Journal of Archaeological Science 118 (2020) 105138

Contents lists available at ScienceDirect

Journal of Archaeological Science

journal homepage: http://www.elsevier.com/locate/jas

# A probabilistic framework and significance test for the analysis of structural orientations in skyscape archaeology

Fabio Silva

*Institute for the Modelling of Socio-Environmental Transitions, Faculty of Science & Technology, Bournemouth University, United Kingdom*

ARTICLE INFO

*Keywords:*
Skyscape
Landscape
Orientation
Modelling
Statistics
Significance testing

ABSTRACT

The assessment of the orientation of built structures is often hampered by the lack of historical or ethnographic records that could be used to support claims of alignments to celestial events such as the rising and settings of sun, moon or stars. This has led to an obsession with surveying large numbers of structures in order to identify patterns in orientation that might betray intentionality - however, no first-principles framework has ever been proposed and the few statistical significance tests used are of limited applicability. This paper addresses these problems by laying out a probabilistic framework for the analysis of structural orientations and using it to develop a test of statistical significance. The framework is based on two simple premises: firstly, that the measurement of a structural orientation can be modelled as a probability distribution; and secondly, that in order to assess the likelihood of celestial alignments such a distribution needs to be coordinate-transformed in a manner not unlike that of radiocarbon calibration. A method that aggregates multiple structural orientations and quantifies their statistical significance in the form of a p-value is then introduced. Finally, the robustness of the presented methodology over previous approaches is demonstrated using real-world datasets of orientations of ancient Egyptian temples and tombs and Scottish Recumbent Stone Circles, and the ensuing conclusions compared with past interpretations.

## 1. Introduction

The search for meaning in the orientation of built structures, particularly in prehistory, is as old as archaeology. Already at the turn of the 20th century there were those that, for example, saw intent behind the careful orientation of Stonehenge and its correlation with the solstice sun (e.g. Lockyer, 1906). This interest eventually led to a peak in the popularity of archaeoastronomy in the 1960s and 70s, despite the contested nature of the claims and interpretations being made (e.g. Hawkins, 1964; Thom, 1967). The field then saw a renaissance of sorts by grounding itself in the analysis of field measurements (as opposed to taking measurements from plan drawings of unknown accuracy) and by paying attention to the unavoidable statistical argument that correlations between structural orientation and celestial objects can occur by chance alone, and therefore should not be assumed *a priori* to encode any prehistoric intent or meaning (Ruggles, 1999; Hutton, 2013).

Further afield, archaeologists interested in how people sit and engage with the landscape also began to pay attention to how people positioned their structures as well as how they orientated them (e.g. Tilley, 1994). This brand of landscape archaeologist was, however, less interested in accurate measurements or analytical techniques and more engrossed in phenomenological observations and post-processual interpretations. It was with the advent of computation that spatial analysis became a well-established and respected archaeological field of inquiry with a robust statistical backbone (e.g. Bevan and Lake, 2016), which has also led to the statistical testing of some of the post-processual interpretations and hypotheses (e.g. Eve and Crema, 2014).

Archaeoastronomy too underwent a flurry of interest in new statistical methodologies in the 1980s (e.g. Freeman and Elmore, 1979; Ruggles, 1984) but failed to maintain this momentum, or to take advantage of developments in the sister field of landscape archaeology. As a result, archaeoastronomy has stalled over the past 40 years (Ruggles, 2011): it still largely relies on data collection techniques first delineated more than one hundred years ago (Lockyer, 1909) and on data visualisation methods that were first used by Thom (1967). For example, despite interest in quantifying the *astronomical* significance of alignments, a widely applicable *statistical* significance test for structural orientations was never devised.

Lately, there is a movement to reframe the field as a *skyscape archaeology*, which applies "a degree of reflexivity well above the norm

*E-mail address:* fsilva@bournemouth.ac.uk.

https://doi.org/10.1016/j.jas.2020.105138
Received 6 November 2019; Received in revised form 11 March 2020; Accepted 23 March 2020
Available online 24 April 2020

of archaeoastronomy in the past" and demands "total immersion in the society we are studying" (Silva and Henty, 2018: 3). Reflexivity applies not only qualitatively to one's interpretation and assumptions, but also quantitatively to the quality and nature of one's data and to the limitations and underlying assumptions of one's analytical methods. This paper proposes to do the latter, by focusing in particular on the over-reliance on what Ruggles called the "qualitative assessment" (2015b: 420) of frequencies of orientations. This widespread approach involves the creation of histograms or *curvigrams*, which are kernel density estimates using a normal kernel with varying bandwidth values, whose peaks are then looked at in search of landscape or skyscape correlates (e.g. Silva, 2017). This involves little to no analytical care with whether those peaks are statistically significant, which is to say whether they cannot be said to arise solely by chance.

A more robust approach started to appear within the Spanish school of archaeoastronomy (Belmonte et al., 2013; González-García and Belmonte, 2011, 2014; González-García and Sprajc, 2016). It consists of comparing the curvigram constructed from the structural measurements with a curvigram based on a random distribution of orientations. The method compares the two and outputs z-values (in units of standard deviation, σ) which indicate how far the empirical curvigram is from the randomly derived one. Following Schaefer (2006), one can then say that only orientations with z-values above the 3σ level can be said to be significant and therefore should form the backbone of the search for a topographic or celestial explanation. Although this method is a significant stride in the right direction, it is not without limitations. Firstly, one is comparing an empirical curvigram constructed from a realistic, and hence small, sample of orientations with a curvigram constructed from an infinite number of random orientations, i.e. with a limiting distribution. Realistic samples of random orientations, with the same characteristics as the empirical dataset, are likely to feature peaks and troughs that are not being taken into account by this method and, therefore, throw serious doubts over its conclusions, especially when used on small datasets. And secondly, this method does not produce a measure of significance for the entire dataset, but rather it estimates z-values as a function of orientation, making it difficult to interpret and reach consensual conclusions regarding a dataset's significance.

In statistics, significance tests are specific cases of the more general Neyman-Pearson hypotheses testing approach (Neyman and Pearson, 1933; Neyman, 1950) where, instead of comparing two competing hypotheses and subsequent selection of one over the other, the purpose of the significance test is to measure the evidence against a single hypothesis – the *null hypothesis*. The result of a significance test is a *p-value*, which indicates the 'strength of the evidence against the [null] hypothesis' (Fisher, 1958, 80) and is more correctly interpreted as the probability of observing an outcome that is as extreme as, or more extreme than, that which is observed empirically, assuming the null hypothesis to be true. In other words, if the null hypothesis does not hold, then the empirical observations will be highly unlikely to occur, and hence the estimated p-value is expected to be very low. A currently accepted cross-disciplinary threshold for significance is that p-values of 0.05 or smaller, corresponding to a mere 5% probability of occurrence under the null hypothesis, are considered statistically significant. This choice is, however, completely arbitrary, and there has been a recent call for this threshold to be changed to 0.005 (Benjamin et al., 2017), as well as Schaefer's already mentioned call to use 3σ in archaeoastronomy, which corresponds to a p-value of 0.001 (Schaefer, 2006). Since its formulation, significance tests have become the most widely used statistical approach crossing such varied fields as the health sciences, the natural sciences and the humanities (e.g. Benjamin et al., 2017).

We can, probably safely, assume that archaeoastronomy's general lack of engagement with significance testing is not due to a lack of awareness on the part of archaeoastronomers (most of whom had training in the exact sciences) but due to the lack of a specific algebraic test that could be applied to the intricacies of structural orientation data, in the same manner that, for example, a Chi-squared test can be used for categorical data. With the advent of the computer age, numerical tests and solutions are now possible where previously they would have taken months, or even years, to conduct manually. Efron and Hastie (2016), have called for 21st-century statisticians to move away from 20th-century algebraic tests, which are based on assumptions and simplifications that allow calculations to be made manually but that can limit their range of applicability, and instead embrace the full computational power that is widely available today. This can be done through the employment of robust, though brute-force, approaches which can be applied in almost every scenario and often do not involve the assumptions and simplifications of the algebraic tests.

This is exactly the aim of this paper: instead of prowling the historical statistical texts in search of the "right" test for structural orientations, it develops a foundational probabilistic framework from which a computational statistical significance test can be developed and employed. The methodology, which is detailed below, was implemented in the R software environment for statistical computing (R Core Team, 2019) and builds upon the R packages *skyscapeR* (Silva, 2019a) and *swephR* (Stubner and Reijs, 2019) which handle the astronomical computations. The source code, as well as all necessary data files to reproduce the figures and tables, are available are in the authorsupplementary material, as are all necessary files to reproduce the figures and tables. When the paper gets accepted for publication the code will be made available open-access and open-source on the author's GitHub page and a link will be added here.

## 2. Measurement as a probability distribution

Any statistical analysis starts from a dataset of observations – in our case the orientation(s) of a given set of similar structures. Once a particular architectural feature whose orientation one wants to analyse is identified, it can be relatively trivial to measure its orientation using standard surveying instruments and techniques. Nonetheless, such a measurement – indeed, any measurement – is but an approximation or estimate of the real value that is being measured (Taylor, 1997). And just like any other estimate it has an associated uncertainty, which often needs to be estimated (e.g. Silva, 2019b). To take uncertainty into account, and to establish the foundations of a probabilistic framework, one must begin by choosing a mathematical model to represent the measurement of structural orientation.

Since measurements are just approximations, every time one is taken it will be different from the previous one. With a very large number of measurements one would be able to construct a *limiting distribution*, which is to say a mathematical function that "describes the proportion of times a repeated measurement yields each of its various possible answers" (Taylor, 1997: 227). In the absence of large numbers of measurements, one can still use a *probability distribution* (or *probability density*) to represent a single measurement, this time describing the distribution of the likelihood of where the true value being measured lies with respect to the measurement itself. The shape of this probability distribution represents the distribution of likelihood and, as such, it is up to the scholar to choose which shape best represents their confidence in where the correct structural orientation and/or alignment lies within the range given by the best estimate plus or minus the uncertainty (Fig. 1). Examples of distribution shapes can be taken from, for example, the many kernels used for kernel density estimation (e.g. Zucchini, 2003: 8). Here, however, we will restrict ourselves to two contrasting examples: the normal and uniform distributions.

