Workshop: Archaeology and Geometric Morphometrics

Abstract

These two workshops are designed to provide an introduction into the application and potential of geometric morphometric (GMM) methodologies for Japanese archaeologists, researchers and enthusiasts, through considering both theoretical and practical components.

The first session will introduce the reader to the fundamental mathematical underpinnings of statistical shape and form, in a digestible and easy-to-follow manner, before detailing the fundamentals of GMM, emphasising their statistical power and coverage in comparison to traditional morphometrics.

In the second session, we will showcase one popular method of examining artefact shape: outline analysis through Momocs (for the R Environment). Using a published dataset, this workshop will showcase the GMM procedure, from outline extraction, collation and transformation, to the analytical workflow (including data visualisation and advanced techniques).

All resources including the presentation, data and code will be available on a GitHub repository ([www.github.com/CSHoggard/-workshopjapan2020](http://www.github.com/CSHoggard/-workshopjapan2020)).

**Instructor**: Dr. Christian Steven Hoggard (University of Southampton, United Kingdom)

**About the Instructor**: Dr. Christian Steven Hoggard is a Visiting Fellow based at the University of Southampton. Christian completed his PhD focusing on Neanderthal technological variability, at the University of Southampton, before undertaking a two-year postdoctoral position at Aarhus University, Denmark. Christian specialises in computational archaeology, specifically, two-dimensional geometric morphometric methodologies, phylogenetic approaches to archaeology, and cultural evolution. Christian has experience working on other artefact beyond lithics, including Bronze Age metalwork, Medieval brooches, and his most recent project (A Joint Effort) which examines differences in biological sex through patella (kneecap) three-dimensional morphology.

Session One: Basics and Research History (45-60 minutes)

*//* : slide break

*Introductory remarks: welcome (format for this session; Q&A component)*

Slide 1: Well good morning (evening) everyone, my name is Dr. Christian steven Hoggard and welcome to the first of two workshops, introducing shape-based approaches to archaeological analyses. Over the next few hours I will introduce you all to the notion of shape analysis, how we as archaeologists catalogue shape variation, visualise this variation, and how we can use geometric morphometrics to understand different aspects of our shared human past. I will detail the geometric fundamentals, the advantages and disadvantages of geometric morphometric methodologies, the sort of questions we can ask, but most importantly detail how we “do” geometric morphometrics. This will lead into our second workshop, where we get “hands-on” with software and conduct some geometric morphometric analyses in the R Environment. If you haven’t found them already then all resources, including a script of this talk (in both English and Japanese) will be found on my GitHub repository ([www.github.com/CSHoggard/-workshopjapan2020](http://www.github.com/CSHoggard/-workshopjapan2020)).

Slides 2/3: We will be using Zoom, Slack and a Google Document throughout the course of this workshop. The Zoom chat will be for immediate reactions and instant feedback. Slack and Google Documents will be where you can ask questions about anything we do throughout this workshop. There will be two ten minute breaks for questions to be proposed and I will answer them following the break. I do wish to highlight from the beginning that there are no “wrong” questions, we’re all intellectually curious and we all do not know everything. So please, do speak up should you have any questions and myself, Atsushi and others will help accommodate. // And, as you are all familiar, here is an example of the Slack environment

Slide 4: So, let’s begin. We all are aware that shape is fundamental to our everyday lives, from the cars we drive, to the kitchen accessories we buy, from the sofa we sit on, and to the bed we lie in, shape is central to our material life. It plays a role in how items can function, how they can be perceived and considered, and can even hint at their possible origins or makers. It is therefore unsurprising that shape has been paramount to how archaeologists comprehend and understand past societies, from our earliest ancestors to historical and contemporary societies. And in every exercise there is usually a necessity to quantify and describe different shapes, and through the data collected, visualise, explore and analyse trends in shape, in order to test suitable hypotheses or models. Recently, in the last two decades, archaeologists have changed how they record and understand the shapes of different artefacts, utilising techniques grounded in the discipline of geometric morphometrics (or GMM).

