Advances in archaeomagnetic dating in Britain: New data, new approaches and a new calibration curve. J. Archaeol. Sci. 85, 66–82.

Clark, A.J., Tarling, D.H., Noel, M., 1988. Developments in archaeomagnetic dating in Britain. J. Archaeol. Sci. 15, 645–667.

Cook, M., Kdolska, H., Dunbar, L., Engl, R., Sagrott, S., Druce, D., Cook, G., 2010. New light on oblong forts: Excavations at Dunnideer, Aberdeenshire.

Fisher, R., 1953. Dispersion on a sphere. Proc. R. Soc. Lond. A 217, 295–305. Gentles, D., 1989. Archaeomagnetic Directional Studies of Large Fired Structures in Britain. Unpublished PhD thesis. University of Plymouth.

Gentles, D., 1993. Vitrified forts. Curr. Archaeol. 133, 18–20.

Hellio, G., Gillet, N., 2018. Time-correlation-based regression of the geomagnetic field from archeological and sediment records. Geophys. J. Int. 214, 1585–1607.

Hunter, F., 2007. Beyond the Edge of Empire: Caledonians, Picts and Romans. Groam House, Inverness.

Jaynes, E.T., 2003. Probability Theory [Electronic Book]: The Logic of Science. Cambridge University Press, Cambridge, UK; New York, NY.

Korte, M., Constable, C., Donadini, F., Holme, R., 2011. Reconstructing the Holocene geomagnetic field. Earth Planet. Sci. Lett. 312, 497–505.

Korte, M., Donadini, F., Constable, C.G., 2009. Geomagnetic field for 0–3 ka: 2. A new series of time-varying global models. Geochem., Geophys., Geosyst. 10.

Licht, A., Hulot, G., Gallet, Y., Thébault, E., 2013. Ensembles of low degree archeomagnetic field models for the past three millennia. Phys. Earth Planet. Inter. 224, 38–67. https://doi.org/10.1016/j.pepi.2013.08.007.

MacKie, E., 1969. Radiocarbon dates and the Scottish Iron Age. Antiquity 43, 15–26.

McFadden, P.L., McElhinny, M.W., 1990. Classification of the reversal test in palaeomagnetism. Geophys. J. Int. 103, 725–729.

Nilsson, A., Holme, R., Korte, M., Suttie, N., Hill, M., 2014. Reconstructing Holocene geomagnetic field variation: new methods, models and implications. Geophys. J. Int. 198, 229–248.

Noel, M., Batt, C.M., 1990. A method for correcting geographically separated remanence directions for the purpose of archaeomagnetic dating. Geophys. J. Int. 102, 753–756.

Pavón-Carrasco, F.J., Maria Luisa, O., Torta, J., Gaya-Piqué, R.L., 2009. A regional archeomagnetic model for Europe for the last 3000 years, SCHA.DIF. 3K: applications to archeomagnetic dating. Geochem. Geophys. Geosyst. https://doi.org/10.1029/2008GC002244.

Ralston, I., 2006. Celtic Fortification. Tempus, Stroud.

Sanderson, D.C.W., Placido, F., Tate, J.O., 1985. Scottish vitrified forts: Background and potential for TL dating. Nucl. Tracks Radiat. Meas. 1982 (10), 799–809.

Skilling, J., 2006. Nested sampling for general Bayesian computation. Bayesian Anal. 1, 833–860.