A normal distribution gives more weight to the mean value and the probability of finding the true value drops as one moves away from the mean (Fig. 1A). It can be used to model an orientation measurement by, for example, saying that its mean value is equal to the measurement's best estimate and that its standard deviation is equal to half of the estimated uncertainty. One must always keep in mind that, because the normal distribution never drops to zero, one is assuming that, however unlikely, it is still possible that the true value is five, ten or a hundred

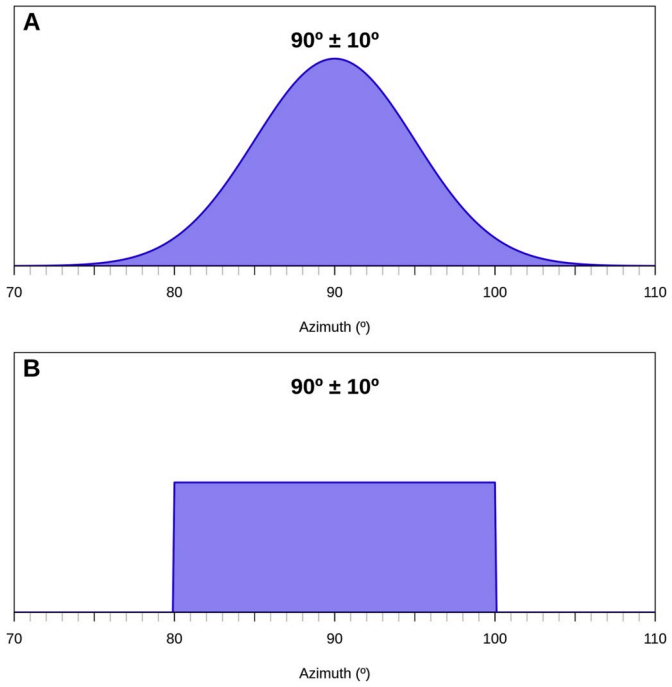


**Fig. 1.** Two different probabilistic models for the same measurement (90° ± 10°). The top one is modelled using a normal distribution with standard deviation equal to half the uncertainty, meaning that one has 95.4% confidence that the real value is within the uncertainty range and that the likelihood peaks around the mean. The bottom model uses a uniform distribution with range given by the uncertainty range, meaning there's 100% confidence that the real value is within the uncertainty estimate and that all orientations inside that range are equally likely to be close to the real value.

times the uncertainty away from the best estimate. An alternative model is to use the uniform distribution. Such a distribution attributes the same likelihood to all azimuths inside a given range, or window, and therefore can represent a less centre- or symmetry-biased assumption (Fig. 1B). This can be easily implemented by, for example, using the uncertainty range as the range of the uniform distribution. Note that, in this case, the probability distribution drops to zero outside of its range – its underlying assumption is therefore that one is 100% confident that the true value of orientation lies somewhere within the range of the distribution. These two distributions, just like any other, have pros and cons and one must therefore carefully weigh them when choosing one to model orientation measurements.

## 3. Coordinate transformation

To explore potential celestial alignments across a number of different sites, one needs to work with quantities that are not tied to a specific location. This is necessary because a given celestial object does not rise nor set at the exact same azimuth when seen from different locations. Two factors affect this: the latitude of the place, which changes the angle between the celestial object's path and the horizontal plane, and the horizon itself, which can block celestial objects from view until they reach a certain altitude, for example in the presence of a mountain (Fig. S1). Hence, structures at different locations can have different orientations but still be intentionally aligned with the rise or set of the same celestial object. For this reason, when analysing the orientations of structures at different locations, one needs to convert one's measurements to a coordinate system that references the location on the celestial sphere that the structures are pointing towards, regardless of horizon or location.

The equatorial, "fixed address" or "Chinese" coordinate system is the standard in astronomy and corresponds to a simple angular system on the projected celestial sphere (Kelly and Milone, 2005: 16–20). In this system, there is a *declination* that is measured along a line connecting the celestial North and South poles (a celestial meridian) and is analogous to latitude on the Earth's surface; and a *right ascension* that is measured along lines parallel to the celestial equator and is therefore analogous to longitude. Because the celestial sphere is perceived to be rotating throughout the day and night, right ascension relates to the time at which a given celestial object will be above the horizon. Declination, on the other hand, is the key quantity to understand where the celestial object will touch the horizon and what trajectory it will follow. It is therefore this quantity that is regularly used in archaeoastronomical studies. Unfortunately, it cannot be directly measured in the field, rather it needs to be calculated from field measurements with recourse to spherical trigonometry. The equation to do this is:

$$\sin \delta = \sin h \times \sin L + \cos h \times \cos L \times \cos A, \quad (1)$$

where $\delta$ is the declination, $A$ is the azimuth, $h$ the horizon altitude and $L$ the latitude of the structure. To this calculation some corrections can be added, such as those due to atmospheric refraction, nutation and parallax (e.g. Kelly and Milone, 2005: 49–70). This equation reveals that the declination is an angular measurement that goes from $-90°$ to $+90°$ and that for a flat horizon with 0° altitude, declination 0° corresponds to azimuths of 90° and 270°, i.e. East and West respectively. Hence, a curve of equal declination on the celestial sphere will, typically, touch the horizon twice: once in the eastern half and again in the western half, corresponding to the rising and setting positions of a celestial object with that declination.

The typical approach within archaeoastronomy is to take the measured azimuth, horizon altitude and latitude and plug them through equation (1) to get a single declination value (e.g. Ruggles, 1999, 22) – an approach that has remained largely unchanged for at least a century (see Lockyer, 1909). An uncertainty around the calculated value of declination, when presented, is often obtained by plugging the uncertainty in azimuth through equation (1). These values are then used to identify one (or more) celestial object(s) with the same declination and that, therefore, would rise or set in alignment with the structure being studied. When in the presence of multiple similar sites, histograms, curvigrams or kernel density estimates can be produced and looked at in search of frequency peaks that highlight patterns in orientation (e.g. Ruggles, 2015). Because this approach is based entirely on singular, or *discrete* applications of the coordinate transformation equation, I will refer to this as the *discrete approach*.

This discrete approach ignores a number of important issues. First and foremost, it can lead new scholars astray by giving off the impression that the uncertainty in declination is the same regardless of azimuth. This is not supported by equation (1), which exhibits a trigonometric relationship between declination and azimuth, meaning that the declination varies differently, for example, around azimuths of 0°/360° (North) and 180° (South). Furthermore, as uncertainty is not fundamentally built into this approach, scholars neglect or underestimate it, which has severe consequences for the accuracy of the methods of the discrete approach (Silva, 2017). Finally, as pointed out by Ruggles (2015b: 418), it fails to note that within the range of azimuths given by the measurement uncertainty, the altitude of the horizon can vary quite dramatically, which leads to significant differences in how azimuths relate to declinations (Fig. 2). Together these observations indicate that a declination is not necessarily normally distributed and, hence, cannot be reduced to a mean and uncertainty values derived from equation (1), which is the underlying assumption of the discrete approach.

Therefore, what is required is a method that performs the coordinate transformation (1) on the full probability distribution of the field measurements (Fig. 1), for which it will need not a single value of horizon altitude but a horizon profile that, at the very least, covers the azimuthal range of interest. Such a method would in many ways be analogous to the process of calibration of radiocarbon dates, wherein there is a

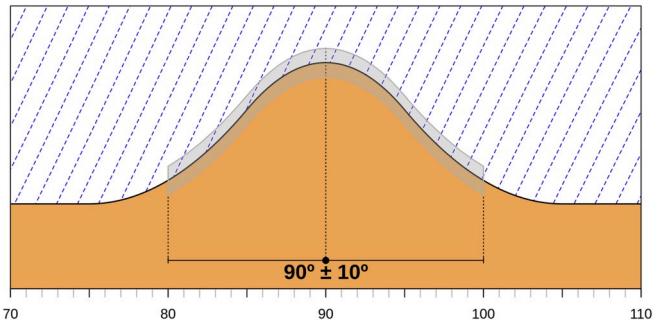


**Fig. 2.** Example of a situation where the range of declinations covered by the horizon band changes dramatically within the uncertainty range of an azimuthal measurement. Orange shaded area corresponds to the ground; dashed blue lines correspond to lines of equal declination separated by one degree. The uncertainty in measurement of orientation (black error bar) and horizon altitude (grey shaded band) are also represented. In this case, the region of the horizon to the left of East spans a mere 4.9° of declination, whereas the region to the right spans 10.3°. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

continuous calibration curve that transforms the probability distributions of the uncalibrated dates (which are normally distributed) into calibrated probability distributions (which are not necessarily bell-shaped).

### 3.1. The transformation process

Whereas for radiocarbon a single atmospheric calibration curve applies to all sites in half of the planet, in the case of structural orientations there must necessarily be one *transformation curve* for each site (sometimes even for each measurement) since equation (1) is tied to the apparent horizon from a particular viewpoint. This requires that a horizon profile be obtained during fieldwork. If this is impossible, one can be virtually constructed with recourse to digital elevation models. The former is however, preferred, as the most readily available digital elevation models are likely to contain spatially variable errors in vertical measurement (e.g. Mukul et al., 2017), which will yield uncontrolled-for errors in horizon profiles and altitude values.

When a horizon profile is obtained it is a simple matter to turn it into a transformation curve by applying equation (1) across the entire 360° panorama at regular intervals (top-right panel in Fig. 3). Since all measurements, however they were obtained, have an associated uncertainty one must take the uncertainty in horizon altitude into account. This can be done by repeating the process two more times for altitude values $h + \delta h$ and $h - \delta h$, where $\delta h$ is the altitude uncertainty (dark blue shaded curves in top-right panel of Fig. 3). Without loss of generality, it is here assumed that the horizon altitude measurement can be represented by a normal distribution with standard deviation equal to the uncertainty of that measurement.

Having, in this way, transformed the horizon profile into what resembles a calibration curve, the process of *coordinate transformation* proceeds just like a radiocarbon calibration (e.g. Bronk Ramsey, 2001). A key difference is that whereas in radiocarbon calibration the distribution that is being calibrated is almost always normally distributed, for present purposes it is desirable to leave the choice of probability distribution open (as discussed above). This means that the algebraic simplification that ensues when convolving two normal distributions, which is implemented in all radiocarbon calibration software packages (Bronk Ramsey, 2001), will not work here. Instead, the density values of the coordinate-transformed probability distribution, at each value of declination, are calculated by numerically convolving the azimuthal distribution with the distribution of said declination that stems out of the transformation curve (see R code in GitHub repositorysupplementary material). Because of its numerical nature this is a slower process than radiocarbon calibration – an unfortunate necessity. This process results in what should be an intuitive output to archaeologists (Fig. 3).