Slides 5/6: But before we discuss GMM, we need to know what shape “actually” is. It is perhaps actually quite difficult to define. We talk about shapes everyday, describing their components with an almost universal vocabulary: square, round, wiggly or pointy for example. However, while we prescribe different objects to different shapes, it is difficult to define what shape *actually* is. // Perhaps the best definition of shape, or the most concise and agreed upon definition, is that by statistician Prof. Christopher Small, who defines shape as “*the total of all information that is invariant under translation, rotations and isotropic scalings*” (Small, 1996: 6). Let’s break this down. What Small is saying here is that shape is all information that does not change whereabout on the page or screen it is, does not change however you rotate it, nor however you resize it. So, six equilateral triangles will have the same shape, as too will six circles, spheres or cubes. In a perfect scenario, they are all the same no matter how you transform them. But as Small (1996) also notes, two objects rarely contain the same information (within a degree of measurement error), with the archaeological record best exemplifying this.

Slides 7/8: In using the above definition, it is important to note that two objects can have the same *shape* but be of different size, a component also of interest to archaeologists: but size can also be quantified and perceived in different forms. There is the classic question of which is bigger for example: a snake, a hippopotamus or a giraffe? We can view size in a variety of ways: 1) a lineal measurement (such as length), 2) a calculated mass, volume or weight, or 3) its centroid size (which we will briefly talk more about later). // So when we are discussing shape we are talking about shape, but when we include size we’re talking about the *form* of an object.

Slide 9: All three variables (*form*, *shape* and *size*) can be quantified, analysed and understood through the discipline of morphometrics. First coined by Professor of Zoology (University College Dublin) Robert Blackith in 1957, morphometrics is quantitative study of size, shape and their variance (or co-variance). For morphometricians, there are commonly two different branches of study: 1) traditional morphometrics and 2) geometric morphometrics. Traditional morphometrics focus on the study of lineal measurements (lengths, widths, angles, ratios and indices), in isolation and in conjunction, through the analysis of scattergraphs or through more complicated analyses e.g. Euclidian Distance Matrix Analysis (EDMA). These methods are deeply rooted in the history of archaeology, providing the basis of artefact classification (along with typology) for over half a century: for the Palaeolithic one can think of the work or Drs. Francois Bordes or Derek Roe concerning handaxe techniques as an example. There are however a number of issues with traditional morphometric approaches. Traditional methodologies are often subjective, and often erroneous in their precision: measurement error can be introduced from a variety of different variables include the equipment, the observer, the orientation, tiredness, etc. GMM on the other hand, focuses on landmarks, outlines and surfaces.

Slide 10) With GMM, points of correspondence can be digitised, automated (if required) and reduce the scope for error. Perhaps the greatest advantage of GMM is their coverage, being able to extract significantly greater levels of shape information in comparison to lineal measurements: points of correspondence, outlines and surfaces can be considered, providing greater topographic information than a few directional measurements. Now, that is not to say that lineal measurements are bad; most often, your hypotheses and models can be considered through lineal morphometrics, but in terms of coverage and the resolution of information then GMM is more advantageous (Refer to Slide 9). Other advantages include the ease with which shape can be extracted: as we may only be considering the shape of an artefact we do not need scaled images and photographs (oriented the same way) can (but not ideally) suffice.

But, with GMM a necessary skill-level which is required, with specialised software packages and a specific vocabulary. Hopefully, by the end of this workshop (and/or the end of the next workshop) this should not be an issue.

Slide 11) Similarly to lineal measurements we can use GMM to explore a number of scenarios. We can:

* Determine on an assemblage level the mean and median shapes, and the distribution of shape variance (i.e. is there a gradual or abrupt change in artefact shape);
* Examine whether two or more artefacts and assemblages are different in terms of their shape
* Investigate whether these assemblages differ in their form (shape plus size);
* Consider how the shape of artefacts relate to size and development (allometry) or other factors e.g. raw material, hominin species, historical period;
* Determine whether differences in the shape of assemblages or artefacts correspond to a particular model of hypothesis e.g. shape change throughout time due to various factors;
* And/or produce network-based models of artefact production (through cluster or phylogenetic analyses of shape).

