Journal of Astronomical History and Heritage, 28(1), 219–239 (2025).

POINT-PATTERN MATCHING AND PROBABILITY IN ARCHAEOASTRONOMY

Alberto Vecchiato, Gabriella Bernardi and Alberto Cora
National Institute of Astrophysics (INAF) – Astrophysical Observatory of Turin, via Osservatorio 20, 10025 Pino Torinese (TO), Italy.
E-mails: alberto.vecchiato@inaf.it ; gabriella.bernadi@inaf.it ; alberto.cora@inaf.it

Abstract: Archaeologists and archaeoastronomers are often faced with the problem of deciding whether a specific archaeological artifact is an intentional representation of an asterism on the sky or rather if such a resemblance is purely coincidental. Sorting out this kind of problems is difficult also because of the lack of a statistical evaluation of the probability of the alleged similarity, and, although in the literature one can find a few attempts in this direction, there is no general consensus about a procedure with a rigorous statistical background.

The aim of this work is to propose one possible rigorous approach, built on the principles of the Point-Pattern Matching (PPM) technique used in astronomy and in other fields of science, and at the same time on those of the axiomatic theory of probability.

We first analyze the similarities and the differences between the PPM in astronomy and archaeoastronomy, showing that a simple least-squares fit can be used to find the best matching between the asterism and the artifact. Then we discuss the interpretation of the residuals of such a fit in other approaches found in the literature, and we propose a different use of these residuals to estimate the matching probability by means of a Monte Carlo simulation. We test our procedure on the well-known case of the three Giza pyramids and their alleged resemblance with the stars of the Orion Belt, and we use the results to stress that the probability found in this way represents just the matching probability that a single random extraction can match a single asterism, while the final probability has to take into account the matching probability with different asterisms by the complete population of similar artifacts. This requires to give a reliable estimation of the number of eligible asterisms n_A and of that of the artifacts n_G.

It is in general very difficult to obtain a precise estimation of n_A and n_G, and although one can finally obtain a reasonable conclusion for the case of the Giza pyramids, each case makes a story on its own and it has to be studied individually, with no guarantee of reaching a conclusive result.

We therefore conclude that the proposed method is certainly based on a solid statistical background and that it gives a reliable estimation of the matching probability for a single extraction. The definition of the overall probability for any asterism and the complete population of artifacts is also straightforward in case of independent extractions, but the difficulty of providing a reliable estimation of n_A and n_G can hamper the chances of deciding on the issue of intentionality solely on the basis of a probabilistic approach. However, it is likely that this analysis can provide useful results that can be used together with other pieces of evidence.

Keywords: archaeoastronomy; statistics; probability.

1 INTRODUCTION

Archaeoastronomy is a subsidiary discipline of both archaeology and astronomy that investigates the astronomical content of archaeological relics (Sinclair, 2006). By bringing up evidence of alignments with specific astronomical phenomena, or other kinds of astronomy-related features, these studies can help us understand the civilizations of the past.

Indeed, astronomy is universally recognized as one of the most ancient among human activities, and it is well known that it played an important role in society since very remote antiquity. It is thus not difficult to accept, even in the absence of written records, that traces of the connection between astronomy and an ancient civilization can be found in its archaeological relics. The debate is not around the existence of the influence of astronomy in ancient civilizations, but rather on its characterization, namely on the possible astronomical interpretation of specific artifacts, or on the usefulness of astronomy in some specific archaeological problems.

For example, when it comes to each specific case, the reliability of the archaeoastronomical results and methods is often criticized, especially among the archaeological community (Polcaro, 2016). According to the cited work, these difficulties have multiple causes; the lack of familiarity of archaeologists with the subject certainly plays an important role, but this paper identifies another important factor also in the quality of some archaeoastronomical works. In particular, it stresses the difficulty of giving a rigorous and accurate statistical assessment to the conclusions of the research. The need for a statistical assessment is indeed included among the list of checks, inspired from the standard practice in pharmacology, that, according to the author, any archaeoastronomical work should meet to be considered scientifically viable.

A case in point of the statistical issue is the problem of deciding upon the intentionality of a

particular astronomical alignment (Polcaro, 2016). In a similar vein, in this work we concentrate on another example, namely the similarity of some artifacts with significant celestial objects and the alleged intentionality of such a similarity. One popular example of this kind of problem is a set of French Gothic cathedrals that would reproduce the shape of the Virgo constellation, and another one, probably even more famous, is the alleged resemblance of the three famous pyramids of Giza, Egypt, with the stars in Orion's Belt, as well as various suggested human-made representations of the Pleiades or of the Little Dipper (Charpentier, 1975; Cora, 2017).

Identifying a specific shape from a series of points is a well-known topic of Information Technology dubbed 'Point-Pattern Matching' (PPM), and for this reason our work took a cue from the existing literature. However, we have to accept that such a problem develops in a different way in archaeoastronomy. This matter is presented in more detail in Section 2.

The aim of the present paper is thus the definition of a statistically sensible way of testing the hypothesis of an intentional correspondence between a human artifact and an asterism in the sky. This has to include both a quantitative definition of similarity between two sets of 2D points and an estimation of the probability that such a similarity can be intentional by checking it against the null hypothesis of a random placement. Both problems are tackled in Section 3, while in Section 4 the method and its results are practically illustrated based on a case study.

Section 5 explains the further problem of making a full statistical evaluation of the probability and the possible limitations of a purely statistical method. Finally, in Section 6 we present our conclusions.

2 THE POINT-PATTERN MATCHING PROBLEM IN ASTRONOMY AND ARCHAEOASTRONOMY

The PPM problem plays an important role in many different disciplines and applications, from computer vision to computational chemistry, and from shape reconstruction to medical imaging (Li et al, 2003). In short, it can be defined as a method to identify an object by searching for correspondences between subsets of points in two different sets. In this, it can be considered a sub-domain of feature-based methods that can use different structures such as, e.g., lines, curves, edges, or surfaces.

The PPM problem is not unknown in astronomy either; quite the contrary, it has been continuously investigated at least during the last four decades (see e.g. Groth, 1986; Murtagh, 1992). The ideas contained in these papers were extensively used and further developed for many purposes in subsequent works. For example, Pál and Bakos (2006) extended the approach of Groth (1986) to tackle the astrometric matching problem in wide-field surveys, where the number of stars is much larger and the contribution of non-linear transformations between the two sets is not negligible because of the amplitude of the field of view. *Astrometry.Net*, instead, (Lang et al., 2010) is used to identify the stars in a given image and to generate its World Coordinate System (WCS) information, namely to position such an image into a standard all-sky reference system. The declared goal of this software is to make sharing astronomical data easier, thus making their exploitation wider and more effective in the scientific community. The problem of bringing two or more images to a common reference system, also called 'image registration', is met also in other projects, like *Montage* (Berriman and Good, 2017; Jacob et al., 2009), set forth to ease the process of astronomical image mosaicking. Finally, an image registration problem that uses PPM techniques is involved also in *Astroalign* (Beroiz et al., 2020), a Python module conceived for more 'observation-related' applications, like managing telescope night-drifting, astrophotography, coaddition and image subtraction, where bringing the images to a common reference frame needs no WCS information.

This list could be longer, and the interested reader can easily find some more examples in the references of the cited papers, but to our knowledge what we present here is representative of the current uses of the PPM techniques in astronomy.

In general, astronomical applications imply an input in the form of images or a list of coordinates that represent stars, and they answer questions like "Where in the sky are these stars?" or "How can I join these images to form a single map?" But suppose for a moment that one of these tools is fed with some fake images of a star field, or some fake lists of coordinates, both randomly generated. It is not difficult to admit that, given the large number of potential targets, these algorithms would likely find a correspondence in the sky, at least in some cases. Fortunately, in the real world of 'common astronomy' we do not have to face this disturbing possibility, because we can be reasonably confident that our data are representations of stars, or other celestial objects, but this is not the case in archaeoastronomy, as we have seen in the Introduction.

The differences between the two cases could hardly be more striking:

- in astronomy we can rest on the fact that our input data are an actual representation of a portion of the sky, whereas in archaeoastronomy the input data are artifacts that are claimed to represent some specific significant stellar asterism;
- in astronomy, having no clue about which stellar pattern is represented in our image is perfectly normal, whereas in archaeoastronomy we are usually pointed towards an astronomical interpretation of the artifact precisely by the similarity of the pattern with a specific asterism; moreover, we can take for granted that the stars allegedly represented are among those visible to the naked eye, which greatly reduces the number of potential targets;
- in astronomy we have different ways of verifying a given correspondence found by means of a PPM method, whereas in archaeoastronomy we usually have just an artifact with a pattern and the stars in the sky; we thus need to give a quantitative estimation of the probability that the matching is not a false positive, namely that it happened just by chance.

The PPM problem in archaeoastronomy, therefore, shares with the astronomical case the need to find the optimal transformation between the 'target' asterism and its alleged representation, that carries one onto the other. For this task, a natural choice is a general linear(-ized) conformal transformation, as also adopted in several earlier works.

Once this task has been accomplished the remaining part of the problem, not present in the astronomical case (at least not in these terms), is to quantify the probability that the matching found with this transformation can be attributed to pure chance. We show below that this requires us to define a quantitative measure of the 'similarity', which can be compared with the analogous measure estimated for a random pattern.

Both of these tasks are tackled in the next section.

3 LEAST-SQUARES MATCHING AND PROBABILITY ESTIMATION

As explained in Sections 1 and 2, the application of the PPM problem to archaeoastronomy requires the solution of two different problems, that is:

(1) to give a quantitative evaluation of the goodness of matching between two sets of 2D coordinates;

(2) to estimate the probability that such a matching can occur by chance.

Each of these two points is tackled in Sections 3.2 and 3.3, respectively. First of all, however, it is worth discussing a preliminary step, namely, how to prepare our inputs, that is, the sets of 2D points.

3.1 Building the Input 2D Coordinate Sets

The sky appears as a spherical surface to our eyes, so in principle it should be mapped with a grid of curved coordinates, and the natural choice for the asterisms that have to be checked would be any celestial longitude–latitude coordinate system, possibly computed at an appropriate epoch. An accuracy of the order of 1 arcmin is quite adequate for archaeoastronomy (an unaided normal human eye has a maximum angular resolution of about one arcminute, corresponding to the 20/20 value of the Snellen visual acuity scale, see Yanoff and Duker, 2009, 54) so one can safely neglect corrections due to the actual 3D distribution of celestial objects, like parallactic shifts.

Unlike the case of astronomical alignments, in the PPM problem changes in the coordinates due to the varying orientation of the Earth, because for example of precession and nutation, can also be neglected, because they are the same for all the objects in the sky, and thus they do not alter the 'shape' of the asterism, i.e. the relative positions of its components.

On the other hand, the need to investigate very ancient artifacts dating back several thousand years, implies that possible corrections due to proper motions should be checked.

Another aspect that can alter the shape of our set of points is the coordinate representation. In principle, the asterism in the sky should be mapped in spherical coordinates, namely those of a curved manifold, but it is common practice, from the plate reductions, to use the so-called standard coordinates (ξ, η)

$$\xi = \frac{\cos\delta \sin\Delta\alpha}{\sin\delta \sin D + \cos\delta \cos D \cos\Delta\alpha} \quad (1)$$

$$\zeta = \frac{\sin\delta \cos D - \cos\delta \sin D \cos\Delta\alpha}{\sin\delta \sin D + \cos\delta \cos D \cos\Delta\alpha} \quad (2)$$

which are obtained (see e.g. Green, 1985) by projecting the spherical coordinates (α, δ) of an object on a spherical surface onto the plane tangent to the unit sphere at a given point with coordinates A, D) (Figure 1), and where $\Delta\alpha = \alpha - A$.