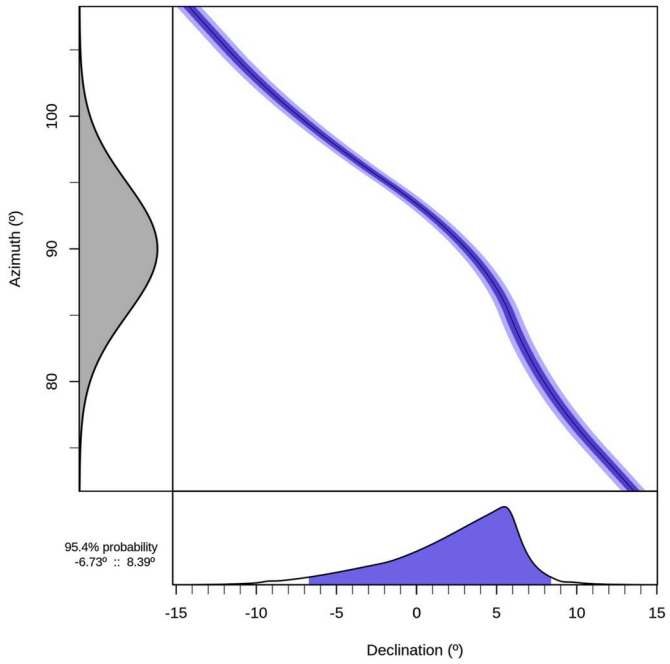


**Fig. 3.** Result of the coordinate-transformation of an azimuthal measurement of 90° ± 10°, modelled as a normal distribution with standard deviation of 5° (grey-shaded curve on the left). The horizon profile from Fig. 2 has been converted into a transformation curve (blue shaded lines on top-right). Finally, the result of the coordinate-transformation is a declination probability distribution (blue shaded curve on the bottom) with 95.4% probability range shown on the bottom-left. Notice how the mound on the horizon considerably distorts the shape of the declination distribution. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

It is now relevant to better understand the impact of this coordinate-transformation and how it affects the shape of probability distributions of individual orientation measurements, especially when compared to the naïve approach of assuming declination distributions to be normal curves. A more thorough exploration of the parameter space is done in the Supplementary Material, however, Fig. 4 below summarises the key results. The first observation is that, due to the changing density of declination isolines per unit of azimuth, the declination distribution changes depending on the orientation, leading to considerably narrower and slightly skewed distributions the closer one is to North or South (Fig. 4A and B). The second observation is that the shape of the horizon does matter quite significantly: if an orientation is targeting a downwards slope, for example (Fig. 4C), then the distribution is wider and shorter due to the fact that more declination isolines are encompassed by the same azimuth range than in the flat scenario. The opposite occurs for an upward slope. For combinations of the two, such as for peaks (Fig. 4D) or notches on the horizon the distribution is compressed on the upwards sloping side and extended on the downwards sloping side, creating a highly non-normal distribution. For more details, see the Supplementary Material, where this analysis is also done for uniform azimuthal distributions.

### 3.2. Aggregating coordinate-transformed probability densities

Coordinate-transformed distributions, regardless of their particular shape, can be aggregated into what would be appropriately called a *Sum of Probability Densities*, to use the same terminology that is in use for

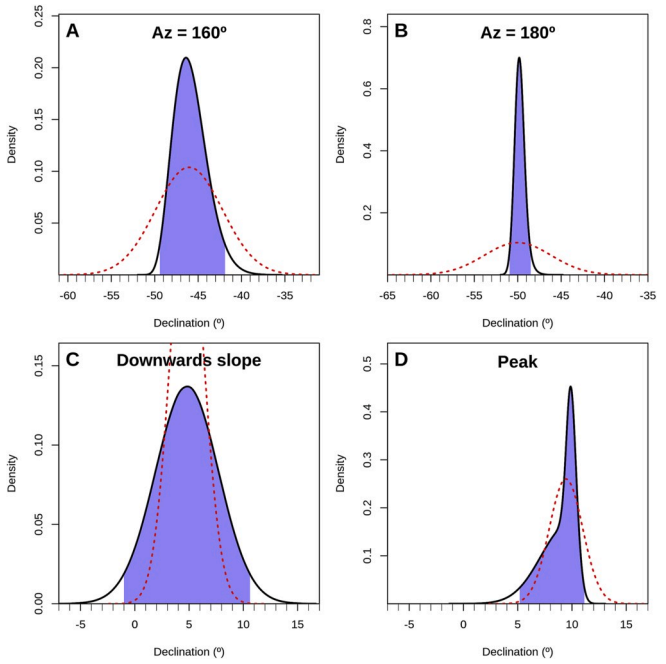


**Fig. 4.** Coordinate-transformation of normal measurement models of azimuths 160° (A) and 180° (B), with equal uncertainty (5°) and a flat 0° horizon scenario at latitude 40° N; and of azimuth 90° but with a downwards slope (C) and a peak (D) on the horizon exactly at that value. The red dotted curves represent the expectation under the discrete approach. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

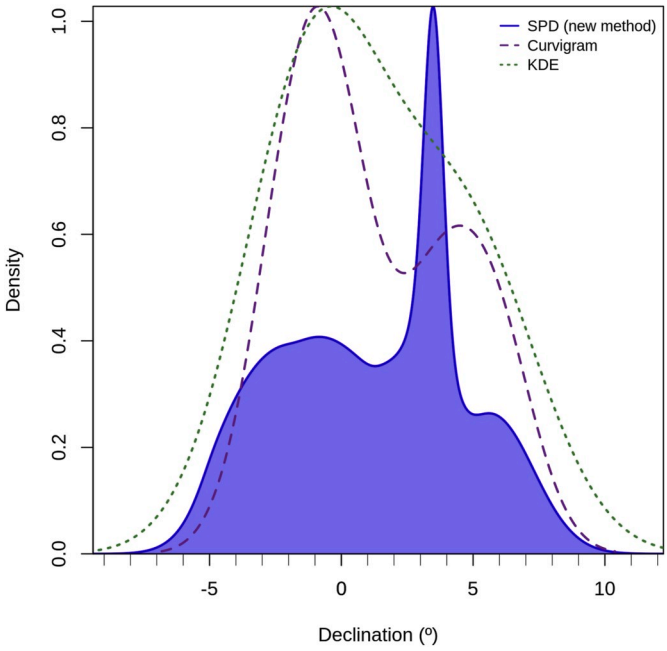


**Fig. 5.** Example comparison of the outputs of the SPD approach that aggregates coordinate-transformed probability distributions developed in this paper (blue-shaded curve) with the discrete approach curvigram (purple dashed curve) and a kernel density estimate (KDE) using Scott's (1992) bandwidth (green dotted curve). For ease of comparison the last two curves were renormalised to match the peak of the SPD. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

radiocarbon analyses (e.g. Bevan et al., 2007). For clarity, and to better distinguish this method from that which assumes all declination distributions to be normally distributed, we will always refer to the latter as the *curvigram*, and to this new method as the *Sum of Probability Densities* or *SPD* for short (contra Silva, 2017 where the two were conflated). In fact, the curvigram is also an aggregated probability distribution, but one where the azimuthal measurements were *not* coordinate-transformed in the way introduced above, but rather modelled as normal curves (i.e. as the red curves in Fig. 4).

Fig. 5 illustrates the differences between the SPD and the curvigram outputs for a set of five random orientations around a 4° peak on the horizon. The azimuthal measurements were modelled using normal distributions with 2° of uncertainty. In this not-unrealistic scenario, the curvigram and SPD present qualitatively and quantitatively very different shapes. Whereas the curvigram features a bimodal distribution around −1° and 5°, the SPD actually downplays the lower declinations and features a narrower peak around 3.5°. This occurs because of the effects discussed above for Fig. 4, where coordinate-transformed declination distributions present a peak towards positive declinations since they are taking into account the varying range of declinations along the horizon (as seen in Fig. 2). Also shown is the result of a kernel density estimate using Scott's (1992) bandwidth, which produces a much smoother curve that peaks close to the SPD peak – and therefore equally misses the mark.

Using the traditional archaeoastronomical approach mentioned in the introduction, one would now look at the declinations of peaks of the curvigram in search of a matching celestial object that could be the target of an alignment. The curvigram and KDE peak around −1°, whereas the SPD peaks at 3.5° so, depending on the accuracy one is imposing on this, scholars using the two approaches would be looking for different targets. Nevertheless, as archaeologists familiar with the use of SPDs in radiocarbon modelling know very well, peaks, troughs and wiggles in aggregated probability distributions can be due to sampling issues (for example if a particular orientation is under-represented in the sample due to sampling bias) or even due to the transformation process (c.f. calibration wiggles in a radiocarbon-derived SPD). Without a significance test one is likely to be over-interpreting the results by reading too much into those peaks.

## 4. Testing significance

### 4.1. Null hypothesis construction

To test for significance, one needs to start off with a null hypothesis that the data is tested against. The most basic one for the study of structural orientations is that the structures being analysed were orientated at random. This translates into a random distribution of azimuths that can be modelled with a uniform distribution, where each and every azimuth value has the same probability of having been chosen (Fig. 6A). The corresponding declination distribution to this null hypothesis, however, is not uniform as can be observed by coordinate-transforming it (Fig. 6B). The distribution peaks at the extremes for the same reasons discussed above for the coordinate-transformation of azimuths close to North and South: the declination distributions of individual measurements get narrower and taller (Fig. 4A and B). This means that, under the null hypothesis of random orientation, we would expect to see more values of declination closer to the extremes than closer to zero.

Such a shape corresponds to a very specific scenario where there is a large number of structural orientations, all of which occur at the same latitude and hence cover the same span of declinations. In a realistic research project however, one is confronted with a finite, and often rather small, sample of orientations, each taken from a different structure on a different location. As such, one needs to consider not the limiting distribution of Fig. 6B but the range of realistic possibilities that, nevertheless, still conform with the null hypothesis. This can be achieved through a brute-force computation method inspired by the approaches to significance testing of radicoarbon SPDs by Shennan et al.

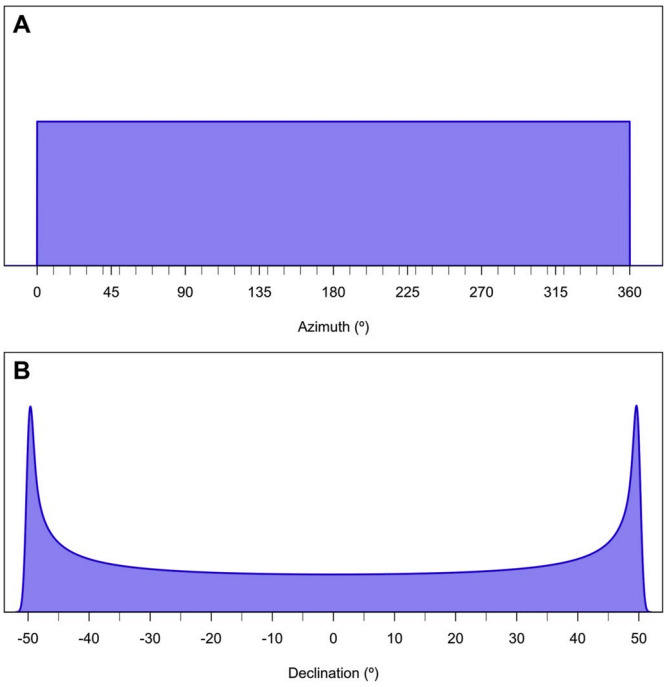


**Fig. 6.** Two models for the null hypothesis of random orientation: (A) shows the null hypothesis azimuthal distribution where each azimuth value is equally likely to be picked; whereas (B) shows the equivalent coordinate-transformed declination distribution which displays peaks around the extremes (corresponding to due North and due South) as being more likely to occur when one is looking at declinations.