Slides 12/13/14/15/16: Before we go into the fundamentals of GMM (landmark geometry etc.) it is important to consider the historical perspective, how we got to where we are today, and the 600-year journey to what we are discussing today. // Perhaps the foundation for GMM can be accredited to painter, printmaker and theorist *Albrecht Dürer (1471-1528).* Dürer is known, in addition to providing the first illustrations of the rhinoceros to western audiences, and for his work in descriptive geometry, detailing the intricacies of complex three-dimensional shapes including helices, conchoids and epicycloids. He’s also known for investigating what we can call the Delian Problem, or “double the cube” problem, where given the edge of a cube you’re asked to determine how to double the volume of said cube. For us, Dürer is important for our discussions on GMM in how he transformed and acknowledged shape, expressing changes in anatomy through a grid-based cartesian system. This became, what we now use as, one of the main descriptors of cataloguing landmark transformation today, and a powerful means of describing shape change. // If we fast-forward three hundred years the next great leap in GMM can be seen through the work of the Scottish biologist, mathematician and classics scholar *Sir D’Arcy Wentworth Thompson (1860-1948)*. Thompson's opus was *On Growth and Form*, an elegantly written book which demonstrated that biological form can reflect physical and mathematical principles. Perhaps the most famous images from *On Growth and Form* are his transformations which, similarly to Dürer, demonstrated that variation in form between species. Importantly, Thompson identified that this shape change could be modelled by consistent deformation. It's this consistency, which is essential here, D'Arcy demonstrated that these specimens stretched one particular pattern. // Note the use of landmarks on these species, and their deformation which provided us with a graphical basis for cataloguing shape. I would urge you to try and find a copy of his book; while Thompson was critical of Darwinism, I would urge you to buy the book for the inspiring language Thompson adopts (which inspired greats such as Huxley, Turing, the epigeneticist Conrad Waddington, Corbusier, and others), and for the mathematical diagrams D’Arcy designs in the later chapters. // If we fast-forward to the last fifty-to-sixty years then the main developments in GMM are in providing a statistical means for cataloguing shape change, this includes the work of Fred Bookstein (the founder of modern-day GMM), the late Dennis Slice, Miriam Zelditch, James Rohlf. Much of their work is what will follow.

Slide 17: So at this point we will have a 10 minute break, if you have any questions please note them down in Slack and the Google Document, and once we’re ready to begin we shall continue.

Slides 18/19: So, following the questions we propose, or the topic we wish to investigate, how do we “do” geometric morphometrics? // Well, first we need to create a catalogue of specimens with which our data can come from. This can be from three-dimensional methodologies, including CT scanning, photogrammetry or structure-from-motion (which I’m a keen enthusiast for), scanners (such as the NextEngine) or microscribes. As highlighted previously, we can analyse shape from two-dimensional formats, drawings and photographs are also applicable here. It’s important in any technique to consider the error associated with each technique; I’ve recently submitted a paper on the error associated with lithic illustrations, and demonstrated that such error is negligible irrespective of illustrative capability, but it would be up to you to account for errors in your own study.

Slide 20/21/22: Once we have collated our dataset, our models, our images etc. it’s time to collate the data which we call shape. As we’ve previously hinted (and as shown in Wentworth Thompson’s work), landmarks (points of morphological correspondence) are central here. A landmark is a coordinate point used to represent a homologous point on a structure. These are quantifiable as cartesian coordinates (in an x/y or x/y/z system) and can be analysed as individual points, or in combination, as surfaces. // It is generally accepted by the morphometric community that there are three different landmark types *sensu* Bookstein (1991). There are first, Type I landmarks, localised mathematical points, defined by an obvious biological structure. These are easy to identify repetitively such as the intersection of specific bones or specific features on bones e.g. posterior tip of a nasal cavity. These are what would be considered the ‘best’ type of landmarks to use where possible. Type II landmarks are mathematical points defined in geometry; these are not biological in nature and can reflect points such as maximum curvature. Type III landmarks, on the other hand, are mathematical points defined with reference to another point. If we consider this rodent jaw (see presentation) we can see all three types of landmarks. We can see Type I, represented by the central part of the tooth, or the change in bone structure to the right of the diagram, Type II landmarks as represented by maximum curvatures on the mandibular hinge and apex at the top, and Type III as where the shape intersect with these Type II landmarks. There are also semilandmarks, landmarks which are placed using an algorithm around an object or between two end-points. For example, we could catalogue a handaxe shape as represented as fifty equidistant (evenly-spaced) points. These would be a special Type III landmark. And here we can see an example (see presentation) of this on archaeological grains, with an algorithm placing semilandmarks (here in hollow) around the edge of the shape. // In their quantity and coverage, (semi-)landmarks:

* Should sample aspects of the shape which are of archaeological interest;
* Should be repeatable and identifiable on all examples (if possible);
* Should cover as much of the shape as possible;
* Should be sufficient as to not increase the weighting of specific areas.