A potential issue of this approach is that the projection, rigorously speaking, alters the shape of a figure, which is exactly the property we want to compare. Thus, we need to determine

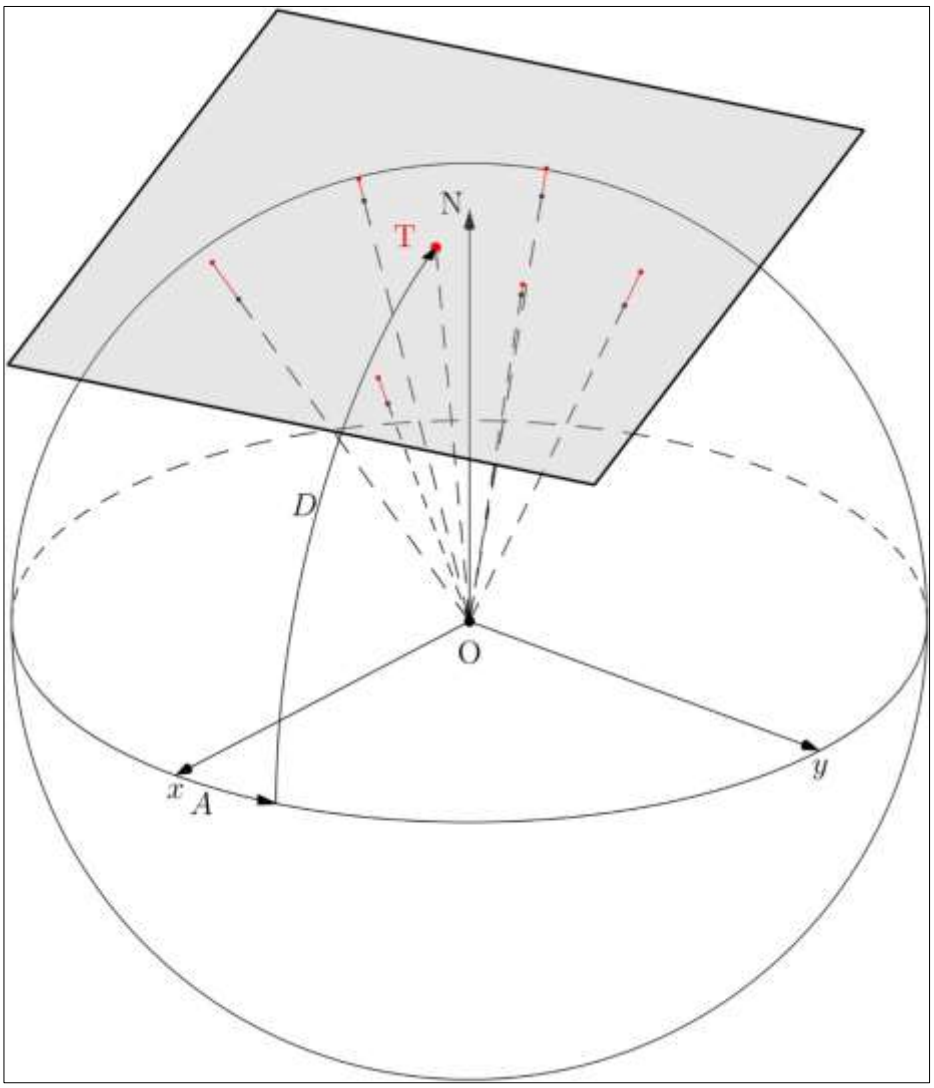


Figure 1: The black points with coordinates (α, δ) on the celestial sphere are projected onto a plane tangent in $\mathrm{T} = (A, D)$ to the sphere. The projection is implemented by Equations (1) and (2), which endow such a plane with the reference grid of the so-called standard coordinates ξ, η. In the figure, the projected points are represented by the red dots.

when these deformations are negligible, that is, the limits of applicability of this method. As it can be immediately understood by comparing a planisphere with a globe, deformations are larger for figures of larger angular size. Moreover, since the shape is determined by the angles between the sides of the figure, the deformation introduced by this projection can be measured from the difference $\Delta\alpha$ between the angles on the curved and on the flat projection surface, respectively.

Equation (A.3) allows for quantitative estimation of $\Delta\alpha$ as a function of the angular size a of the figure, showing that, for example, $\Delta\alpha \simeq 1.1$ arcmin for $a \simeq 2.25$ deg. Therefore, projections can be safely adopted for most asterisms, except possibly for large constellations.

Similar considerations apply for the candidate on-ground representations. The archaeological artifacts having a claimed astronomical correspondence in the sky can be of various types. Their size can range from less than 1 m^2 to monumental representations covering distances of a few kilometers, or even hundreds of kilometers for the most extreme claims. However, their angular sizes depend on the ratio between the physical size and that of the Earth. The previously mentioned angular size of 2.25 deg, for example, corresponds to a distance of about 262 km on the surface of our planet. Once again, this guarantees the applicability of the projection approach for the vast majority of cases.

It is also worth noting that what is actually relevant is not only the absolute size of the asterism ion the sky or of its representation, but

also the ratio between the two angular sizes, because figures covering similar angular areas are deformed in similar ways. In other words, if we had an asterism and an alleged representation of the same angular size, we could still use the standard coordinates in case their deformations are significant.

Finally, we are not strictly constrained in the choice of the tangent point of our projection. However, it is clear that it would be convenient to minimize the deformations coming from the projection. This naturally calls for using the barycenters of each set, namely

$$A \triangleq \alpha_B = \frac{1}{n}\sum_{i=1}^{n} \alpha_i \quad (3)$$

$$D \triangleq \delta_B = \frac{1}{n}\sum_{i=1}^{n} \delta_i \quad (4)$$

In conclusion, for the stars the required input set of 2D coordinates on a flat surface is obtained by the above-mentioned projection, a procedure that can be applied to the artifacts as well, unless they are small enough to start directly with flat coordinates.

3.2 Description of the Matching Method

The general input of a PPM problem is constituted by two sets of points with different cardinality, say n and m, and the algorithm tries to match a subset of one of the sets with cardinality $k < \min(n, m)$, with an equal number of points of the other set. For what was mentioned in Section 2, however, the problem under investigation can be restricted by considering two sets with an identical number of points, because we start from an initial guess about the correspondence between a set of stars and a set of different objects. In other words, in a classical PPM problem we start by knowing that a set of points represents a specific kind of object, and we try to find the one that best matches this set among a list of possible candidates by comparing, in general, subsets of points of these candidates. In archaeoastronomy, as already mentioned, we start 'from the other side of the problem', namely we have already identified a correspondence, and we need to prove this claim.

Regarding the cardinality of the sets, in a certain sense subsets are used also in archaeoastronomy, but this cannot be considered a pure PPM problem. Probably, the best way to understand such a statement is by example. When it is stated that the three pyramids of Giza represent the three stars in Orion's Belt, it would be natural to wonder why the ancient Egyptians should have picked out just these stars from among all stars. In fact, these were part of a larger constellation named 'Sah' that included stars from Orion and Lepus (Shaltout and Belmonte, 2005). Obviously, there could be many reasons for such a choice, but usually this matter has to be understood *a priori* with specific motivations that have nothing to do with a PPM problem. For example, one might say that the pyramids were aligned with the stars on a significant date, or that these three stars had a specific meaning in Egyptian mythology. In other words, we need solid justifications for the choice of a specific subset, without which even a perfect match would make little sense.

For this reason, we worked with the hypothesis that the PPM problem of choice in archaeoastronomy always takes place between two sets of equal cardinality. We then adopted a method similar to that followed by Wang and Zhang (2012), which finds the best possible match between the two configurations by modeling it with a conformal Euclidean transformation.

In our case, in fact, we need to identify a transformation that allows a geometric matching between two 'shapes'. Such a matching is necessarily based on a comparison of angles; two patterns match one another when corresponding triangles formed with any three corresponding points of the two sets are similar, that is, when they form the same angles between their sides. For this reason, the transformation between the two sets has to keep angles unaltered, or in mathematical language, it has to be conformal. In a 2D Euclidean framework, the most general conformal transformation uses 4 parameters (see e.g. Vecchiato, 2017 and references therein) expressing a combination of: a common-factor scaling (1 parameter), a rotation (1 parameter), a translation (2 parameters) and, possibly, a parity transformation (no parameters), but we can make it simpler.

The parity transformation implies an inversion of the symmetry of the object that we need to match with. We decided to skip it for the moment because it would mean considering cases of a 'mirrored' representation of the object. We cannot exclude that, in specific cases, this might be meaningful, but in this paper we are more interested in establishing the principles of a general approach that can be extended in future works. We can also neglect the translation and its 2 parameters as long as we make the assumption that such a transformation is implicitly taken into account by 'overlapping' the two tangent points, namely the origin of the two standard coordinates' reference systems (see Appendix 2). This implies a sub-optimal realization of the matching transformation that may be potentially relevant, since a bad choice of the two tangent points would completely disrupt the matching. However, in our case, a first visual correspondence between pairs of artifacts and

stellar objects is established by eye, and in such a situation, a convenient choice is the barycenter of the two sets as in Equations (3) and (4). These two points, as mentioned in the previous section, minimize the deformations generated by the projection onto the tangent plane of the standard coordinates. It is worth stressing again that choosing to estimate the translation in this way does not produce the best possible matching, so the outcome we are obtaining has to be intended as an upper limit of, although probably quite close to, the best result.

After having converted the spherical coordinates of the stars and, if needed, of the artifacts in their standard coordinates $(\bar{\xi}_i, \bar{\eta}_i)$ and (ξ_i, η_i), with $i = 1, \dots, n$, the remaining scaling and rotation parameters are obtained by solving for the unknown parameters a and b in the least-squares sense the following system of equations

$$\begin{cases} \bar{\xi}_1 &= \xi_1 a - \eta_1 b \\ \bar{\eta}_1 &= \xi_1 b + \eta_1 a \\ \vdots & \quad \vdots \\ \bar{\xi}_n &= \xi_n a - \eta_n b \\ \bar{\eta}_n &= \xi_n b + \eta_n a \end{cases} \tag{5}$$

where, as described in Appendix 2, we formulated the problem in such a way that the resulting system is linear and the scale factor s and the rotation angle θ can be obtained from a and b with Equations (B.9) and (B.10).

The result is the best match between the two sets in the least-squares sense, namely, as mentioned in Appendix 2, a new set of points $\{\tilde{P}_i\} = \{(\tilde{\xi}_i, \tilde{\eta}_i)\}$ that minimizes the sum of the squares of the distances (and obviously also the sum of the distances)

$$\begin{aligned} S_{\text{sq}}(d_i) &= \sum_{i=1}^{n} d_i^2 \\ &= \sum_{i=1}^{n} \left(\tilde{\xi}_i - \bar{\xi}_i\right)^2 + (\tilde{\eta}_i - \bar{\eta}_i)^2 \end{aligned} \tag{6}$$

of this approximate reconstruction with the 'target' points $\{\bar{P}_i\} = \{(\bar{\xi}_i, \bar{\eta}_i)\}$.

3.3 Matching Probability

Once the best matching parameters have been estimated with the procedure of the previous section, it is necessary to give a quantitative estimation of the probability of such a matching. This provides an evaluation of the probability that the similarity between these ancient monuments and the stellar cluster (or asterism) is due to pure chance.

In previous years, other researchers, including one of us, have already tried to develop a mathematical method to quantify the probability that ancient artifacts show a by-chance correspondence with celestial objects. It is thus worth spending some time giving an overview of the authors' current knowledge of the matter. This will give us a chance to explain why we decided to choose a different way for this task.