(2013), Timpson et al. (2014) and Crema et al. (2017). The goal is to use the information in the empirical dataset – in this case the sample size, uncertainty profile and spatial locations – to create a confidence envelope around the null hypothesis that represents the range of possibilities of the null hypothesis. It is from this envelope that the statistical significance can then be quantified in the form of a p-value.

To explain the process, consider the set of measurements of the orientation of five similar structures that were used above to produce Fig. 5. The goal is to compare this empirical SPD with the null model in order to assess whether or not it constitutes evidence against the null hypothesis of random orientation. One should therefore ask the question of what would the SPD of a randomly orientated set of five structures, sited at the same locations and measured with the same uncertainty, look like. This can be answered as follows: (1) pick five random azimuths from the uniform distribution of Fig. 6A; (2) randomly attribute to them the same measurement uncertainties, locations and horizon profiles of the empirical set; (3) coordinate-transform them; and (4) aggregate them to produce a simulated SPD (grey curve in Fig. 7B). Comparing the empirical SPD with the simulated one isn't particularly informative since there is a very large number of different possibilities contained within the null hypothesis, especially for such small sample sizes – as is clear when this process is repeated a few more times (Fig. 7C). A much more helpful approach is to repeat this a great number of times so as to get a sense for the 95% confidence envelope around the null hypothesis (Fig. 7D and E). This is to say that we can be confident that a dataset that conforms to the null hypothesis – i.e. a dataset that is randomly orientated – will produce an SPD that, 95% of the time, falls inside of the confidence envelope.

From this, one can calculate a value that quantifies the level of statistical significance (Fig. 7E). This is often referred to as a *p-value* and is defined as the probability of observing, under the null hypothesis, the empirical dataset or an even more extreme dataset. This can be calculated by comparing the empirical SPD with the simulated ones. Following Shennan et al. (2003: 6), we calculate a statistic defined by the total area of the empirical SPD that lies outside of the confidence envelope. The same statistic is also calculated for each of the simulated SPDs originated by the random picks from the null hypothesis. Finally, the p-value is calculated as (following North et al., 2002):

$$p = 1 - \frac{r+1}{n+1}, \tag{2}$$

where $r$ is the number of simulated SPDs with the statistic greater than, or equal to, that obtained from the empirical SPD, and $n$ is the total number of simulated SPDs. This amounts to saying that the probability of observing an SPD like that of the empirical dataset under the null hypothesis is given by the probability of finding a similar SPD among the simulated ones. We are therefore implementing a *global p-value* along the lines suggested by Shennan et al. (2013) as a measure of global significance, i.e. a measure of the statistical significance of the empirical dataset as a whole.

Furthermore, the regions of the empirical SPD that lie outside of the confidence envelope – the very regions whose area is used to calculate the global p-value – can be highlighted. These *regions of significance* comprise those declination ranges that are deemed statistically significant and they are precisely the regions which should be of interpretive interest. As one is working with a 95% confidence envelope, one should still expect 5% of false positives to lie outside of the envelope. To try to separate the wheat from the chaff, as it were, different approaches have been taken. Timpson et al. (2014: 551) introduced a heuristic *false positive remover,* whereas Crema et al. (2017) used a *false-discovery rate* to address the high risk of incorrectly rejecting a true null hypothesis. A third approach, albeit equally heuristic and arguably more useful for the present concern, is here introduced. It consists in the calculation of a local measure of significance – the *local p-value* – which takes only the region of significance into account for its calculation. Essentially, the algorithm loops through the identified regions of significance and repeats the p-value calculation using only the identified declination range to calculate the area statistic for the empirical as well as the simulated SPDs. The local p-value is then also calculated using equation (2). This results in a measure of significance that correctly attributes more significance (lower p-values) to regions that go well beyond the confidence envelope, whereas lower peaks or troughs, or those in regions where high deviations are more likely (such as near northern and southern azimuths or near peaks and notches on the horizon), are attributed lower significance (higher p-values). Local significance can help identify false positives, which will have larger local p-values compared to true positives (see Fig. 7E).

## 5. Applications

The statistical framework just introduced can now be used to look at real-world studies of structural orientations. In particular, this framework will prove useful in not only explicitly comparing measurements to the expectation under the null hypothesis but also in quantifying significance (the global p-value) and identifying ranges of statistically significant orientations that are potentially intentional. In the process, these applications will also highlight flawed readings of the data (such as the over-interpretation of frequency peaks) that disappear when this more robust methodology is employed.

### 5.1. Ancient Egyptian tombs and temples

The tombs and temples of ancient Egypt were among the first structures to be studied for their orientation using a surveying methodology that has remained largely unchanged to this day. In what is largely considered to be a key turning point in the history of the field of archaeoastronomy, Lockyer (1894) set out a more rational and scientific approach to the measurement of structural orientations, which he promptly applied to the study of Egyptian temples. Since Lockyer's

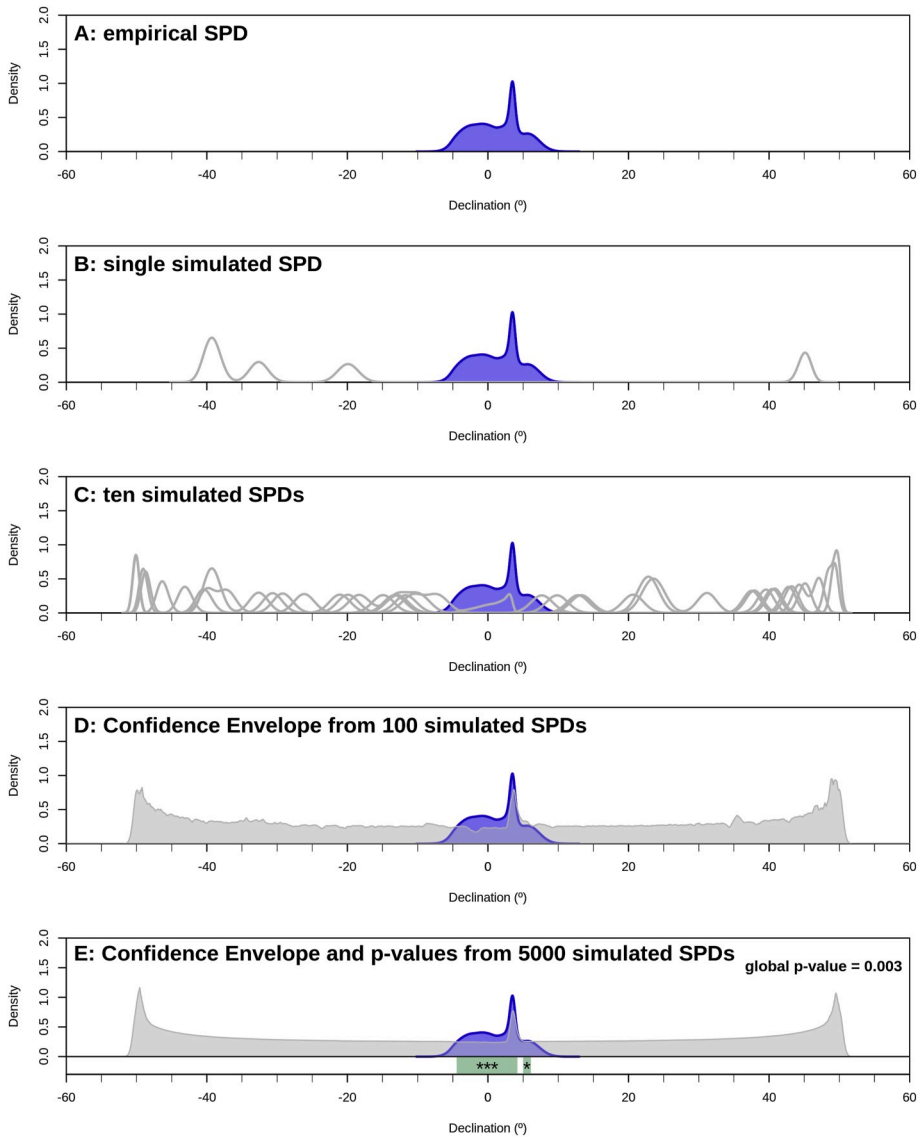


**Fig. 7.** The process of constructing a confidence envelope for the null hypothesis of random orientation by randomly sampling from a uniform azimuthal distribution and randomly attributing the same uncertainties and horizon profiles from the empirical dataset. When the process is repeated a few hundred times or more a 95% confidence envelope can be obtained (grey band in panels D and E) as well as a global p-value (panel E). Regions of significant positive deviations from the null hypothesis are marked by the green bands at the bottom of panel E, with stars denoting their level of significance. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

impressive survey of dozens of ancient Egyptian structures, no large-scale study was attempted until 2004 (Shaltout and Belmonte, 2005). The so-called Egyptian-Spanish Mission had for its aim the measurement of the orientation of the vast majority of ancient Egyptian buildings so as to obtain 'sufficient fieldwork data that could prove, or disprove, through statistical studies, all the speculations concerning temple orientation' (Belmonte et al., 2009: 218). Over its five-year period the mission surveyed over 300 pyramids, hypogea, chapels, sanctuaries and temples (Shaltout and Belmonte, 2005; Belmonte and Shaltout, 2006; Shaltout et al., 2007; Belmonte et al., 2008; Belmonte and Shaltout, 2009; Belmonte et al., 2010). After identifying the direction of the Nile as a key concern in the orientation of the temples, Shaltout, Belmonte and collaborators explored the possibility of celestial alignments (Shaltout and Belmonte, 2005; Belmonte et al 2009, 2010). They employed an analytical framework which largely relied on the creation of curvigrams and the identification of celestial objects that match their peaks. Using this, they identified seven families of orientation by looking at the peaks whose frequency was higher than a relative normalized frequency (see Figs. S8 and 8). These were given numbers and interpretations: equinoctial (I), solstitial (II), seasonal (III), Sirius (IV), Canopus (V), meridional (VI) and intercardinal (VII), largely by attaching cardinal and celestial interpretations to the peak orientations (see Belmonte et al., 2009 for details).