Slide 23: Irrespective, landmarks (and I cannot emphasise this enough) have to be plotted in the same order, or else subsequent transformations will not consider the correct parts of the shape. For bioarchaeologists, there are a greater number of points of morphological points of correspondence (skulls have explicit landmarks and are of known orientation), permitting easy three-dimensional shape studies. For archaeologists studying non-biological material greater creativity and thought is needed in orienting specimens and placing landmarks; we have fewer case studies in peer-reviewed journals to compare to too, so it’s difficult to understand what is correct or deemed suitable. Irrespective of what type of archaeology you study issues of fragmentation and sample size (as a representation of its overall population) are important considerations when conducting your study.

Slide 24: To provide one example of where greater creativity is necessary, here (see presentation) is a study I conducted with a colleague last year (and published in the *Danish Journal of Archaeology*), detailing Nordic Bronze Age “tutuli”, small metal circular places with a protruding tip. We were interested in the appropriateness of classifications used to catalogue their variability (as certain pieces are thought to represent certain periods). We had to overcome issues of degradation, but more interestingly (to me) was the fact that this was novel, no other similar study existed and so we had to devise a framework for cataloguing shape variation. We therefore used six landmarks at points of convexity and shape change, and placed semilandmarks between these landmarks. This allowed for an analysis of tutuli cross-section and differences in tutuli type (see presentation).

Slide 25) At this stage, does anyone have any questions regarding landmarks? We’ll have plenty of time for discussion at the end too, if people do not wish to ask questions now.

Slide 26/27) Great, let’s now consider how we would digitise these landmarks. There are a variety of different open-source and proprietary software available, depending on whether you are conducting two- or three-dimensional shape analysis. Personally, if I was doing two-dimensional landmark analysis, I would use the TPS Suite for plotting landmarks, there are loads of instructions and guides for this software online. But this is because I first got introduced to GMM through the TPS Suite. A variety of other better packages are available. // With an increasing number of packages for creating, manipulating and analysing shape coordinates, in addition to a more code-literate discipline, R is the ideal environment for archaeologists. The packages *Geomorph* and *Momocs*, and the features in *shape*, are perhaps the most widely used packages, but they differ depending on what type of analysis you wish to conduct: *Geomorph* is more for three-dimensional biological analyses while *Momocs* focuses on two-dimensions.. The next workshop will focus on the *Momocs* package which is designed for two-dimensional outline analysis (a method we’ll come to soon…). For those who are R literate already, I prefer Momocs because of its compatibility with the *Tidyverse* (a commonly used suite of packages for data manipulation, transformation, modelling and visualisation. It is important to remember that you can always mix and match too, so you can generate data in the TPS Suite and import the data to R for analysis (something I regularly do).

Slide 28) So we have covered quite a tricky component of geometric morphometrics, so we’ll have another ten minute break, if you have any questions please note them down and I’ll answer them after the break.

Slide 29) Once you have digitised landmarks you will need to check your data (either as a notepad file or through R). Ask yourself four questions: 1) Do all of my specimens have the correct number of landmarks? 2) Are all my landmarks in the correct order? 3) Are the ID labels correct (these you amend afterwards), 4) Are they to scale? (If you’re conducted size-integrated analyses).

Slide 30) So, at this point you’ve collected your landmark data: you’ve placed landmarks on places you’ve found convenient, or you’ve placed semilandmarks around an object to capture its shape. Now you need to decide on what type of analysis you wish to conduct. Here I’ve summarised methods of landmark analysis into three different methods, the first two grounded in classic GMM (*sensu stricto*), the third not.

1. Landmark Analysis (if your points are individual and represent specific landmarks)
2. Outline Analysis (if you are using semilandmarks and wish to consider these as one general shape)
3. Other (including alternative methods e.g. Polynomial Curve Fitting, we won’t consider this today)

Slide 31/32/33) If we wish to pursue landmark analysis, we first have to extract the actual ‘shape’ from our data: the photographs and models are in different morphospaces, may be rotated incorrectly and may be of different size. Here we perform what is known as the Generalised Procrustes Analysis (also known as Procrustes Superimposition or Generalised Least Squares), a procedure to isolate shape. // This means the raw coordinates (landmarks) are translated to a common centroid (as seen A-B), scaled so they are of the same centroid size (squared distance measure), as seen in B-C, and rotated to minimise the distance between landmarks (C-D). The new landmarks, often called Procrustes coordinates describe the shape and are what we use in our multivariate analyses. It is with these with which we can do all subsequent analyses if we wish. We now have the ‘shape’ of an object. For a bit of historical context, in Greek mythology Procrustes was a son of Poseidon (as was basically everyone in Greek mythology) and lived in a sacred passage between Athens and Eleusis. There he had a bed, in which he invited every passer-by to spend the night. It was then when he set to work on them with his blacksmith’s hammer, stretching them to fit the bed. Procrustes continued his reign of terror until he was captured by Theseus, travelling to Athens, who ‘fitted’ Procrustes to his own bed.