Gaspani (2012) suggested a method based on the Bernoulli distribution. The advantage of such a method is that its only input is the cardinality of the set. This is allowed by the characteristic of the adopted probability distribution, which admits just the two equiprobable states of 'success' and 'failure'. As long as such a correspondence can be determined *a priori*, one can derive a matching probability of the whole set with a very simple formula. In this way, the level of confidence of the matching increases with the cardinality of the set, and the author concluded that at least 9 elements were needed to reach a 95% confidence level, and 14 elements for the 3σ (99.7% confidence) level. The weak point of this approach, in our opinion, is that the author does not give a criterion for establishing the success or the failure of a correspondence. In particular, it does not provide a way to give a quantitative measure of the correspondence. Therefore, each investigator might make a different evaluation of an alleged correspondence. Moreover, the advantage of the method, namely that it only depends on the number of objects, is also its drawback, because it is intuitive to understand that a 'good' matching is less probable than a 'fair' matching, which in turn is less probable than a 'bad' matching. So the matching probability, in principle, should depend also on the goodness of the matching, and therefore a result that also takes into account this factor might very well differ significantly from the one based purely on the cardinality. In conclusion, this method seems useful to get a first rough evaluation of the matching probability. It can reasonably exclude some cases, and it can give an approximate idea of the amount of work needed to have a reliable result in other cases.

One of the present authors introduced for the first time the use of standard coordinates in this problem (Cora et al., 2018). He found the best match between the two sets of points by means of a transformation derived from the astronomical use of plate coordinates, while Mulas and Sanna (2014 in Italian, or 2023 in English), in a work aimed at establishing the matching probability with the Pleiades, matched the two sets using a slightly different method. They set up a common reference system by imposing the matching between a pair of objects with the pair of corresponding stars. That gave a reference direction and a normalized unit of measure, to which the positions of all the remaining stars and objects could be referred. The matching pair was chosen to be the one, among all the combinations, that minimized the

distances between the other stars and the corresponding objects. This procedure, of course, is not completely equivalent to a least-squares fit, but in practice it gives similar results.

Afterwards, both authors exploited a similar idea to compute the matching probability, which is based on a sort of 'darts and targets' model. Without providing too much detail, the artifacts are like 'darts' aiming at the 'target' stars placed on a dartboard. The probability that a single dart i reaches a distance d_i from one target is the ratio $P_i = d_i^2/R^2$ between the area of the circle of radius d_i and that of the dartboard, which is assumed to be a circle of radius R. Since all d_i are considered independent events, the matching probability for *n* objects at their specific distances results in

$$P_{\text{tot}} = \prod_{i=1}^{n} P_i \quad (7)$$

in the approach of Cora, and

$$P_{\text{tot}} = (\prod_{i=1}^{m} P_i)m! \quad (8)$$

in the work of Mulas and Sanna, where $m = n - 2$. It is beyond the scope of this paper to explain the reason for the difference between these two formulas; the reason they are reported here is to stress that in both approaches the matching probability is the result of a product of single probabilities, a fact that we will use later in this paper.

In fact, such a procedure for estimating the matching probability, shows a number of shortcomings.

First of all, the simple formula $P_i = d_i^2/R^2$ is based on the ratio between two circles, thus implying that the 'dartboard' region has a circular shape. Although one can figure out a way out of this issue in the case of different shapes, this type of formula is valid only if the circle of radius d_i is entirely contained into the area of the dartboard. Again, one can try to figure out a way to consider the case in which the small circle has just a partial intersection with this area, but the most general case would depend on the exact shape of the dartboard, which in many cases has a great deal of arbitrariness. For example, in the case of the three pyramids of Giza it would surely be hard to agree on a well-defined shape and size that they could have been distributed across. And even if the claimed representation is on a well-defined artifact, say a stone disk of a given shape and size, it could well have been different when it was made. We could continue along this line of reasoning, but this is more than sufficient to understand that this estimation strongly depends on the choice of the area where we decide to put the targets, which might easily be arbitrary.

Moreover, it is worth stressing again what we mentioned in Section 3.1, namely that the shape of a figure is determined by the angles between the sides of the figure, rather than by its size. In other words, any set of points can be decomposed into a set of triangles, and a shape matching can always be reduced to a relation of similarity between the corresponding triangles. Two triangles are similar if all their angles are equal, which does not depend on the size of the triangle, so one may wonder whether the matching probability should instead be completely independent of R. This issue, however, is worth a more in-depth investigation.

I fact, let us suppose that we have just one target point ($n = 1$) at the center of the dartboard, and that we launch a dart at random; in this case, the probability of reaching a certain distance d_i from the target point is precisely $P_i = d_i^2/R^2$. As already observed above, if the target is not at the very center of the board this definition already seems less safe, but we will not investigate this point further. Rather, we prefer to consider the case of a random launch on a board with $n > 1$ target points. In this case, we need to find the target point that is nearest to the launched dart.

Indeed, the problem of finding the distance of the nearest neighbor in a randomly placed distribution of *N* points is well studied. It is relevant also in astrophysics (Chandrasekhar, 1943), and was investigated by van Wamelen et al. (2004) for the 2-dimensional case. Without entering into the mathematical details, which can be found in the cited works, it is intuitive to understand that, with respect to the case of the single target, it will be much easier to reach the same distance d_i from any of the n points, because of the increased density of targets over the same area. In statistical jargon, the expectation value of d_i, namely 'the average distance of the nearest neighbor', becomes smaller for a larger n.

In other words, the typical distance d_i from a target point decreases when the density of these points increases, and therefore the value of P_i decreases as well. It can be shown that now it is possible to make such a value independent of the area of the board, but this can hardly be considered a success. In fact, in our interpretation, P_i should be a probability, which however has the odd characteristics of being smaller (i.e. more unlikely) when actually, as we have just noticed, it is easier (i.e. more likely) to obtain smaller values.

This difficulty propagates to the value of the matching probability. Indeed, keeping our reasoning at an intuitive level, we are reminded that the matching probability P_{tot}, regardless of the method we consider, is the product of many P_is, each of which is less than unity. If with many

target points we have to expect, on average, smaller P_is for any launch, it is easy to understand that we have to expect an even smaller P_{tot}. What we interpret as the matching probability, therefore, is affected by the same conundrum of the single P_i: it is apparently easier to obtain smaller values of P_{tot} with more points, when one should expect exactly the opposite.

Once again, it is possible to find a possible way out of this other difficulty by using a definition of P_i that takes into account the expectation value of the nearest neighbor, but this modified definition will take back the dependence on the area. With all these issues, we believe that continuing to search for a correct definition of the matching probability along these lines is not worth the effort, and that the problems we have highlighted are intrinsic to this general definition.

The above considerations, instead, suggest that it might be more convenient to adopt a different approach that interprets Equation (7) and the like as an *event*, rather than an actual probability, an approach that we will explain in the next section.

3.4 Monte Carlo Estimation of the Matching Probability

We have already thoroughly stressed the problem of having a probability that depends on the area chosen to represent the domain of possible events, while a similarity between two shapes should be independent of this parameter. We can (apparently) get rid of this dependence (at least for a circular surface) by considering the statistical expectation value of the nearest-neighbor distance, but this operation creates more problems than it solves. In particular, it highlights the fact that the resulting expectation value of P_{tot} for a random 'launch' of *n* 'darts' can easily become very small just by increasing the number of points to be matched.

Finding a result that is at odds with the behavior of a probability, given the above considerations, should not be a surprise. To the contrary, this is a clear indication that P_{tot} cannot be interpreted as a probability. Indeed, an expectation value should not be mistaken for a probability; the probability is the chance that a certain event can occur, while an expectation value is the most probable event among all the possible ones. We stress again that our goal is to find the probability of a specific matching against the null hypothesis of a 'random launch'. Thus, we have to find a way to compare the P_{tot} of our actual matching against the expectation value of the n points of the asterism.

For example, let us say that for a 10-point asterism, we obtained $P_{\text{tot}} = 10^{-10}$ with the method of Mulas and Sanna. If we interpreted this value as a probability one should be tempted to conclude that such a small value is a strong indication against an accidental matching, but an approximate estimation of the expectation value of the distance of the nearest neighbor with n targets yields a corresponding expectation value of $\langle P_i \rangle \sim 1/4(n+1) = 1/44$. This is, on average, what we should obtain for each of the 10 P_i s, therefore an accidental matching would give, on average, a value of $\langle P_{\text{tot}} \rangle = (1/44)^8 \cdot 8! \simeq 3 \cdot 10^{-9}$, which is not so different. Even worse, just imagine that $P_{\text{tot}} = 10^{-6}$. Even in this case, without knowing the expectation value for our case, such a small probability would be in favor of an intentional matching, but we know instead that this matching result is much *worse* than one should expect from a random case. Should one conclude, then, that this is a case of intentional mismatch?

The answer is obviously that we cannot know. Rather, this comparison apparently tells us two things:

(1) That the first matching is better than the second one;
(2) That (with respect to our rough estimations) the first matching is better than the one expected on average by a random attempt, and the second one is worse.

It is worth stressing that these conclusions are naturally implied by the use of a least-squares approach to find the best possible matching. The least-squares fit, in fact, entails the above cited criterion that the best matching is identified by the smallest sum of the squared residuals S_{sq} given by Equation (6). Therefore, it also means that a random 'launch' with sum S_{sq} is better or equal than the one of the artifact $\tilde{S}_{\text{sq}}$ if and only if $S_{\text{sq}} \leq \tilde{S}_{\text{sq}}$.

Such an interpretation, forced by the least-squares fit, implies that the function S_{sq} can be naturally interpreted as an event in the axiomatic approach to probability (Kolmogorov, 1950). In this context, an event is an element of a set *E* that contains all possible outcomes of a process; each event is associated with a probability, a number between 0 and 1, and all the probabilities must sum to 1. This last property expresses the concept that the set *E* contains all possible events, so we are certain that any possible outcome of our process is contained in *E*. For example, the launch of a fair 6-faced die has a set of 6 possible outcomes. The probability of a specific face (event) is 1/6 because it is a fair die, and it is obtained as the ratio between the number of 'favorable' events and the number of all possible events. In this example,

the elements of *E* are represented by integer numbers, but we can imagine a process in which the events are represented by real numbers.

This is the case of our launch-based model, in which each launch is associated with an S_{sq}. In mathematical jargon, this quantity is a 'continuous random variable', and favourable events can be unambiguously defined as those for which $S_{\mathrm{sq}} \leq \tilde{S}_{\mathrm{sq}}$, according to the 'smaller is better' criterion stemming from the least-squares fit. For this reason, it is reasonable to dub the random variable S_{sq} as 'matching goodness', with the agreement that a '0.1' goodness is better than a '0.2' goodness. If we wish to know the probability of having a given matching goodness $\tilde{S}_{\mathrm{sq}}$, we can simply imagine doing a large number N of launches, and counting the number of favourable events $\tilde{N}$ for which $S_{\mathrm{sq}} \leq \tilde{S}_{\mathrm{sq}}$. The probability, by definition, is the ratio $\tilde{N}/N$. This process is known as the 'Monte Carlo method'.

3.4.1 On the Operational Definition of the Random Variable

In our model the random variable is a real number that measures the matching goodness of an event, namely the proximity of the outcome of a launch to a set of stars identified with their angular coordinates on the sky. It has to be stressed that the way adopted to define such a number is somewhat arbitrary. The natural choice, as anticipated above, is clearly the quantity S_{sq}, but any quantity that meets the 'smaller is better' criterion implied by the least-squares approach is appropriate for our needs.

For example, another valid random variable alternative to S_{sq} is clearly $S = +\sqrt{S_{\mathrm{sq}}}$, or even $S_m = -\sqrt{S_{\mathrm{sq}}}$, if we agree to consider the inverse ordering of the events and to count as favorable those for which $S_m \geq \tilde{S}_{\mathrm{m}}$. The actual choice is not important if we limit ourselves to the matching probability estimation with the simple Monte Carlo procedure.