To explore whether all of these families of orientations are indeed statistically significant, and hence worthy of further consideration, the probabilistic framework developed here has been employed. The published dataset of 330 measurements of the Egyptian-Spanish Mission (Belmonte and Shaltout, 2009: 347–352) was used, and each azimuthal measurement was modelled using a normal distribution with 2° standard deviation. This uncertainty is slightly larger than the 1.5° bandpass used by Shaltout and Belmonte (2005: 281) in their methodology, which is itself mysteriously larger than the claimed 0.5° precision (2005: 279), but yields results that are closer to those published by the authors (see Fig. S8). Due to the lack of precision in the published georeferences, accurate locations for all structures could not be obtained, which prevented horizon profiles from being reconstructed from a Digital Elevation Model. Therefore, flat horizons with an altitude equal to that measured by the Spanish-Egyptian team, and with an associated uncertainty of 0.5°, were assumed throughout. The SPD constructed in this way is not expected to vary much from the traditional curvigram since the effects of horizon shape (i.e. the ones depicted in Fig. 4C and D) are effectively being ignored. Any marked differences are therefore likely to stem from typos in the published data and from the azimuthal effects described in Fig. 4A and B. Such differences are therefore deemed conservative and would be expected to be magnified by future studies that use full horizon profiles. Such studies are beyond the remit of this

methods-driven paper, but the present analysis is still relevant as it will show the importance of performing the significance test even in the absence of full horizon profiles. After coordinate transforming the orientation measurements and summing them, the resulting SPD can be subjected to the significance test, giving rise to Fig. 8 and Table 1.

Firstly, the global p-value is rather small ($p < 0.0002$) attesting to the overall statistical significance of this dataset which can therefore be said to exclude the null hypothesis of random orientation. This, however, doesn't mean that every structure in the dataset does not conform to the null hypothesis. Rather, what can be seen are peaks that step out of the confidence envelope (grey shaded), indicating regions of significant deviation from the null hypothesis. The regions of significant positive deviation (i.e. higher than the confidence envelope) directly correlate to orientation families I, II and III- (see Table 1). The local p-values reinforce what Fig. 8 qualitatively states: that peak III- is considerably less significant than the others and may potentially be a false positive — indeed, if one were to use the more strict significance thresholds suggested by Schaefer (2006) or Benjamin et al. (2017) then this peak would not be deemed significant at all. In other words, only two (or possibly three) of the families inferred by Belmonte et al. (2009), are validated by this more robust method – all other peaks fall within the confidence envelope and therefore conform to the expectation of random orientation.

What then does the new method permit one to say about the orientation of Ancient Egyptian tombs and temples? Despite the existence of statistically significant peaks and troughs, when looking at the entire dataset, most surveyed structures appear to follow a pattern of random orientation. The only structures that can be said to be statistically aligned to some target are those that fall in the two or three regions of significance. The most statistically significant corresponds to an easterly orientation (family I), which may have been a simple cardinal orientation to East, an alignment to the Nile which generally runs south-north or, less likely, an alignment to the equinoctial sun (see discussion in Belmonte et al., 2009: 227-8). This is followed by a second peak (family II) which very tightly matches the orientation of the December solstice sunrise (Belmonte et al., 2009: 229-30). The potential targets of family III require more complex arguments, involving the orientation of the sun around the beginning of the Ancient Egyptian seasons of *Peret* and *Shomu* around 1500 BC (Belmonte et al., 2009: 231-32). As previously mentioned however, this last family of orientations may very well be a false positive and actually not be statistically significant.

These two or three families can therefore be said to represent the dominant orientations for ancient Egypt as a whole, possibly reflecting a long-term interest in these celestial objects and/or events, rather than more regionally or temporally localised intentions. Similarly to Belmonte et al. (2009: 244–249) one can split the dataset into meaningful chronological segments to assess whether these families played more prominent roles in different historical periods, or even whether other families become statistically significant when looked at in this way. The dataset was therefore split into four chronological periods of political, social and religious stability (Shaw, 2000); ignoring orientations belonging to temples or tombs of other, less stable, periods, as well as those which belonged to structures of unknown or dubious chronology. The results are shown in Fig. 9 and Table 2, where a chronological pattern becomes apparent.

**Table 1**
Regions of significance identified by the method developed in this paper, including Belmonte et al.'s family number, declination range, local p-value and potential targets discussed in Belmonte et al. (2009).

| Family | Declination Range | | Local p-value | Potential targets |
|---|---|---|---|---|
| | Min | Max | | |
| I | −4.7° | +2.7° | <0.0002 (***) | East, Nile river, Equinox sun |
| II | −26.2° | −21.1° | <0.0002 (***) | December solstice sun |
| III- | −11.9° | −9.5° | 0.0138 (*) | Sunrise at beginning of Egyptian seasons of *Peret* and *Shomu c.* 1500 BC |

Old Kingdom (2686 - 2160 BC, Shaw 2000) structures were by far and large built with an easterly orientation - potentially facing the Nile river although this should be assessed on a case-by-case basis - whereas no other orientation family was statistically significant in this period. Family I continued to be of importance in the Middle Kingdom (2055 - 1650 BC, Shaw 2000), when the family of orientations to the December solstice sun (II), became significant. In the New Kingdom (1550 - 1069 BC, Shaw 2000), family I disappeared completely, whereas family II became the most statistically significant one, including also related orientations marked as II⊥ to represent directions perpendicular to the December solstice sun. This may indicate the special importance attributed to the sun-god in the New Kingdom as seen, for example, in the Books of the Netherworld that feature exclusively within the royal tombs of this period (Assmann, 2008: 186–189). Family VII (both + and -) appeared for the first time in the New Kingdom, being entirely absent in previous periods, and is deemed significant by the employed statistical method. Belmonte et al. (2009: 238–244) suggest this family involves a 45° shift from the cardinal directions, which they suggest was a clever solution to ensure 'simultaneous astronomical and Nile orientations' (238). Finally, in the Ptolemaic period (332 - 30 BC, Shaw 2000) only family VII- juts above the confidence envelope of the null hypothesis, and indeed the global p-value does not indicate statistical significance in this subset. It is also worth highlighting that family III- is not statistically significant in any of these periods, adding further weight to the above interpretation of it being a false positive.

This case study illustrates an important issue with archaeoastronomical approaches that rely solely on the construction of a curvigram and the analysis of its peaks: without the significance test one is prone to over-interpret the many peaks, most of which, as the above analyses demonstrate, are not statistically significant. Nevertheless, this dataset deserves a lot more attention. On the one hand, the chronological dynamics revealed above needs to be socio-historically contextualised to be fully understood. Further subsetting of the data, for example into regional groupings and site-types similarly to Belmonte et al. (2009: 247), may shed further light. On the other hand, and perhaps more importantly, future analyses should also incorporate full horizon profiles, as well as take into account variations in both azimuthal and altitudinal uncertainty, which should improve on these results considerably.

### 5.2. Scottish Recumbent Stone Circles

The second case-study focuses on the Recumbent Stone Circles of Aberdeenshire, most recently surveyed for potential celestial alignments by Clive Ruggles (1984, 1999). The circles are believed to have been built during the period 2500-1750 BC (e.g. Ruggles and Burl, 1985: S25) and have been at the centre of archaeoastronomical debate for a long time (see the review by Henty, 2015). Following Thom and Thom (1980) and Ruggles (1984), their orientations have been studied largely by focusing on the orientation towards the midpoint of the recumbent stone as seen from a putative centre of the circle or, alternatively, along a line perpendicular to the recumbent itself (Ruggles, 1999: 96). Having surveyed all extant monuments in this way, and creating histograms of their orientations, Ruggles concluded that they were likely oriented to observe the full moon passing low over the recumbent around midsummer each year (Ruggles, 1999, 98). This view is at least partially endorsed by Richard Bradley, who led the excavation of three of these monuments and largely agreed with Ruggles' conclusion when he stated that they 'may have been directed towards the moon and in some cases they may also have faced the winter sun' (Bradley, 2005: 111). On the other hand, the Royal Commission on the Ancient and Historical Monuments of Scotland recently produced a major synthesis work on these megalithic monuments, which noted that such celestial explanations do

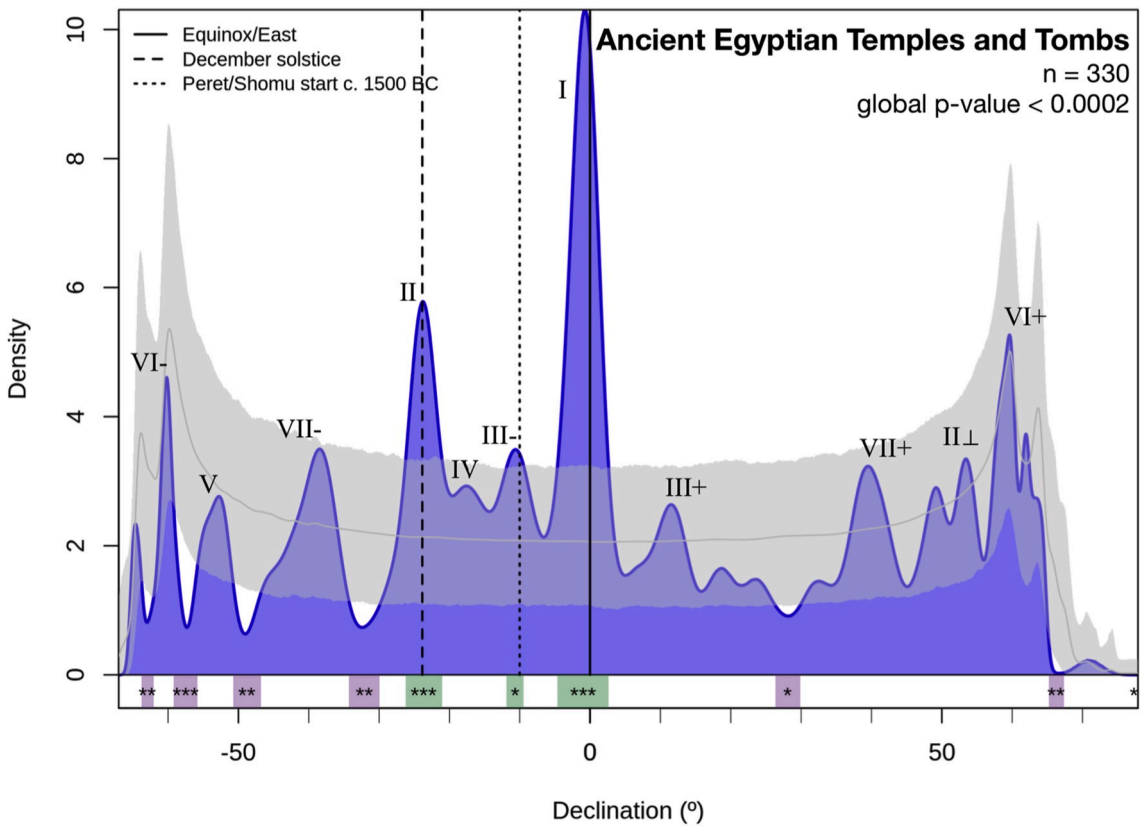


**Fig. 8.** Result of the significance test applied to the Ancient Egyptian temple and tomb orientation dataset, running 5000 simulated SPDs. Roman numerals indicate Belmonte et al's (2009) orientation families. Also shown as vertical black lines are the declinations of the sun at the December solstice (dashed), at the start of the ancient Egyptians seasons of *Peret* and *Shomu* around 1500 BC (dotted), and at the equinoxes (solid line). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

not apply to all the stone circles, leading to the conclusion that 'archaeoastronomy has failed to provide a convincing and coherent theoretical model […] with respect to their orientation' (Welfare, 2011a: 28).