Slide 34) With these new coordinates we can now consider shape change (from one shape to another or to a mean shape of one group to another group mean shape), we can also now perform a variety of different statistical and analytical procedures. In visualising shape change we can use a variety of illustrative techniques including grids (like Wentworth Thompson), lollipop sticks, vectors and contours - at this point it is incredibly customisable.

Slide 35/36/37/38/39) More importantly, we can now explore the data through a variety of multivariate approaches. Sadly, I do not have time to go through them all in detail here, but I’ll highlight some of the key methods. Perhaps the first technique all archaeologists employ is a **Principal Component Analysis** (or PCA for short). Through a PCA we can examine the main sources of shape change (as represented by a series of different axes) and examine how individual artefacts relate to one another, or how groups of artefacts relate to one another. This, however, does not consider the groupings when conducting the analysis, we overlay the group data afterwards. // If we wished to examine the group data explicitly (age, location, sex, etc.), we can conduct a **Discriminant Function Analysis** (DFA/ LDA/ CVA). Through a DFA we determine whether groups defined *a priori* can be distinguished based on their maximum group separation. We can obtain percentages to see how well different artefacts can successfully be discriminated into our groups, or produce a training dataset to compare a series of artefacts with unknown grouping. I’m currently employing these techniques for example to build a dataset of patella of known sex, to test in archaeological case studies. // Both of these techniques are exploratory, and usually the first things archaeologists do with their data; these don’t however provide us with a statistic for a particular hypothesis, but as a means for looking at shape variation. We can conduct a **MANOVA (or Procrustes ANOVA)**to provide a statistical framework for examining shape. If we have a null hypothesis that the two groups or populations are the same we can use a MANOVA to test for particular groupings or factors. // If we wished to examine shape against hypotheses involving quantitative data e.g. size or a symmetry score we can produce a variety of **Regression**-based methodologies. In these (and in certain MANOVAs) we would use the Principal Component scores produced from the PCA. // We may want to continue examining our classifications through various Cluster-based analyses, whether these be quite simple **Hierarchical Clustering** or computationally-intensive techniques including **Maximum Likelihood**. Finally, we may want to adopt techniques from **Machine Learning** and produce **Artificial Neural Networks** and (like other previous techniques) produce new typologies or clusters. In sum, it all depends on what you wish to explore, and the hypotheses you wish to test.

Slide 40) Are there any questions? It is a lot to take in during this one hour, and we don’t expect you to remember this straight away. We’re providing a flavour of what you can do with GMM.

Slide 41) Finally, I want to briefly go over outline-based methodologies in GMM. These are as common as landmark methods and are grounded on semi-landmarks plotted around the outline or feature of an object. These techniques are useful for structures which are comparable in a geometric sense but where individual homologous landmarks are difficult to pinpoint e.g. stone tools. There are a number of advantages in adopting outline analysis methods, particularly that we do not need the same amount of points for each artefact (many more will be necessary for more complex objects, for example) and we do not need to start at exactly the same position. I’ve published quite a few articles on outline analysis and I would encourage you to look them up for further information; there is some really exciting work coming out of Japan at the moment, using many of these outline-based methods too, and I would encourage you all to read these. The main difference in outline methods, in comparison to traditional landmark methodologies, is in that the semilandmarks are fed through a set of parametric equations to fit a curve (grounded on Fourier techniques), and it is the coefficients (which produce these curves) which we examine, unlike the Procrustes coordinates. We do not need to perform a Procrustes Superimposition as the Fourier-based methodologies account for rotation and size within these parametric equations. This will all make more sense when we get to the coding practical, where we conduct an outline analysis. With the coefficients we extract from the outlines (chains of semilandmarks) we can employ all the above techniques, it is only the transformation of the data and the substance which we analyse which differs.

Slide 42) So that is a quick overview of GMM, and how it can be employed in the archaeological record. Looking towards the future we will see a greater application of GMM in non-biological archaeologists, we’ll observe developments in the automation (of models and landmarks), the adoption of more powerful analytical methods (such as Bayesian and Machine Learning techniques), and most-exciting for me, a more open-access, code-sharing, reproducible GMM.

Slide 43) So, I thank you for your attention and I hope you have all have learned something new.