One choice that is not valid is anything that comes from the definition of P_{tot}, namely either from Equations (7) or (8). The reason, simply, is that P_{tot} is based on a product of residuals, rather than a sum, and the product does not necessarily preserve the ordering of the events established by S_{sq}. This property is essential to guaranteeing the correct computation of the matching probability with the Monte Carlo approach, and it is implicitly established by using the least-squares algorithm used to find the best superposition.

3.4.2 Estimation of the Matching Probability with the Monte Carlo Method

The Monte Carlo method is based on the possibility of having an operational definition of a 'pure chance event' for its extractions, namely our 'launches'. We define it as a random set generated by placing *n* objects (points) in a circle of radius 1 centered at the origin of the reference system of the standard coordinates described in Section 3.1. All objects in the set are placed independently of each other, and any location in the circle has the same probability of being selected; in mathematical jargon, the extraction of any point follows a uniform probability distribution within the area of the circle.

With this definition in mind, we can summarize the Monte Carlo procedure with the following steps:

(1) We run a Monte Carlo test that produces a large number N of random sets of n points, as described above.

(2) Each set is matched with the asterism under investigation as described in Section 3.2, thus producing a sample of by-chance (random) events e_i, $i = 1, \ldots, N$, and the corresponding values x_i of their random variables. Such a variable can be either S_{sq}, S, or any other formula that preserves the ordering of the matching goodness implied by the least-squares fit. It is important to remember that the algorithm for the fit assumes (see requirement 1 in Appendix 1) that the points of the target set and those of the random set are already ordered by correspondence. In a random extraction, obviously, we cannot take this requirement for granted, so we need to order the random set in advance. We take the order that gives the minimum value of x among all possible correspondences because we are searching for the best matching, and by definition this is the one that gives the smallest value of the residuals between the target and the reconstructed points, that is the ordering that produces the minimum x. These extractions will produce a sample of values of the random variable in the range $X = [x_{\mathrm{min}}, x_{\mathrm{max}}]$.

(3) The matching goodness $\tilde{x}$ of the artifacts and the asterism is evaluated in the same way as for the random events, under the condition that $x_{\mathrm{min}} \leq \tilde{x} \leq x_{\mathrm{max}}$.

(4) Finally, we estimate the matching probability by computing the ratio

$$P(\tilde{x}) = \frac{\tilde{N}}{N} \quad (9)$$

between the number $\tilde{N}$ of favorable events, namely those for which $x_i \leq \tilde{x}$, and the total number of extractions. An alternative way of computing the matching probability, which makes use of the so-called probability density function (PDF) estimated by a technique called Kernel Density Estimation (KDE, see e.g. Parzen, 1962 or Rosenblatt, 1956), is also included in the code that we implemented.

Finally, it is worth stressing that $P(\tilde{x})$ has to be interpreted as the probability that *one* by-chance extraction gives a matching with the asterism as good as $\tilde{x}$ or better. We return later to the importance of this detail.

3.5 Estimation of the Accuracy of the Matching Probability

The matching probability P is computed from two different quantities, the distribution of the matching goodness of the random launches with the asterism, and the estimation of the matching goodness of the actual artifact, $\tilde{x}$. Therefore, the accuracy of P depends on both of these quantities.

In principle, one can perform an infinite number of launches. In the language of statistics they are called the population of the events, and we will indicate it with $\mathcal{X}$. The entire population reproduces accurately the distribution of the matching goodness, and therefore it can give the true value $P_{\mathcal{X}}(\tilde{x})$ of the probability for $\tilde{x}$. However, we can perform only a finite number of launches, namely, we consider a sample X of the entire population. This will give an approximate representation of the distribution, and thus an approximate estimation of $P_{\mathcal{X}}(\tilde{x})$, that we name $P_X(\tilde{x})$, and it is intuitive to understand that the larger the number of launches, the more accurate the approximation is.

Leaving the mathematical details to Appendix 3, it can be shown that, in our case where $P_X(\tilde{x}) \ll 1$, the accuracy of this approximated value attained by a Monte Carlo run with N launches is

$$\sigma_{P_X(\tilde{x})} = \frac{\sqrt{\tilde{N}}}{N} \tag{10}$$

Another standard way to estimate the accuracy of a Monte Carlo run consists of repeating a sufficient number of times M, a Monte Carlo simulation with $n < N$ launches. Each of the M simulations gives an estimation $P_i^{(n)}, i = 1, \dots, M$ of the probability $P_{\mathcal{X}}(\tilde{x})$ for a sample of size n, and a more accurate estimation of the true probability and its accuracy can be computed with Equations (C.5) and (C.6) reported in Appendix 3.

The process is repeated for several increasing values of n, so that the estimation of the true probability eventually converges to a constant value, while that of the accuracy decreases following the law $\sigma_{P^{(n)}} = k\sqrt{n}$. We then fit the pairs $(n, \sigma_{P^{(n)}})$ to estimate the parameter k, and finally the accuracy for N is obtained by extrapolating the fit to $n = N$.

Both methods should give similar results for reasonably large values of M, n, and N, e.g. as long as we observe the known 'rule of thumb' for which the sample size should be $n \geq 30$.

It is worth noting that these methods give an *a posteriori* estimation of the accuracy of $P_X(\tilde{x})$. However, the Monte Carlo simulation can easily be long and computationally demanding, so it would be convenient to have an *a priori* estimation of the number of Monte Carlo iterations needed to attain a specific accuracy. Many such methods can be found in the literature (see, e.g. Newcombe, 1998; Wald, 1943; Wallis, 2021 and references therein), but the need to use them, and determine which is the best choice, has to be evaluated on a case-by-case basis.

Although this solves the problem of computing the accuracy of the distribution, it has to be recalled that the matching goodness $\tilde{x}$ of the artifact is computed via least-squares matching between the set of (standard) coordinates of the artifact itself (ξ_i, η_i) and that of the asterism $(\bar{\xi}_i, \bar{\eta}_i)$. However, all these quantities are affected by their own measurement errors, therefore in principle the corresponding error $\sigma_{\tilde{x}}$ should be taken into account, as a different value of $\tilde{x}$ will obviously affect the estimation of $P_X(\tilde{x})$.

A straightforward way to tackle the problem is to compute the variances σ_a and σ_b, and the covariance σ_{ab} of the matching problem of Equation (5) and to get $\sigma_{\tilde{x}}$ by propagating these errors in Equation (6), or whichever function we have decided to adopt to define the matching goodness. The effect of $\sigma_{\tilde{x}}$ on the estimation of $P_X(\tilde{x})$, finally, can be evaluated by direct calculation of its confidence interval $[P_X(\tilde{x} - \sigma_{\tilde{x}}), P_X(\tilde{x} + \sigma_{\tilde{x}})]$, where we simply have to count the number of favorable events for $\tilde{x} - \sigma_{\tilde{x}}$ and $\tilde{x} + \sigma_{\tilde{x}}$. The detailed formulas to implement the calculation of $\sigma_{\tilde{x}}$, again, are reported in Appendix 3.

This, however, does not fully solve the archaeoastronomical problem, because we also must consider such a similarity in a more general case, as we are going to explain better in Section 5.

4 THE PYRAMIDS OF GIZA: A CASE STUDY

Before proceeding with the discussion, we will now show the results of the proposed method

Table 1: Input data for the Giza pyramids study case. As explained in the main text, the accuracy of the stellar coordinates in 2550 BCE is $\sim 10^{-3}$ deg, thus much better than the precision, because of the error propagating in the precession and proper motion estimations.

Star (°)			Star in 2550 BCE (°)		Pyramids (°)		
Name	α	δ	α	δ	Name	λ	φ
Mintaka	83.283667	–0.285056	27.210042	–14.525222	Khufu	29.9791667	31.1341667
Alnilam	84.333583	–1.190167	28.553042	–15.007139	Khafre	29.9761111	31.1308333
Alnitak	85.468417	–0.285056	29.923917	–15.304028	Menkaure	29.9725000	31.1283333

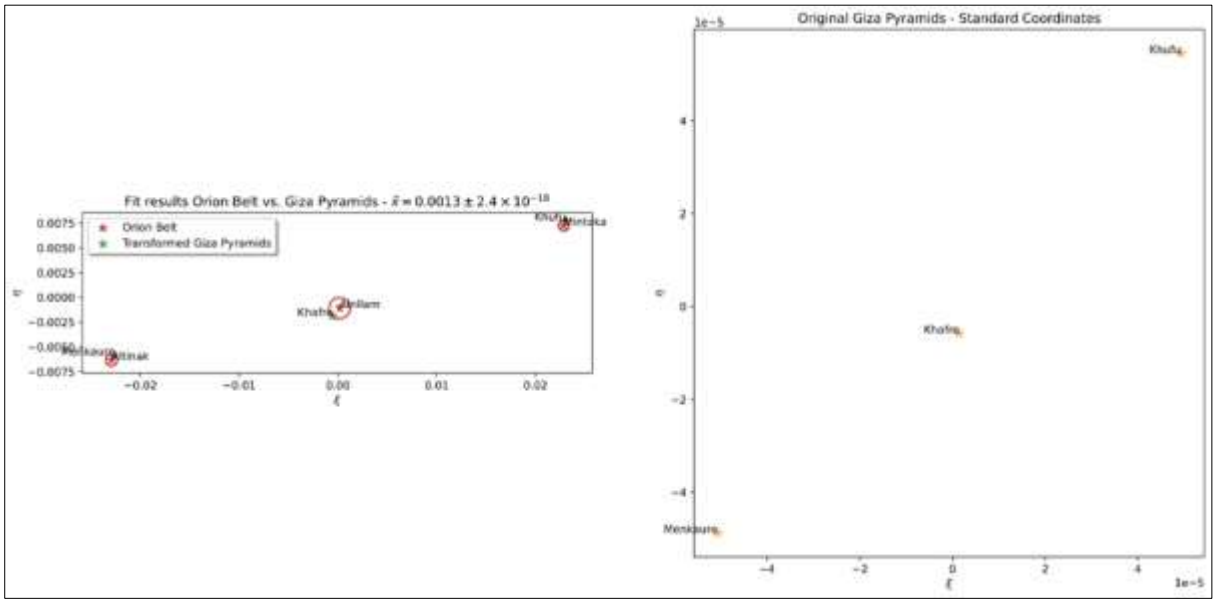


Figure 2: Results of the least-squares fit of the Giza pyramids onto the stars in Orion's Belt. The right panel shows the original positions of the pyramids, magnified to ease the comparison with the transformed objects, while the left panel contains the fit results. The red circles around the stars are placed to highlight the distance between the targets and the fitted points.

in the famous case of the above-mentioned Giza pyramids. This has to be considered as a study case, chosen because it uses just three objects, so that the Monte Carlo computation can be completed very quickly. Moreover, the scientific community generally agrees that their similarity with Orion's Belt is not intentional (Krupp, 2015), so this will be useful to gauge the results of the method for practical purposes.

Table 1 shows the input data used for the problem, namely the present coordinates of the stars in Orion's Belt, those at the approximate time of construction of the Giza Pyramids, and the corresponding coordinates of the three pyramids. For the reader's convenience, the corresponding objects are listed on the same row. As mentioned above, however, the software was 'blind' to this correspondence, which came out automatically as the configuration giving the best match. The large difference between the two sets of stellar coordinates is obviously due to precession, but we moved the stars to that epoch in order to take into account the changes in the relative positions of the three stars caused by their proper motions. We used Cartes du Ciel/SkyChart 4.2.1 (Chevalley, 2019, hereafter CDC) to evaluate the past coordinates of the stars in 2550 BCE. According to the documentation of this software, CDC uses the Extended Hipparcos Compilation (Anderson and Francis, 2012) as default catalog, and the algorithm of Vondrák et al. (2011) for the computation of the precession. These features ensure an accuracy of the order of a few arcseconds throughout the historical period, which is more than adequate for our purposes. Moreover, CDC turned out to be the most accurate program for archaeoastronomical applications in an extended comparison involving several free and commercial astronomical software packages (De Lorenzis and Orofino, 2018).