One of the great things about high-quality open data is that it can be endlessly analysed by alternative and, at the time of publication, yet-to-be-devised methodologies. Ruggles' data (1999: 212-6) can now be approached using the probabilistic framework developed in this paper. This dataset contains orientation measurements for 37 stone circles, taken from the circles' putative centre towards the midpoint of each recumbent. The most recent survey of recumbent stone circles (Welfare, 2011b) was used to obtain accurate georeferences for each of these circles. Horizon profiles were obtained from the SRTM digital elevation model through the HeyWhatsThat website (Kosowsky, 2019). The orientation measurements were then modelled in three different ways (Fig. 10):

A) as normal distributions centred around the middle of the recumbent and with 0.5° of uncertainty;
B) as normal distributions centred around the middle of the recumbent and with an uncertainty given by the width of the recumbent arrangement;
C) as uniform distributions whose width is given by the recumbent arrangement.

Model A is closer to the traditional approach of archaeoastronomers, except that the coordinate transformation will be applied across the entire probability distribution. The second model (B) is an attempt to take into account the fact that the entire recumbent arrangement could act as a window to any alignment targets, whether topographic or celestial, a point more poignantly made by Henty (2014). This model, however, still assumes that the middle of the recumbent is more likely to be the backsight for any alignment – i.e the probability density is highest at this location. Model C ditches this assumption and rather suggests that any location within the recumbent arrangement is equally likely, an approach closer to Silva's research among the Portuguese passage graves (Silva, 2014, 2019). These models, depicted in Fig. 10, form the basis for the ensuing analysis.

Each measurement was then coordinate-transformed and summed, leading to the SPDs of Fig. 11. Model A produces several peaks and wiggles, which is to be expected from the very narrow uncertainty margin. Model B, on the other hand, results in a smoother SPD, especially with respect to the peaks around −25° and −17°. The model that employs a uniform probability distribution (model C) yields a SPD that features a single distinctive peak towards the southernmost extreme of the declination range and a long tail in the other direction. This shows that, when the full width of the recumbent is taken into account those minor peaks are dissolved. This is not surprising when more attention is given to the fact that the recumbent arrangements are quite wide: the mean angular width being 21.7° and the maximum being 35.5° (for Strichen circle). In itself, this suggests that the 0.5° uncertainty of model A is likely to be underestimating the actual uncertainty in these measurements (cf. Silva, 2019b for a general discussion of the dangers of uncertainty underestimation). As such, any interpretative attempts to identify alignment targets from those peaks in model A are likely to be a wild goose chase.

To explicitly test whether or not that is the case, the next step is to perform the significance test and obtain the declination ranges that present significant deviations from the null hypothesis of random orientation. The results are shown in Fig. 12 and Table 3.

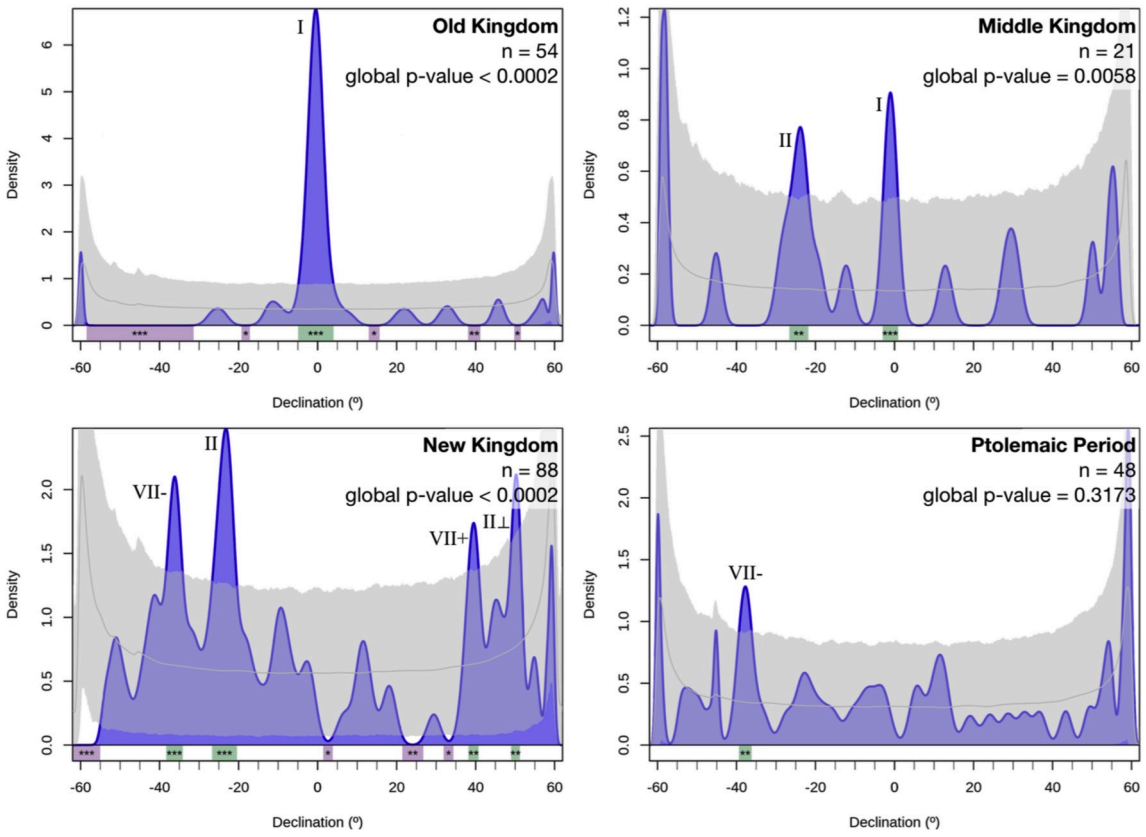


**Fig. 9.** Result of the significance test applied to the Old Kingdom, Middle Kingdom, New Kingdom and Ptolemaic Periods subsets, running 5000 simulated SPDs. The numbers of the orientation families deemed statistically significant in each period are shown.

**Table 2**
Regions of significance identified by the method developed in this paper, when applied to chronological subsets of Belmonte et al. (2009) dataset.

| Period | Family | Declination Range | | Local p-value | Potential targets |
|---|---|---|---|---|---|
| | | Min | Max | | |
| **Old Kingdom** | I | −4.9° | +4.0° | <0.0002 (***) | East, Nile river, Equinox sun |
| **Middle Kingdom** | I | −3.0° | +0.9° | 0.0006 (***) | East, Nile river, Equinox sun |
| | II | −26.6° | −21.8° | 0.0032 (**) | December solstice sun |
| **New Kingdom** | II | −26.7° | −20.4° | <0.0002 (***) | December solstice sun |
| | VII- | −38.3° | −34.2° | 0.0004 (***) | Intercardinal direction |
| | VII+ | +38.2° | +40.8° | 0.0054 (**) | |
| | II⊥ | +49.0° | +51.2° | 0.0006 (**) | Perpendicular to II (December solstice sun) |
| **Ptolemaic Period** | VII- | −39.4° | −36.2° | 0.003 (**) | Intercardinal direction |

The global p-values indicate that indeed the orientations of the Recumbent Stone Circles significantly deviate from random, with a clear clustering around the southernmost orientations (i.e. lowest possible declinations). But whereas model A has three peaks that are significantly above the confidence envelope of the null hypothesis, model C features only two and model B merely one. Those secondary peaks are very close to the confidence envelope, which can also be seen by their large local p-values shown in Table 3, which are considerably higher than those of the main peak. In fact, if one were to use Schaefer (2006) or Benjamin et al. (2017) tighter significance thresholds then none of those secondary peaks would be deemed significant.

This case study demonstrates the importance of carefully choosing a model for the measurements by illustrating the impact it has on interpretation. The azimuthal probability distribution represents one's *a priori* assumptions about the measurement's relationship to potential alignments and this choice affects the shape of the aggregated distribution – whether it be the SPD as defined here or the curvigram. Moreover, this case study also demonstrates how the significance test can attenuate this problem by highlighting the lack of significance of those secondary peaks.

Finally, this brief study of the Recumbent Stone Circles cannot close without looking at potential alignment targets. Fig. 13 overlays the declination of previously considered celestial objects (for the period 2500-1750 cal BC) on top of the results of the significance test. It shows that the preferred interpretations of Ruggles and Bradley, where the recumbent arrangement frames the setting summer full moon or winter sun, are insufficient to explain the orientation of these structures. The vertical black line marks the declination of the sun at its most southern extreme during the December solstice, and it is clear from the SPD that most of the recumbents lie too far south to catch the setting sun at any point in the year. In fact, this would be possible in only 8 out of the 37 stone circles. The orange vertical band marks the range of the summer full moon over its 18.6-year cycle, and although this band partly covers the region of significance it only does so for about a third of it. When the dataset is looked at more closely, it becomes clear that at 19 out of the 37 circles it would be impossible to observe the summer full moon set within the recumbent window.

This could be interpreted as evidence against any kind of celestial alignment – which indeed may be true – however, it is worth exploring any underlying assumptions in the dataset that may need to be reassessed or otherwise incorporated into the measurement distribution. In this particular case, three assumptions need to be highlighted. Firstly,

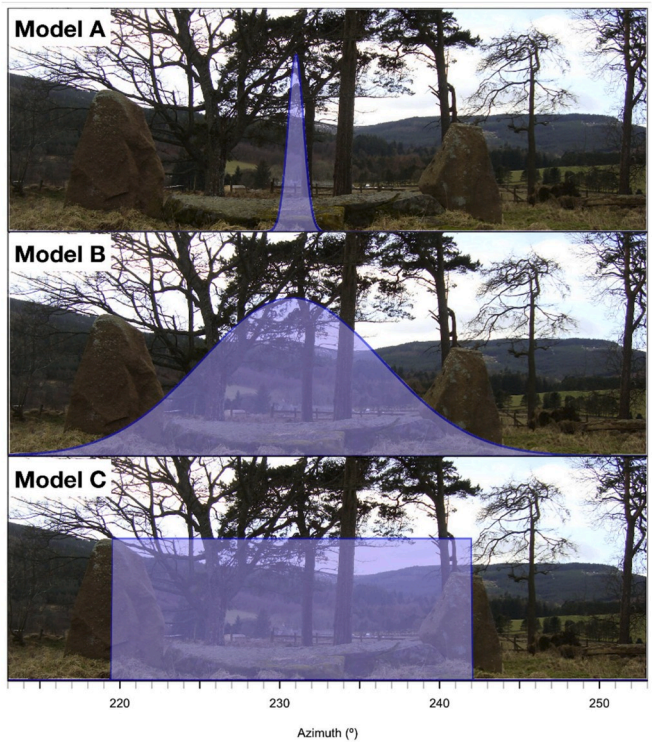


**Fig. 10.** Three different ways to model the orientation of the recumbent arrangement of a Recumbent Stone Circle: (A) as a normal distribution tightly concentrated around the centre of the recumbent; (B) as a broader normal distribution to encompass the entire recumbent arrangement; and (C) as a uniform distribution. The background image is for Sunhoney circle, courtesy of Liz Henty (with permission).