We set up a Python code to process the data and run the Monte Carlo computations. First, the co-ordinates of the Pyramids and those of the stars dated back to 2550 BCE were converted into standard coordinates using Equations (1) and (2), then the former were fitted to the latter by solving the equation system (5). The matching goodness was then computed from Equation (6), which gave $\tilde{x} = 1.3 \times 10^{-3} \pm 2.4 \times 10^{-18}$ (Figure 2).

As pointed out in the previous sections, the matching probability comes from a Monte Carlo

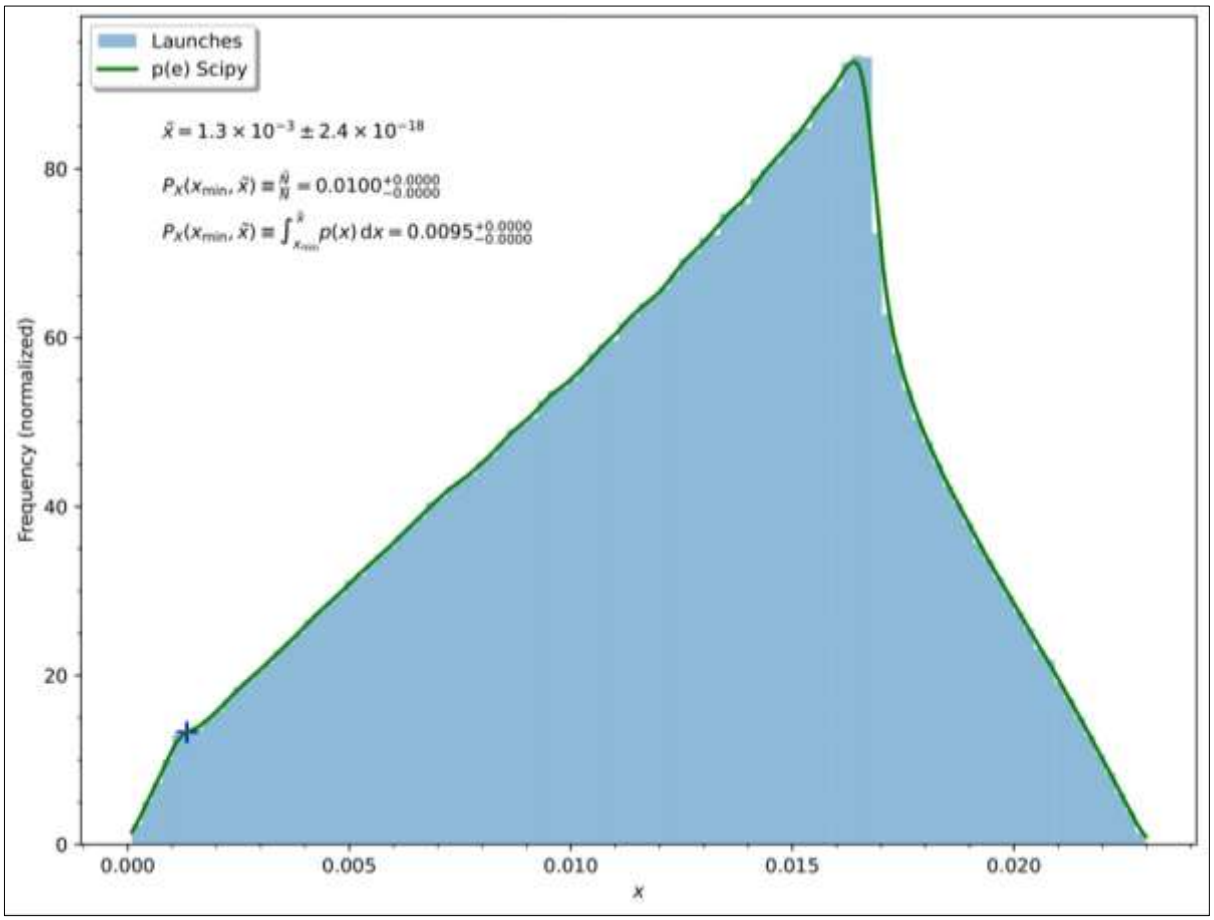


Figure 3: Results of the Monte Carlo run for the Giza Pyramids. The distribution of the matching goodness is fitted with a KDE method using Gaussian kernels, which gives a numerical estimation of the PDF (green line). The matching probability $P(\tilde{x})$ is estimated both as $\tilde{N}/N$ and as the (numerical) integral of the PDF in the range $[x_{\text{min}}, \tilde{x}]$. The contribution of $\sigma_{\tilde{x}}$ to the probability error is negligible, because of its small value. The blue cross identified the point on the PDF with abscissa $\tilde{x}$.

run that constitutes the bulk of the workflow. In this case we used 2 million launches, generated as uniformly distributed triplets of points within a circle of radius 1 and centered at (0,0). Their coordinates were interpreted as standard coordinates, and thus used to compute the matching goodness of each triplet as made for the Pyramids. The sample of $N = 2 \times 10^6$ events obtained in this way was finally used to estimate both the PDF and the matching probability as described in Section 3.4.2. For the PDF reconstruction we used the Gaussian KDE method implemented in the scipy.stats library, namely a KDE algorithm that uses Gaussian kernels.

The probability computed using Equation (9) is $P = 0.009974$, while that from the integration of the PDF results in $P = 0.009467$ (Figure 3), where for ease of notation we write P for $P_X(\tilde{x})$. As regards the accuracy of P, it is evident that the contribution of $\sigma_{\tilde{x}} = 2.4 \times 10^{-18}$ is negligible, while that of the Monte Carlo run itself can be estimated from Equations (10) or (C.4), and for $N = 2 \times 10^6$ iterations yields $\sigma_P = 7.03 \times 10^{-5}$ and $\sigma_P = 7.06 \times 10^{-5}$ with the complete and the approximate versions of the equation respectively. For the sake of completeness, we also performed an estimation following the 'Monte Carlo-plus-fit' procedure described in Section 3.5 using $M = 100$ samples with replacement for each size n, obtaining $\sigma_P = 7 \times 10^{-5}$. The complete results of the 'Monte Carlo-plus-fit' procedure are shown in Figure 4, and in Table 2.

The results of the 'simple' Monte Carlo and of the 'PDF-Monte Carlo' runs show a difference that, although small, is likely not compatible with the estimated errors. However, as is well known, the latter depends on the choice of the kernel type and its bandwidth. Further experiments, indeed, showed that the matching probability estimated with PDF reconstruction varies at the 10^{-4} level for different values of the bandwidth, and the algorithms for estimating its optimal value were quite computationally demanding. Moreover, the results of the trivial ratio of Equation (9) had the further advantage of being independent of the specific choice of the random variable, because it obviously gives the

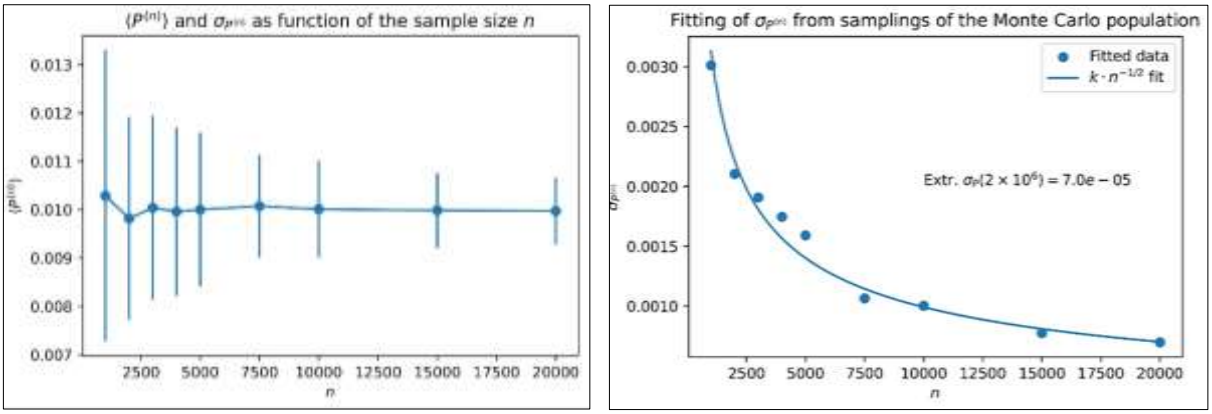


Figure 4: Results of the 'Monte Carlo-plus-fit' procedure. The left panel shows the evolution of $\langle P_i^{(n)} \rangle$ (Equation (C.6)), with the error bars reporting the corresponding $\sigma_{P_i^{(n)}}$ (Equation (C.7)). The right panel shows the fit of the estimated errors with a $k\sqrt{n}$ law and its extrapolation to $N = 2 \times 10^6$.

Table 2: Results of the 'Monte Carlo-plus-fit' estimation of $\sigma_{P_{X_N}(\bar{x})}$ compared with those obtained from Equations (C.4) (fourth column) and (C.5) (last column). The last row shows the results for $N = 2 \times 10^6$, in which the second column does not give the value of $\langle P_i^{(n)} \rangle$ for $M = 100$ samples but rather the result of the single full Monte Carlo run, and $\sigma_{P_i^{(n)}}$ is the extrapolation of the fitting from the previous rows.

n	$\langle P_i^{(n)} \rangle$	$\sigma_{P_i^{(n)}}$	$\sigma_{\bar{y}_n}$ (1)	$\sigma_{\bar{y}_n}$ (2)
1000	0.010290	0.0030128	0.0031913	0.0032078
2000	0.009820	0.0021040	0.0022049	0.0022159
3000	0.010040	0.0019071	0.0018202	0.0018294
4000	0.009963	0.0017456	0.0015703	0.0015782
5000	0.010004	0.0015904	0.0014074	0.0014145
7500	0.010075	0.0010632	0.0011532	0.0011590
10000	0.010012	0.0010015	0.0009956	0.0010006
15000	0.009988	0.0007724	0.0008119	0.0008160
20000	0.009974	0.0006954	0.0007027	0.0007062
2×10^6	0.009974	0.0000700	0.0000703	0.0000706

same results with $x = S_{\text{sq}}$ or $x = \pm\sqrt{S_{\text{sq}}}$. With the KDE algorithms, instead, a different random variable influences the actual distribution of the data, and thus the optimal kernel needed to fit them and the estimation result.

Despite these small differences, this case study gives with any of the adopted methods the quite impressive probability $P \simeq 1\%$ of a random matching between the pyramids of Giza and the stars in Orion's Belt. This seems to suggest an intentional representation, contrary to what is generally accepted by the scientific community, and therefore it deserves a further explanation.

5 MATCHING PROBABILITY AND 'COMPOUND' PROBABILITY

The proposed method allows one to estimate the probability that a given set of points can match by chance an asterism in the sky. Such a set of points, for the purposes of archaeoastronomy, can be obtained in different ways, depending on the types of artifacts under investigation. In the case of structures with sizes up to several tens or even hundreds of kilometers, in the sense discussed in Section 3.1, one can conveniently use a procedure based on the standard coordinates used in astrometry.

However, as briefly mentioned in Section 2, the probability estimated in this way is not sufficient to solve the original question of determining whether a seemingly apparent similarity can be actually the result of an intentional deed or rather just a matter of chance. To this end, in fact, one must consider that often, in the field of archaeoastronomy, the investigated artifact is just one sample among the many potentially available from a complete population of similar objects. In such a case, it can be argued that it is not unlikely to have a matching just by chance. Moreover, the similarity with a specific asterism might be a matter of chance also because there are many stars in the sky, and a set of randomly placed points, in principle, has the chance of reproducing the shape of any of the many combinations of an equal number of stars.

In practice, the matching probability described in the previous sections is akin to that of a single extraction when multiple ones are possible. In order to find the answer we are seeking

we have to estimate how the matching probability changes when considering all possible events that can generate a similar match. It is worth stressing once more that this requires considering not only the possibility of matching a specific asterism with any reasonable combination of artifacts from the complete population, but also that the matching with that specific asterism might be accidental.

This problem, as already mentioned in Cora (2018) and in Mulas and Sanna (2023), is just a standard problem of finding a compound probability. In general, in fact, the probability of a specific event is the probability that we can obtain such an event in a single attempt, or with a single 'extraction'. But if we know that we have many independent extractions available, then the compound probability theorem applies, which states that the probability of n_e independent events e_i, with $i = 1, \ldots n_e$, each having a probability P_i, is just $P = \prod_{i=1}^{n_e} P_i$.

In our case, we know the probability $P(\tilde{x})$ that one extraction gives a matching goodness $x \leq \tilde{x}$ (we recall that, for our definition, 'smaller is better') and we need to know the probability that at least one among n_e independent extractions gives such a matching. By definition, the probability that a single extraction gives a matching *worse* than $\tilde{x}$ ($x > \tilde{x}$) is $1 - P$, so, by means of the compound probability theorem (for independent events) the probability that all n_e extractions give a matching worse than $\tilde{x}$ is $\prod_{i=1}^{n_e}(1 - P) = (1 - P)^{n_e}$.

Finally, requiring that at least one extraction gives the desired matching means that the previous event must be false, so its probability is

$$P_{\text{comp}} = 1 - (1 - P)^{n_e} \tag{11}$$

We thus need to estimate how many independent extractions one can produce for the case under investigation. Giving a reliable assessment of this number, however, can easily become a quite difficult task.

In general, $n_e = n_A \cdot n_G$, where n_A is the number of possible asterisms that can be matched with the artifacts, and n_G are the number of possible groups of n objects that can be formed with all possible artifacts. As regards the latter, a better explanation is needed, and we will return to this below.

It is easier, instead, to show the possible difficulties encountered in the estimation of n_A. Indeed, one could choose a completely 'blind' criterion and say that any combination of *n* stars among the total number n_* visible with the unaided eye from a certain location on the Earth is a valid candidate for the matching. This immediately leads to

$$n_A = \binom{n_*}{n} \tag{12}$$

Such a criterion easily produces huge numbers. For example, if we consider the case of the three pyramids of Giza and take, say $n_* = 3000$, then $n_A \simeq 4.5 \cdot 10^9$. In the case of the Gothic cathedrals allegedly representing the constellation of Virgo, $n = 6$ (for the main stars only), and so $n_A \simeq 10^{18}$, while if we need to match the seven main stars of the Pleiades, then $n = 7$ and $n_A \simeq 4.3 \cdot 10^{20}$.

The above choice can be easily criticized, as it admits among the candidates also asterisms formed by stars extending over the whole celestial sphere! This clearly absurd result, however, is not useless, as it stresses the need to provide a better definition of 'eligible candidates', which makes it easier to give more reliable numbers.

The key point of the problem is to acknowledge that the question itself, namely whether a specific artifact is an intentional representation of an asterism in the sky, would not even arise were we not able to recognize a familiar asterism among a set of points. At the same time, if we can say that the asterism that we are familiar with was not known to the population that made the artifact, one can surely conclude that the similarity was due to pure chance.

In summary, the asterism must be known to the observer (either because it is part of 'our' sky or because the observer is familiar with the sky of the ancient population that made the artifact) and the asterism must belong to the set of constellations created by that civilisation. If the former condition is not met the question simply does not exist in first place, and if the latter condition is not met the matching can be safely considered pure chance.

These considerations can considerably reduce the number of candidates, however, they do not completely settle the issue about n_A, for a few reasons:

(1) Very often, we are not able to precisely identify the constellations recognised by an ancient civilisation.
(2) Our way of grouping the stars could be partially overlapping with 'their' way; as in the previously cited example of the Egyptian constellation that included stars from Orion and Lepus.
(3) The question could be posed also in the case that the artifact reproduces just in part a recognized asterism, as in the case of the Giza pyramids.

We are thus unable to give a precise estimation of n_A, which moreover could vary case

by case, but we can make some order-of-magnitude considerations.

Our sky comprises about 90 asterisms, but only 48 of the current 88 official constellations were considered by the ancient Greeks who, in turn, had largely adopted the Babylonian denominations. This is just to show how the number of eligible asterisms naturally shrinks if we go back in time, and it does not even consider the case of other civilizations, like the Chinese, or those of the other continents. In general, it is reasonable to assume that the number of eligible candidates reduces simply because of our lack of knowledge about the sky of the ancient civilisations, and as a matter of fact, we are limited to a number of very well-known cases, like the Pleiades, the Big Dipper and a few others that we know, with certainty, were widely recognized by practically all past civilizations.

On the other hand, the possibility of considering subsets of known objects can easily increase the number of possibilities that might be considered in the future because, for example, someone will notice the similarity with a specific subset that has never been noticed before. It is difficult to give a precise estimation in this sense, but it looks reasonable to set a tentative range of about $10 \lesssim n_A \lesssim 100$, where the upper limit coincides with that given by Mulas and Sanna (2023). This is certainly an improvement with respect to the many orders of magnitude of uncertainty that we started with, but such a range still leaves a broad margin of uncertainty.

We easily find similar difficulties when we try to estimate the number of groups n_G that can be eligible for the matching. As a first example, let us consider once again the problem of the Giza pyramids. There are more than 100 pyramids in Egypt (Lehner, 2008); how many groups of three pyramids would be sensible to consider as eligible matching candidates? Similarly, in the case of the number of asterism, it seems reasonable to deny the possibility of choosing any among the $\binom{100}{3} = 161{,}700$ possible combinations of three pyramids, but it is much harder, if not impossible, to restrict our choice to only a few selected groups of artifacts like we did for the asterisms. So, an estimation of n_G would most probably give a very uncertain result.

The seminal work of Mulas and Sanna (2014; 2023) can give an idea of the kind of difficulty that prevents us from giving a general criterion to find a reliable estimation of n_G. In this work, the authors claim that a particular group of nuraghes (a kind of megalithic structure erected in Sardinia, Italy, by an ancient civilization between the second and first millennium BCE) is an intentional representation of the Pleiades, intended as an asterism formed by 8 stars. In the same work it is reported that about 8000 of such monuments are known on this island, and the authors estimate $n_G \sim 8000/8 = 1000$ on the basis of the purposely simple hypothesis that these monuments are uniformly distributed. They acknowledge that this is not exactly the case, but they also noted that a different criterion, that considered 'groups of nuraghes' as a given number of these monuments relatively close to each other and well-separated by others, would surely give a number smaller than 1000, so that their choice can be regarded as a conservative one. Although it is beyond the scope of this paper to further discuss this criterion, it is worth stressing that one might also choose a kind of 'nearest neighbours' criterion to form $n_G \lesssim 8000$ groups of 8 nuraghes by choosing, for each monument, its 7 nearest neighbors. This is another arbitrary choice, and we do not claim that it is more reasonable than the previous one in general. Rather, we observe that such a choice cannot be excluded *a priori*, and that even if it might not be viable for the case of the nuraghes, it may well be admissible in other cases.

Indeed, this criterion seems much more acceptable for representations with a small number of components, just like the pyramids of Giza. So, if we blindly follow the original criterion of Mulas and Sanna (2023) we obtain $n_G \simeq 30$, while from the nearest neighbors criterion we have $n_G \lesssim 100$. The uncertainty is even larger if we again accept the idea that the only monuments relatively close to each other and well-separated by others are eligible as groups. So let us believe without any further investigation that, as in the case of the nuraghes, this gives a number smaller than 100/3. We put $n_G \simeq 10$ just to have $10 \lesssim n_G \lesssim 100$, namely the same range for the number of asterisms. Table 3 shows that, for $P = 0.009974$, $0.63 \lesssim P_{\text{comp}} \lesssim 1$. In other words, even with the most restrictive choice for n_A and n_G, the probability that the apparently impressive similarity between the three pyramids of Giza and the stars in Orion's Belt is a matter of chance is about 63%. Moreover, it increases very quickly, going up to $P_{\text{comp}} \simeq 87\%$ for $n_{A/G} = 10$ and $n_{G/A} = 20$ and so on, thus confirming the widely accepted opinion of a matching having occurred by chance.

Once again, however, dealing with an actual case of archaeological interest is beyond our scope here, and indeed we chose this specific problem just to have an already reasonably established result to test our method. Rather,

Table 3: P_{comp} for the three Giza pyramids for different values of n_A and n_G in the ranges mentioned in the text.

n_A \ n_G	10	20	30	40	50	60	70	80	90	100
10	0.633	0.865	0.951	0.982	0.993	0.998	0.999	1.000	1.000	1.000
20	0.865	0.982	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000
30	0.951	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
40	0.982	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
50	0.993	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
60	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
70	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

the main takeaway of this discussion is about the difficulty of setting a reliable confidence interval for n_A, n_G, and P_{comp}. Obviously, each specific case is a story on its own, but in general it is likely that a purely probabilistic approach would not be able to provide alone an unquestionable answer. Rather, it seems more reasonable to state that it can be one of the pieces of evidence that can be provided, from among all the other kinds needed to establish such a bold hypothesis with a better confidence.

6 CONCLUDING REMARKS

In this work we recalled the role of archaeoastronomy as a tool that can support archaeological research by means of an astronomical analysis of ancient artifacts, and at the same time we highlighted how the lack of a reliable statistical evaluation can undermine its results. We considered in particular the case of the alleged similarity between some archaeological artifacts and a few relevant asterisms on the sky that, according to some interpretations, are not accidental, but rather the result of a deliberate effort. We thus reviewed the attempts made by Cora (2018), Gaspani, (2021) and Mulas and Sanna (2023), to propose methods to decide if such intentionality, based on a statistical procedure that can estimate the probability that a specific matching between the asterism and the artifacts is due to pure chance.

We recognized that the problem belongs to the general class of the so-called PPM techniques, and at the same time we discussed the advantages and the difficulties of the proposed methods. This discussion was used to propose a different statistical procedure that determines the best matching between an asterism and its alleged on-ground representation by means of a least-squares fit. The residuals of such a fit, in other methods, are linked directly to the matching probability, whereas in our analysis we pointed out that this interpretation is at odds with the definition of probability. We thus proposed a different approach in which the residuals are used to define a measure of the similarity between the asterism and the artifacts, that we dubbed 'matching goodness', while the probability is obtained by a Monte Carlo run that reproduces the probability distribution of a random extraction of points, and is thus able to compare the actual matching goodness with that of the random hypothesis. In this way, the probability that a by-chance extraction gives a matching goodness equal to or better than that of the artifact can be estimated simply by the usual ratio between the numbers of favorable and the total extractions, or by an approximate integration of the PDF. We then pointed out that the accuracy of the probability depends on two separate factors, namely the accuracy of the Monte Carlo simulation and that of the matching goodness of the artifact $\tilde{x}$, and we showed how these can be estimated by standard techniques.

Then, we tested this procedure by applying it to the well-known case of the three pyramids of Giza, because of their resemblance to the stars in Orion's Belt, and we found that the matching probability is indeed quite small (about 1%). This result is easily explained by considering, as already pointed out in Cora et al. (2018) and Mulas and Sanna (2023), that this number represents the probability that a single random extraction can give such a matching goodness with one specific asterism. Therefore, this probability does not take into account that there are many asterisms on the sky that a random extraction could match with, and that one can usually have many artifacts of the same kind that can be considered independent extractions, each of which could reproduce the target asterism. The correct formula for the compound probability P_{comp} is therefore recalled on the hypothesis that each combination can be interpreted as an independent extraction, and its application to the Giza pyramids shows that $P_{\text{comp}} \geq 63\%$ even in the most favorable case, thus confirming the reliability of our procedure.