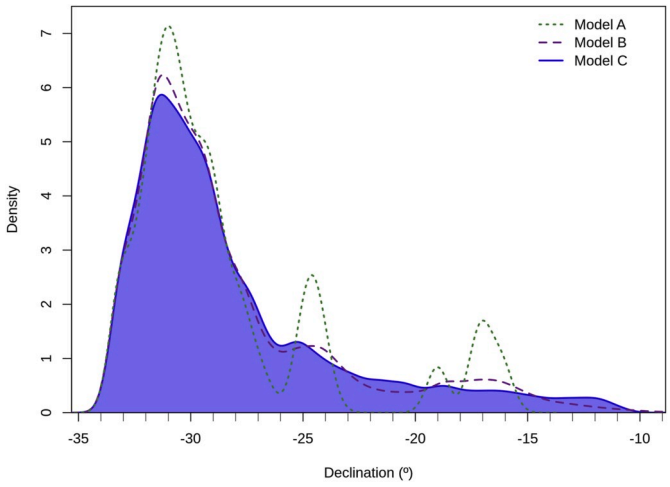


**Fig. 11.** SPDs for each of the three measurement models considered.

the assumption that these circles have a centre-point from which these measurements were taken might be untenable, especially considering that they are not perfect geometric circles nor are there any visible marks or finds at their geometric centres (Henty, 2012, 2014). Secondly, as Bradley (2005: 99) remarked and Henty explored (2014), it is possible that the circle's internal area was not entered at all, but instead that the space between two stones opposite the recumbent was used as viewpoint from which to look at the recumbent arrangement. If that was the case – or indeed if the framing device of the recumbent was meant to be observed from any other location, either within or outside the circle – then new measurements must be obtained and this analysis performed once more. And finally, the above analysis explores only the possibility of celestial alignments at the horizon, i.e. when celestial objects are so low as to touch the horizon, leaving open the possibility of alignments to celestial objects within the recumbent arrangement but higher up in altitude. To tackle these points is outside the scope of this paper, but they are highlighted here as avenues for future exploration, which should focus on alternative ways of thinking about these circles, the placement of their recumbent stones, and any potential alignments.

## 6. Concluding remarks

This paper introduced a probabilistic framework for the quantitative analysis of structural orientations. It began by defending the use of probability distributions to model orientation measurements, and then used those distributions to infer statistically significant orientations that are unlikely to be due to chance alone. As such, it not only formulated a quantitative foundation for further theoretical and methodological developments but also a robust analytical framework for the analysis of orientation datasets. The latter can be summarised as the following four-stage process:

i. decide on a probability distribution to model the orientation measurements;
ii. coordinate-transform the azimuthal distributions into declination distributions;
iii. aggregate the probability distributions into an SPD; and
iv. perform the significance test.

Having done this, one will be in a much stronger position to consider, and interpret, any correlations between statistically significant deviations from the null hypothesis and the risings or settings of celestial objects. In fact, it is not merely skyscape archaeologists that will be interested in this methodology, but also landscape archaeologists interested in alignments to geographic features. The same framework and methodology can be easily deployed, except that step ii above can be skipped when one is interested in landscape targets, therefore further bridging the gap between landscape and skyscape archaeology.

This approach has the potential to raise the analysis of the orientations of archaeological structures to a new standard, comparable to that of other quantitative fields that routinely use p-values to support their finds and claims. This method, although somewhat computationally intensive, is robust enough to be applicable to datasets of any size, from very small local studies to larger regional, national and international projects. With the right amount of reflexivity over one's data and assumptions, it will not only allow for an independent reassessment of past archaeoastronomical claims but also form a robust foundation for future skyscape research.

Two important theoretical points are also foregrounded by the methodology developed here. On the one hand, it highlights the role played by *a priori* assumptions that can significantly affect one's inferences. In the proposed framework these assumptions form the basis for the choice of measurement distribution (step i) and, therefore, need to be explicitly reported and justified. The same applies for how the coordinate-transformation is achieved when in the absence of 360° horizon profiles (step ii), as well as the significance level that is chosen (step iv). These should not be seen as limitations or hindrances but rather as features: instead of blurring and hiding researchers' biases and assumptions, the proposed framework exposes and foregrounds them. This has the advantage of making the scholarly endeavour more open and transparent which, as a by-product, will lead to more robust peer reviews and enhance cross-disciplinary understanding of the process of skyscape archaeology.

On the other hand, the proposed methodology deals explicitly and exclusively with the issue of statistical significance and not with intentionality. Intentionality has been a major concern for archaeoastronomers who attempt to demonstrate that the alignments they

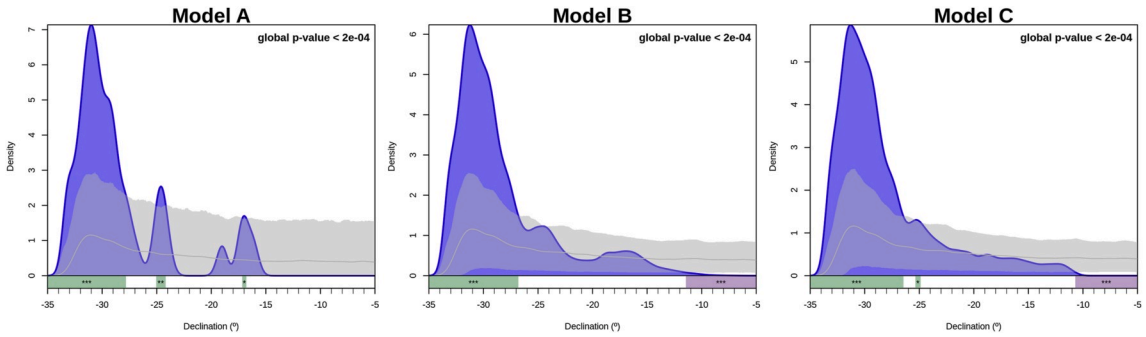


**Fig. 12.** Results of the significance test of the Ruggles (1999) dataset of the centre orientations of the Scottish Recumbent Stone Circles, showing the coordinate-transformed SPD (blue shade and line), the confidence envelope around the null hypothesis of random orientation (grey shade and lines), the global p-values obtained (top-right) and the regions where there is statistically significant deviation from the null hypothesis (coloured shaded regions at bottom of each image). Plots shown are for the three models discussed in the text. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

**Table 3**
Regions of significance identified by the method developed in this paper, including their declination range and local p-values, for each of the three models considered.

| Model | Declination Range | | Local p-value |
|---|---|---|---|
| | Min | Max | |
| **A** | −35.0° | −27.8° | <0.0002 (***) |
| | −25.1° | −24.2° | 0.0060 (**) |
| | −17.2° | −16.8° | 0.0244 (*) |
| **B** | −35.3° | −26.8° | <0.0002 (***) |
| **C** | −35.3° | −26.5° | <0.0002 (***) |
| | −25.4° | −24.9° | 0.0238 (*) |

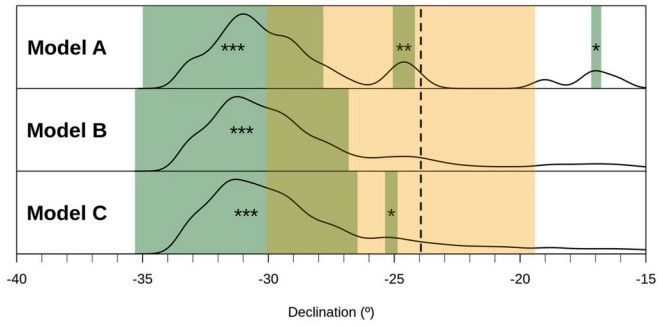


**Fig. 13.** SPDs (black curves) and regions of significance (green-shaded areas) for the three models considered, compared to the declinations of the December solstice sun (vertical dashed line) and summer full moon (orange-shaded area). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

observe today were intentionality incorporated into the structures they study (e.g. Ruggles, 2011). One methodological school – the so-called *green archaeoastronomy* — has relied on statistics as a proxy for intentionality under the largely unquestioned reasoning that a statistically significant pattern in orientation is likely to be due to an intentional desire to orient structures in the same way. The concepts of intentionality and statistical significance thus became conflated. Intentionality, however, is beyond the scope of inferential statistics, and certainly beyond the scope of this probabilistic framework. This framework can only be used to infer patterns of orientation that are unlikely to be due to chance — but those patterns might have emerged for a variety of reasons, including, but not exclusive to, celestial intentionality. Therefore, intentionality should not be argued for solely with recourse to a significance test, but also with evidence that is independent from the orientations themselves, and the principles of skyscape archaeology tells us that one should always start by seeking it in the wider archaeological record and the socio-historic context of the structures being studied (Silva and Henty, 2018).

This probabilistic framework should not be regarded as the endgame of quantitative orientation studies – this is but a stepping stone to bringing archaeoastronomy to the 21st century. In fact, statisticians across a variety of fields, including quantitative-minded archaeologists, are moving away from significance testing and p-value estimation towards the exploration of more informative alternatives such as approximate Bayesian computation and likelihood-based model selection (e.g. Burnham and Anderson, 2002). These more advanced methods allow not only for the empirical data to be statistically compared against the null hypothesis of randomness but also against other hypotheses that can explicitly incorporate landscape and skyscape targets. It is hoped that the framework presented in this paper might provide a robust baseline for the statistical analysis of structural orientations, but also act as a springboard to future development and application of even more robust statistical tools.

## 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.

## Acknowledgments

The author would like to thank James Steele for very early discussions on this topic, Victor Reijs for putting an early iteration of the framework presented here to the test, A. César González-García for further tests and discussions, and the anonymous reviewer for constructive feedback. Any remaining errors and opinions are the author's sole fault.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jas.2020.105138.

## References

Assmann, J., 2008. Death and Salvation in Ancient Egypt. Cornell University Press, Ithaca and London.
Belmonte, J.A., Shaltout, M., 2006. On the orientation of ancient Egyptian temples: (2) new experiments at the oasis of the western desert. J. Hist. Astronomy xxxvii 173–192.
Belmonte, J.A., Shaltout, M., 2009. In Search of Cosmic Order: Selected Essays on Egyptian Archaeoastronomy. Supreme Council of Antiquities.