These final considerations led us to discuss the potential usefulness of this or any other purely probabilistic method in the context of arch-

aeoastronomical research. We conclude that it is possible and relatively straightforward to estimate the matching probability of a single extraction P with matching goodness $\tilde{x}$. However, estimating the compound probability is a much more difficult task because it depends on the possibility of giving a precise estimation of the number of asterisms that the artifact can reproduce, and of the number of extractions that can be realized with the complete population of similar artifacts. In the case of the Giza pyramids this problem, in our opinion, can be solved within a reasonable degree of accuracy, in the sense that it is possible to give a reasonable lower limit to P, but it is not possible to give a general answer, and each problem has to be considered individually.

Anyway, the method rests on a solid theoretical background, and it can be used as an additional tool in these kinds of problems, if complemented by other evidence of a different kind. Finally, it is worth adding that the Python code we used to run the Monte Carlo will eventually be made available on GitHub.

7 ACKNOWLEDGMENTS

The authors wish to thank all the referees for their valuable comments and suggestions.

8 REFERENCES

Anderson, E., and Francis, Ch., 2012. XHIP: an extended Hipparcos compilation. *Astronomy Letters*, 38(5), 331–346.

Beroiz, M., Cabral, J.B., and Sanchez, B., 2020. Astroalign: a Python module for astronomical image registration. *Astronomy and Computing*, 32, 100384.

Berriman, G.B., and Good, J. C., 2017. The application of the montage image mosaic engine to the visualization of astronomical images. *Publications of the Astronomical Society of the Pacific*, 129(975), 058006.

Chandrasekhar, S., 1943. Stochastic problems in physics and astronomy. *Reviews of Modern Physics*, 15(1), 1–89.

Charpentier, L., 1975. *The Mysteries of Chartres Cathedral*. New York, Avon.

Chevalley. P., 2019. Cartes du Ciel / SkyChart (https://www.ap-i.net/skychart/en/start; accessed 2 February 2025).

Cora, A, 2017. Il mistero delle piramidi... può essere spiegato dalla statistica. *Giornale di Astronomia*, 43(1), 27–30 (in Italian).

Cora, A., Franchi, M., and Morosin, R., 2018. *Warp Experiment, un Esperimento di Alternanza Scuola Lavoro Condotto durante Il Servizio Civile*. Pino Torinese, INAF–Osservatorio Astrofisico di Torino (*Technical Report 181*)(in Italian).

De Lorenzis, A., and Orofino, V., 2018. Comparison of astronomical software programs for archaeoastronomical applications. *Astronomy and Computing*, 25, 118–132.

Devore, J.L., 2016. *Probability and Statistics for Engineering and the Sciences* (*9th edition*). Boston, Cengage Learning.

Gaspani, A., 2021. La Presunta Corrispondenza Euristica Tra Gli Elementi Archeologici Con le Stelle e le Costellazioni. *Academia* (https://www.academia.edu/40492747/; accessed 28 September 2021)(in Italian).

Green, R.M., 1985. *Spherical Astronomy*. Cambridge, Cambridge University Press.

Groth, E.J., 1986. A pattern-matching algorithm for two-dimensional coordinate lists. *Astronomical Journal*, 91, 1244–1248.

Jacob, J.C., Katz, D.S., Berriman, G.B., Good, J.C., Laity, A.C., et al., 2009. Montage: a grid portal and software toolkit for science-grade astronomical image mosaicking. *International Journal of Computational Science and Engineering*, 4(2), 73–87.

Kolmogorov, A.N., 1950. *Foundations of the Theory of Probability*. New York, Chelsea Publishing Company.

Krupp, E.C., 2015. Archaeoastronomical concepts in popular culture. In Ruggles, C. (ed.). *Handbook of Archaeoastronomy and Ethnoastronomy*. New York, Springer. Pp. 263–285.

Lang, D., Hogg, D.W., Mierle, K., Blanton, M., and Roweis, S., 2010. Astrometry.net: blind astrometric calibration of arbitrary astronomical images. *Astronomical Journal*, 139(5), 1782–1800.

Lehner, M., 2008. *The Complete Pyramids.* London, Thames and Hudson.

Li, B., Meng, Q., and Holstein, H., 2003. Point pattern matching and applications—a review. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, *Volume 1*. Washington (DC), IEEE. Pp. 729–736.

Mulas, A., and Sanna, M., 2023. *The Nuraghes and the Pleiades: An Extraordinary Hypothesis on the Layout of a Group of Bronze Age Monuments in Sardinia*. Sardinia, Condaghes (in Italian).

Mulas, A., and Sanna, M., 2014. *In Terra Come in Cielo: I Nuraghi e le Pleiadi*. Sardinia, Condaghes.

Murtagh, F., 1992. A new approach to point pattern matching*. Publications of the Astronomical Society of the Pacific*, 104, 301–307.

Newcombe, R.G., 1998. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in Medicine*, 17(8), 857–872.

Oberle, W., 2015. *Monte Carlo Simulations: Number of Iterations and Accuracy*. US Army Research Laboratory (Technical Report ARL-TN-0684).

Pál, A., and Bakos, G.Á., 2006. Astrometry in wide-field surveys. *Publications of the Astronomical Society of the Pacific*, 118(848),1474–1483.

Parzen, E., 1962. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3), 1065–1076.
Polcaro, V.F., 2016. The credibility of archaeoastronomy: a suggestion from pharmacology*? Mediterrane–an Archaeology and Archaeometry*, 16(4), 1–5.
Rosenblatt, M., 1956. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3), 832–837.
Shaltout, M., and Belmonte, J.A., 2005. On the orientation of ancient Egyptian temples: (1) Upper Egypt and Lower Nubia. *Journal for the History of Astronomy*, 36(124), 273–298.
Sinclair, R.M., 2006. The nature of archaeoastronomy. In Bostwick, T.W. and, Bates, B. (eds.). *Oxford VII International Conferences on Archaeoastronomy: Viewing the Sky Through Past and Present Cultures*. Phoenix, Parks, and Recreation Department (Pueblo Grande Museum Anthropological Papers). Pp. 13–26.
van Wamelen, P.B., Li, Z. and, Iyengar S.S., 2004. A fast expected time algorithm for the 2-D point pattern matching problem. *Pattern Recognition*, 37(8), 1699–1711.
Vecchiato, A., 2017. *Variational Approach to Gravity Field Theories*. Cham (Switzerland), Springer.
Vondrák, J., Capitaine, N., and Wallace, P. 2011. New precession expressions, valid for long time intervals. *Astronomy and Astrophysics*, 534, A22.
Wald, A., 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54(3), 426–482.
Wallis, S.A., 2021. *Statistics in Corpus Linguistics: A New Approach*. New York, Routledge.
Wang, X., and Zhang, X., 2012. Point pattern matching algorithm for planar point sets under Euclidean Transform. *Journal of Applied Mathematics*, 2012, 139014.
Yanoff, M., and Duker, J.S., 2009. *Ophtalmology*, *Third Edition*. Amsterdam, Elsevier.

9 APPENDIX 1: SHAPE DEFORMATIONS INDUCED BY STANDARD COORDINATES

An approximate idea of the deformations induced by the projection to standard coordinates, namely of the differences $\Delta\alpha$ between the angles between the sides of the original figures and those of the projected one, can be given by considering the so-called *spherical excess* E of a spherical triangle. We know that the sum of the internal angles of a flat triangle is 180°. That of a spherical triangle, instead, is larger, and the amount exceeding 180° is called spherical excess. The Gauss–Bonnet theorem states that the spherical excess of a spherical triangle is equal to the solid angle Ω_T subtended by this figure, and thus is proportional to the size of the figure. More precisely, if $\alpha_i, i = 1,2,3$ are the internal angles,

$$E \equiv \sum_{i=1}^{3} \alpha_i - \pi = \Omega_T = \frac{A_T}{R^2} \tag{A.1}$$

where A_T is the area of the spherical triangle and R the radius of the sphere. L'Huilier's theorem, instead, gives E as function of the angular length of the sides, and in the case of an equilateral triangle with side a

$$\tan\left(\frac{E}{4}\right) = \sqrt{\tan\left(\frac{3a}{2}\right)\tan^3\left(\frac{a}{4}\right)} \tag{A.2}$$

In this case the difference $\Delta\alpha_i$ between any α_i and its flat counterpart is clearly $E/3$, and for small spherical triangles $a \ll 1$ and $E \ll 1$, so a simple Taylor expansion gives

$$\Delta\alpha_i \simeq \frac{a^2}{3}\sqrt{\frac{3}{8}} \tag{A.3}$$

10 APPENDIX 2: MATHEMATICAL DESCRIPTION OF THE MATCHING METHOD

As explained in the main text, we have 2 sets with n 2D points each: $S = \{\mathrm{P}_i \equiv (x_i, y_i), i = 1, \dots, n\}$ (dubbed 'representation') and $\bar{S} = \{\bar{P}_i \equiv (\bar{x}_i, \bar{y}_i), i = 1, \dots, n\}$ ('target'), and we want to find a transformation that produces the best matching between S and $\bar{S}$. We make three basic assumptions:

(a) The two lists are already ordered by correspondence points, i.e., the transformation sought must bring the ith point of S onto the ith point of $\bar{S}$.
(b) The points of the two sets belong to an intrinsically flat surface.
(c) The transformation between S and $\bar{S}$ must be conformal.

As is clear in the following, the first assumption allows solving the problem with standard least-squares techniques.

The second assumption is simply a matter of having a simpler mathematical treatment, and in our case, where the two sets of original points belong to two intrinsically curved surfaces (namely the Celestial Sphere and the Earth Ellipsoid) such an assumption is guaranteed by their prior transformation into standard coordinates by means of Equations (1) and (2). Mathematically, this translates into the problem of finding a function depending on m parameters, $T(\cdot; \{p_l\}), l = 1, \dots, m$, such that:

(1) $\tilde{\mathrm{P}}_i = T(\mathrm{P}_i; \{p_l\})$. $\tilde{S} = \{\tilde{P}_i\}$ can be called the 'transformed representation'.

(2) It minimizes the sum of the squares of the residuals between $\{\bar{P}_i\}$ and $\{\tilde{P}_i\}$, namely $\sum_{i=1}^{n}(\tilde{x}_i - \bar{x}_i)^2 + (\tilde{y}_i - \bar{y}_i)^2$ is the smallest possible sum.

It is worth observing that minimizing the residuals as described in point 2, implies that also the sum of the Euclidean distances $d_i = \sqrt{(\tilde{x}_i - \bar{x}_i)^2 + (\tilde{y}_i - \bar{y}_i)^2}$ between the transformed representation and the target are minimized by the transformation.

The third assumption implies, as explained in the main text, that $T(\cdot;\cdot)$ is obtained at most by the combination of:

1. A scale transformation with a unique scale factor s in common between the two axes
$$\begin{pmatrix}\tilde{x}\\ \tilde{y}\end{pmatrix} = s\begin{pmatrix}x\\ y\end{pmatrix} \tag{B.1}$$
2. A rotation $R(\theta)$ of an angle θ around the origin
$$\begin{pmatrix}\tilde{x}\\ \tilde{y}\end{pmatrix} = R(\theta)\begin{pmatrix}x\\ y\end{pmatrix} = \begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix} \tag{B.2}$$
3. A translation $t \equiv (t_x, t_y)$
$$\begin{pmatrix}\tilde{x}\\ \tilde{y}\end{pmatrix} = \begin{pmatrix}x + t_x\\ y + t_y\end{pmatrix} \tag{B.3}$$
4. A parity transformation that changes the sign of the two coordinates.

In this way, the resulting transformation is conformal, i.e., it does not change the angles between any pair of sides of the figure that can be obtained by connecting the points of the sets.

As explained in the main text, here we neglect the parity transformation, therefore it can be shown that the general transformation between (x, y) and $(\bar{x}, \bar{y})$ can be written as a set of two equations, linear with respect to the 4 parameters a, b, c, d:

$$\begin{pmatrix}\tilde{x}\\ \tilde{y}\end{pmatrix} = \begin{pmatrix}ax - by + c\\ bx + ay + d\end{pmatrix} \tag{B.4}$$

where

$$\begin{pmatrix}a\\ b\end{pmatrix} = s \cdot R(\theta) \tag{B.5}$$

$$\begin{pmatrix}c\\ d\end{pmatrix} = \begin{pmatrix}t_x\\ t_y\end{pmatrix} \tag{B.6}$$

Obviously, in general there exist infinite exact solutions of the equation system (B.4) for a single pair of points *P* and $\bar{P}$; for 2 points (four equations) we have at most one solution, and for more than 2 points there is no exact solution. In the latter case, a least-squares fitting of the equation system provides the estimation of a set of parameters that satisfies the minimization criterion of point 2.

Moreover, the two sets are first transformed into standard coordinates, thus satisfying the requirement to have a flat manifold, and the use of the barycenter of the two sets as the projection point allows us to neglect the translation without introducing serious discrepancies.

With this expedient we assume $t_x = t_y = 0$, and the equations of the transformation then reduce it to

$$\begin{cases}\tilde{x} &= xs\cos\theta - ys\sin\theta\\ \tilde{y} &= xs\sin\theta + ys\cos\theta\end{cases} \tag{B.7}$$

where, as in Equation (B.4), we can make the equations linear by putting $a = s\cos\theta$ and $b = s\sin\theta$. In this way a set of *n* points $P_i = (x_i, y_i)$ produces the system of equations

$$\begin{cases}\bar{x}_1 &= x_1 a - y_1 b\\ \bar{y}_1 &= x_1 b + y_1 a\\ \vdots & \quad\vdots\\ \bar{x}_n &= x_n a - y_n b\\ \bar{y}_n &= x_n b + y_n a\end{cases} \tag{B.8}$$

which, as already pointed out, has to be solved in the least-squares sense. Such a system can be written in matrix form as $\mathbf{b} = A\mathbf{x}$, where $\mathbf{b}^{\mathrm{T}} = (\bar{x}_1, \bar{y}_1, \ldots, \bar{x}_n, \bar{y}_n)$, $\mathbf{x}^{\mathrm{T}} = (a, b)$, and

$$A = \begin{pmatrix}x_1 & -y_1\\ y_1 & x_1\\ \vdots & \vdots\\ x_n & -y_n\\ y_n & x_n\end{pmatrix}$$

Obviously, we can recover the values of s and θ from those of a and b by inverting Equation (B.5), which gives

$$s = \sqrt{a^2 + b^2} \tag{B.9}$$

$$\theta = \tan^{-1}\left(a/\sqrt{a^2 + b^2}, b/\sqrt{a^2 + b^2}\right) \tag{B.10}$$

11 APPENDIX 3: CALCULATION OF THE MONTE CARLO ACCURACY

The matching goodness of random launches, as mentioned in the main text, can be regarded as the population $\mathcal{X}$ of a random variable $x \in \mathbb{R}$ defined in the range Δx of the values admissible for $\mathcal{X}$. It thus admits a true value of the probability that $x \leq \tilde{x}$ for a specific value $\tilde{x} \in \Delta x$, which we indicate with $P_{\mathcal{X}}(\tilde{x})$.

A Monte Carlo simulation, by extracting a sample X_N N elements from $\mathcal{X}$, provides an approximation $P_{X_N}(\tilde{x})$ of the probability $P_{\mathcal{X}}(\tilde{x})$. The larger the number N of launches is, the more accurate the representation is, namely

$$P_{\mathcal{X}}(\tilde{x}) = \lim_{N\to\infty} P_{X_N}(\tilde{x}) \tag{C.1}$$

It is well known that the Central Limit Theorem (CLT) can be used to estimate how accurately $P_{X_N}(\tilde{x})$ estimates $P_{\mathcal{X}}(\tilde{x})$ (Devore, 2016). Indeed, the result of the actual Monte Carlo run is the outcome of just one among the infinite possible samplings of size N of the total population, and by the CLT it can be proved that

$$P_{X_N}(\tilde{x}) = \tilde{\bar{y}}_N \tag{C.2}$$

where $\tilde{\bar{y}}_N$ is one extraction from a population $\bar{\mathcal{Y}}_N$ of a random variable $\bar{y}_N$ whose distribution has mean $\mu_{\bar{y}_N} = P_{\mathcal{X}}(\tilde{x})$ and standard deviation (Oberle, 2015)

$$\sigma_{\bar{y}_N} = \sqrt{\frac{P_{\mathcal{X}}(\tilde{x})\left(1 - P_{\mathcal{X}}(\tilde{x})\right)}{N}} \tag{C.3}$$

Moreover, the CLT states also that the larger the value of N is, the better the distribution of $\bar{\mathcal{Y}}_N$ approximates a normal distribution.

Therefore, as an element of the population $\bar{\mathcal{Y}}_N$, $\tilde{\bar{y}}_N = P_{X_N}(\tilde{x})$ lies with 68% probability in the range $\left[P_{\mathcal{X}}(\tilde{x}) - \sigma_{\bar{y}_N}, P_{\mathcal{X}}(\tilde{x}) + \sigma_{\bar{y}_N}\right]$. This is equivalent to saying that the accuracy of $\tilde{\bar{y}}_N = P_{X_N}(\tilde{x})$ as an approximation of $P_{\mathcal{X}}(\tilde{x})$ is $\sigma_{\bar{y}_N}$.

We cannot compute directly $\sigma_{\bar{y}_N}$ because we have just one element of the population $\bar{\mathcal{Y}}_N$, however, if we make the approximation $P_{\mathcal{X}}(\tilde{x}) \simeq P_{X_N}(\tilde{x})$, we can write from Equation (C.3)

$$\sigma_{P_{X_N}(\tilde{x})} = \sigma_{\bar{y}_N} \simeq \sqrt{\frac{P_{X_N}(\tilde{x})\left(1 - P_{X_N}(\tilde{x})\right)}{N}} \tag{C.4}$$

which, if $P_{\mathcal{X}}(\tilde{x}) \ll 1$, and recalling Equation (9), becomes

$$\sigma_{P_{X_N}(\tilde{x})} = \sigma_{\bar{y}_N} \simeq \sqrt{\frac{P_{X_N}(\tilde{x})}{N}} = \frac{\sqrt{\tilde{N}}}{N} \tag{C.5}$$

thus giving the Eq. (10) estimation of the accuracy attained by the Monte Carlo run after N steps.

The other way to estimate the accuracy of a Monte Carlo run by performing M times a Monte Carlo simulation with $n < N$ iterations, is similarly justified by the CLT. This theorem, in fact, states that each $P_i^{(n)}$ belongs to an approximately normal distribution with mean $\mu_{\bar{y}_n} = P_{\mathcal{X}}(\tilde{x})$ and standard deviation given by Equation (C.4). The M Monte Carlo runs are thus a sampling of such a distribution for n iterations, so we can estimate $\mu_{\bar{y}_n}$ and $\sigma_{\bar{y}_N}$ by the actual average and standard deviation of the sample

$$P_{\mathcal{X}}(\tilde{x}) = \mu_{\bar{y}_n} \simeq \langle P_i^{(n)} \rangle = \frac{1}{M}\sum_{i=1}^{M} P_i^{(n)} \tag{C.6}$$

$$\sigma_{P_{X_N}(\tilde{x})} = \sigma_{\bar{y}_n} \simeq \sigma_{P_i^{(n)}} = \frac{1}{M}\sum_{i=1}^{M}\left(P_i^{(n)}\right)^2 - \left(\langle P_i^{(n)} \rangle\right)^2 \tag{C.7}$$

Finally, for the sake of completeness, we recall in more detail the standard way of computing the accuracy of the least-squares estimation of $\tilde{x}$ that contributes, together with that of the Monte Carlo distribution, to the overall accuracy of the estimation of $P_{\mathcal{X}}(\tilde{x})$.

It is well known that the quantities σ_a, σ_b, and σ_{ab} of the matching problem of Equation (5) represented in matrix form as $\mathbf{b} = A\mathbf{x}$, are the elements of the normal matrix $\left(A^{\mathsf{T}}A\right)^{-1}$ except for a factor σ_0. This factor, in the absence of an external characterization, can be estimated from the residuals $r_i = \left(b_i - \tilde{b}_i\right)$ of the solution of the equation system as

$$\sigma_0^2 = \frac{1}{N-M}\sum_{i=1}^{N}\left(b_i - \tilde{b}_i\right)^2 \quad \text{(C.8)}$$

with N the number of equations of the system and M the number of estimated parameters. In our case $N = 2n$ and $M = 2$, thus, following the notation of Section 3.2,

$$\sigma_0^2 = \frac{1}{2(n-1)}\sum_{i=1}^{N}\left(\tilde{\xi}_i - \bar{\xi}_i\right)^2 + (\tilde{\eta}_i - \bar{\eta}_i)^2 \quad \text{(C.9)}$$

Dr. Alberto Vecchiato is a researcher at the Astrophysical Observatory of Turin, which is part of the Italian National Institute for Astrophysics (INAF). His research interests are in the area of astrometry and gravity studies, with particular emphasis on the reduction of global astrometric data from space satellites, and on high-precision testing of gravity theories, and in archaeoastronomy.



He has been working for about twenty years on the ESA Gaia Mission, and is also interested in the field of the education and popularization of astronomy. He is the author of several scientific publications, of the textbook *Variational Approach to Gravity Field Theories*, and co-author with Gabriella Bernardi of the book *Understanding Gaia: A Mission to Map the Galaxy*, both published by Springer. He is a member of the Internatonal Astronomical Union (IAU), of the Società Italiana di Archeoastronomia (SIA), of the Società Astronomica Italiana (SAIt), and of the Italian Society of Gravitation (SIGRAV).



Gabriella Bernardi is a writer and science journalist associated with the INAF (National Institute of Astrophysics) in Italy. She graduated in Physics from the University of Turin, and since 2008 has been a member of the European Union of Science Journalists' Association (EUSJA). She is a member of IAU Commission C3 (History of Astronomy) and also serves on the IAU Executive Committee Working Group on 'Women in Astronomy'.

She has published in several newspapers and magazines specializing in the popularization of astronomy and is the author of six astronomy books, three of which were published by Springer, including *The Unforgotten Sisters: Female Astronomers and Scientists before Caroline Herschel*; *Giovanni Domenico Cassini: A Modern Astronomer in the 17th Century*. Gabriella has also contributed to *Domestic Astronomy in the Seventeenth and Eighteenth Centuries* in the Palgrave series on 'History of Women and Science since 1660'.



Alberto Cora was born in Turin (Italy) in 1963. A former amateur astronomer he completed a BSc in Physics at Turin University in 2004. In 1989 he began work at the Astrophysical Observatory of Turin (INAF-OATo) and collaborated in the development of the 5-channel photopolarimeter. Since 1995 he has been working in the Solar Physics Group at the Observatory participating in observations of the solar corona with the UltraViolet Coronagraph Spectrometer (UVCS) on board on the SOlar Heliosperic Observatory (SOHO). He has also developed the SOho Long-term data ARchive (SOLAR) at Turin Observatory.

Alberto also was in charge of public educational and outreach activities at the Observatory from 2011 to 2018, and from 2015 to 2017 he participated in drafting Piedmont regional law 3/2018 to control light pollution. He is a member of the Società Italiana di Archeoastronomia and of the IAU.