Belmonte, J.A., Shaltout, M., Fekri, M., 2008. On the orientation of ancient Egyptian temples: (4) Epilogue in Serabit El Khadim and overview. J. Hist. Astron. xxxix, 181–211.
Belmonte, J.A., Shaltout, M., Fekri, M., 2009. Astronomy, landscape and symbolism: a study of the orientation of ancient Egyptian temples. In: Belmonte, J.A., Shaltout, M. (Eds.), In Search of Cosmic Order: Selected Essays on Egyptian Archaeoastronomy. Supreme Council of Antiquities Press, Cairo, pp. 215–283.
Belmonte, J.A., Fekri, M., Abdul-Hadi, Y.A., González García, A.C., 2010. On the orientation of ancient Egyptian temples: (5) testing the theory in Middle Egypt and Sudan. J. Hist. Astronomy xli 65–93.
Belmonte, J.A., González-García, A.C., Polcaro, A., 2013. Light and shadows over petra: astronomy and landscape in Nabataean lands. Nexus Netw. J. 15, 487–501. https://doi.org/10.1007/s00004-013-0164-6.
Benjamin, D.J., et al., 2017. Redefine statistical significance. Nat. Hum. Behav. https://doi.org/10.1038/s41562-017-0189-z.
Bevan, A., Lake, M., 2016. Computational Approaches to Archaeological Spaces. Routledge, Oxon and New York.
Bevan, A., Colledge, S., Fuller, D., Fyfe, R., Shennan, S., Stevens, C., 2007. Holocene fluctuations in human population demonstrate repeated links to food production and climate. Proc. Natl. Acad. Sci. Unit. States Am. 114 (49), E10524–E10531. https://doi.org/10.1073/pnas.1709190114.
Bradley, R., 2005. The Moon and the Bonfire: an Investigation of Three Stone Circles in North-east Scotland. Society of Antiquaries of Scotland, Edinburgh.
Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, second ed. Springer, New York and Berlin.
Bronk Ramsey, C., 2001. Development of the radiocarbon calibration program. Radiocarbon 43 (2), 355–363.
Crema, E.R., Bevan, A., Sienna, S., 2017. Spatio-temporal approaches to archaeological radiocarbon dates. J. Archaeol. Sci. 87, 1–9. https://doi.org/10.1016/j.jas.2017.09.007.
Efron, B., Hastie, T., 2016. Computer Age Statistical Inference: Algorithms, Evidence and Data Science. Cambridge University Press, Cambridge.
Eve, S.J., Crema, E.R., 2014. A house with a view? Multi-model inference, visibility fields, and point process analysis of a Bronze Age settlement on Leskernick Hill (Cornwall, UK). J. Archaeol. Sci. 43, 267–277. https://doi.org/10.1016/j.jas.2013.12.019.
Fisher, R.A., 1958. Statistical Methods for Research Workers, thirteenth ed. Hafner, New York.
Freeman, P.R., Elmore, W., 1979. A test for the significance of astronomical alignments. Archaeoastronomy 1, S86–S96 supplement to the Journal for the History of Astronomy 10.
González-García, A.C., Belmonte, J.A., 2011. Thinking hattusha: astronomy and landscape in the Hittite lands. J. Hist. Astron. 42, 461–494. https://doi.org/10.1177/002182861104200404.
González-García, A.C., Belmonte, J.A., 2014. Sacred architecture orientation across the Mediterranean: a comparative statistical analysis. Mediterr. Archaeol. Archaeometry 14 (2), 95–113.
González-García, A.C., Sprajc, I., 2016. Astronomical significance of architectural orientations in the Maya Lowlands: a statistical approach. J. Archaeol. Sci.: Report 9, 191–202. https://doi.org/10.1016/j.jasrep.2016.07.020.
Hawkins, G.S., 1964. Stonehenge: a neolithic computer. Nature 202, 1258–1261.
Henty, L., 2012. A voyage around the recumbent stone circles of north-east Scotland. In: Pimenta, F., Ribeiro, N., Silva, F., Campion, N., Joaquinito, A., Tirapicos, L. (Eds.), SEAC 2011 Stars and Stones: Voyages in Archaeoastronomy and Cultural Astronomy. BAR (International Series 2720), Oxford, pp. 164–169.
Henty, L., 2014. The archaeoastronomy of tomnaverie recumbent stone circle: a comparison of methodologies. Pap. Inst. Archaeol. 24 https://doi.org/10.5334/pia.464. Art 15.
Henty, L., 2015. An examination of the divide between archaeoastronomy and archaeology. In: Silva, F., Campion, N. (Eds.), Skyscapes: the Role and Importance of the Sky in Archaeology. Oxbow Books, Oxford, pp. 20–31.
Hutton, R., 2013. The strange history of British archaeoastronomy. J. Stud. Relig. Nat. Cult. 7 (4), 376–396. https://doi.org/10.1558/jsrnc.v7i4.376.
Kelly, D.H., Milone, E.F., 2005. Exploring Ancient Skies: an Encyclopedic Survey of Archaeoastronomy. Springer, New York.
Kosowsky, M., 2019. HeyWhatsThat [online]. http://www.heywhatsthat.com.
Lockyer, J.N., 1894. The Dawn of Astronomy: A Study of the Temple-Worship and Mythology of the Ancient Egyptians. Cassell and Company Limited, London.
Lockyer, J.N., 1906. Stonehenge and Other British Stone Monuments Astronomically Considered. Macmillan and Co, Ltd, London.
Lockyer, N., 1909. Surveying for Archaeologists. Macmillan and Co., Ltd, London.
Mukul, M., Srivastava, V., Jade, S., Mukul, M., 2017. Uncertainties in the shuttle radar topography mission (SRTM) heights: insights from the Indian Himalaya and Peninsula. Sci. Rep. 7, 41672. https://doi.org/10.1038/srep41672.
Neyman, J., 1950. First Course in Probability and Statistics. Henry Holt and Company, New York.
Neyman, J., Pearson, E.S., 1933. On the problem of the most efficient tests of statistical hypotheses. Phil. Trans. Roy. Soc., Ser. A 231, 289–337.
North, B.V., Curtis, D., Sham, P.C., 2002. A note on the calculation of empirical P values from Monte Carlo procedures. Am. J. Hum. Genet. 71, 439–441. https://doi.org/10.1086/341527.
R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. URL: https://www.R-project.org/.
Ruggles, C.L.N., 1984. Megalithic Astronomy: A New Archaeological and Statistical Study of 300 Western Scottish Sites. BAR (British Series 123, Oxford.
Ruggles, C.L.N., 1999. Astronomy in Prehistoric Britain and Ireland. Yale University Press, London.
Ruggles, C.L.N., 2011. Pushing back the frontiers or still running around the same circles? 'Interpretative archaeoastronomy' thirty years on. Proc. Int. Astron. Union 7 (S278), 1–18. https://doi.org/10.1017/S1743921311012427.
Ruggles, C.L.N., 2015. Analyzing orientations. In: Ruggles, C.L.N. (Ed.), Handbook of Archaeoastronomy and Ethnoastronomy. Springer Science, New York, pp. 411–425. https://doi.org/10.1007/978-1-4614-6141-8_26.
Ruggles, C.L.N., Burl, H.A.W., 1985. A New Study of the Aberdeenshire Recumbent Stone Circles, 2: Interpretation. Archaeoastronomy (supplement to the Journal for the History of Astronomy) 8, S25–S60.
Schaefer, B.E., 2006. Case studies of three of the most famous claimed archaeoastronomical alignments in north America. In: Bostwick, T.W., Bates, B. (Eds.), Viewing the Sky through Past and Present Cultures; Selected Papers from the Oxford VII International Conference on Archaeoastronomy. Pueblo Grande Museum, Phoenix, pp. 27–56.
Scott, D.W., 1992. Multivariate Density Estimation: Theory, Practice, and Visualization. Wiley, New York.
Shaltout, M., Belmonte, J.A., 2005. On the orientation of ancient Egyptian temples: (1) Upper Egypt and lower Nubia. J. Hist. Astron. xxxvi, 273–298.
Shaltout, M., Belmonte, J.A., Fekri, M., 2007. On the orientation of ancient Egyptian temples: (3) key points in lower Egypt and siwa oasis, Part 1. J. Hist. Astron. xxxviii, 141–160.
Shaw, I. (Ed.), 2000. The Oxford History of Ancient Egypt. Oxford University Press, Oxford.
Shennan, S., et al., 2013. Regional population collapse followed initial agriculture booms in mid- Holocene Europe. Nat. Commun. 4, 2486. https://doi.org/10.1038/ncomms3486.
Silva, F., 2014. 'A tomb with a view': new methods for bridging the gap between land and sky in megalithic archaeology. Adv. Archaeol. Pract. 2 (1), 24–37. https://doi.org/10.7183/2326-3768.2.1.24.
Silva, F., 2017. Inferring alignments I: exploring the accuracy and precision of two statistical approaches. J. Skyscape Archaeol. 3 (1), 93–111. https://doi.org/10.1558/jsa.31958.
Silva, F., 2019a. skyscapeR: skyscape archaeology data reduction, visualization and analysis. R package version 0.5.9001. URL: https://github.com/f-silva-archaeo/skyscapeR.
Silva, F., 2019b. On measurement, uncertainty and maximum likelihood in skyscape archaeology. In: Henty, L., Brown, D. (Eds.), Visualizing Skyscapes: Material Forms of Cultural Engagement with the Heavens. Routledge, Oxon, pp. 55–74.
Silva, F., Henty, L., 2018. Editorial. J. Skyscape Archaeol. 4 (1), 1–5.
Stubner, R., Reijs, V., 2019. swephR: high precision Swiss ephemeris. R package version 0.3.0. URL: https://CRAN.R-project.org/package=swephR.
Taylor, John, 1997. An Introduction to Error Analysis: The Study of Uncertaintites in Physical Measurements. University Science Books.
Thom, A., 1967. Megalithic Sites in Britain. Oxford University Press, Oxford.
Thom, A., Thom, A.S., 1980. Megalithic Rings. BAR (British Series 81, Oxford.
Timpson, A., et al., 2014. Reconstructing regional population fluctuations in the European Neolithic using radiocarbon dates: a new case-study using an improved method. J. Archaeol. Sci. 52, 549–557. https://doi.org/10.1016/j.jas.2014.08.011.
Tilley, C., 1994. A Phenomenology of Landscape. Berg, Oxford.
Welfare, A., 2011a. Great Crowns of Stone: the Recumbent Stone Circles of Scotland. Royal Commission on the Ancient and Historical Monuments of Scotland, Edinburgh.
Welfare, A., 2011b. Great Crowns of Stone: the Recumbent Stone Circles of Scotland – Gazetter and Appendices. Royal Commission on the Ancient and Historical Monuments of Scotland, Edinburgh. Available online: https://www.historicenvironment.scot/archives-and-research/publications/publication/?publicationId=cc24fec5-c34c-4b97-9d79-a58600e19733.
Zucchini, W., 2003. Applied smoothing techniques Part 1: kernel density estimation. Available online: http://staff.ustc.edu.cn/~zwp/teach/Math-Stat/kernel.pdf. (Accessed 1 June 2019).