CLASSIFICATION OF SHAPES

GREGORY J. MATTHEWS ET. AL.

1. Introduction

Statistical analysis of shape plays an important role in many areas of science such as biology (O'Higgins [1989], O'Higgins and Dryden [1993], Goo [1989]), chemistry (Dryden and Melville [2007], Czogiel and Brignell [2011]), medicine (Bookstein [1996], Brignell [2010]), bioinformatics (Green and Mardia [2006]), genetics (Horgan and Fenton [1992]), geology (Lohman [1983]) and anthropology Brophy [2014](Matthews 2017). Formal techniques for statistical shape analysis have been developed to extend classical statistical methods to shapes, such as computation of a mean shape and shape variability, extensions of Hotelling's T^2-test for inference about mean shapes, and principal components analysis of shapes (PCA), to name just a few Dryden and Mardia [1998]. Statistical shape analysis techniques generally assume that the entire shape is observed. However, in practice, there are instances where the shapes of interest are only partially observed. For example, the shape of the occlusal surface of isolated fossil teeth from the Family Bovidae is useful to biological anthropologists for identifying the taxa of specimens, which in turn is used to reconstruct paleoenvirnoments. In order to classify these isolated fossil tooth remains, scientists rely on, among other factors, the shape of the outline of the occlusal surface of the tooth to make accurate taxonomic classifications. Currently, only complete (or nearly complete) teeth are generally considered in classification analyses, however, there are numerous specimens of fractured fossil Bovid teeth that are not considered as strongly or ignored entirely in the analysis process because the full shape is not observed. There are countless examples in the missing data literature where dropping observations with missing data (i.e. complete case analysis) has been demonstrated to produce biased results, therefore, it is scientifically beneficial to have methods for including partial observations in an analysis of interest, in this example the classification of fossilized teeth from the family Bovidae.

1.1. **Statistical Shape Analysis.**

Date: September 5, 2019.

1.2. **Landmarks.** Much of the early work in shape analysis was based on defining shapes by a series of landmarks (Bookstein [1986], Mardia and Dryden [1989], Kendall [1984], Le and Kendall [1993]) where each point represents a meaningful location on a shape and there is a natural correspondence between points across shapes in a sample. This allows for a natural correspondence of points across shapes and greatly simplifies analysis in many cases, such as computing a mean shape or variability of a collection of shapes. The use of landmark data also allows for a natural methods for the alignment of shapes to one another using a method called Procrustes analysis Green [1952], Goodall and Mardia [1991], which aligns shapes subject to translations, rotations, and scaling. Additionally, shapes can be defined by unlabeled points where there is some true labeling, but it is unobserved. This setting is more complicated than labeled landmarks since the correspondence between points across shapes is unknown. In many situations, however, there are no reasonable choices for landmarks on the shape. Later,the concept of sliding landmarks Green [1996], Bookstein [1997] was proposed. Sliding landmarks are the same as landmarks except that they are allowed to "slide" along the contour of the curve when being aligned with a reference set of landmarks.

1.3. **Continuous Curves1.** Rather than describing a closed shape with a discrete set of points, a function can be used:

$$C : \mathbb{S} \to \mathbb{R}^n$$

where $\mathbb{S}$ is the unit sphere and guarantees that the function has the same start and end point (i.e. is a closed curve). The function C is a closed curve but scalings, translations, rotations and reparameterizations of this function will usually result in different functions, which means that this framework does not offer invariance across these transformations. Rather than using this function directly, the first derivative could be used, which results in a translation invariant representation.

This representation is the foundation of the square root velocity function (SRVF) Srivastava and Klassen [2016b] framework, which allows invariance across tranlsation, rotation, (optional) scaling, and reparameterization. Frameworks for representing shapes that have invariance over reparameterization are said to be *elastic*.

To define the SRVF start by considering an absolutely continuous function $\beta : \mathbb{S} \to \mathbb{R}^n$. The SRVF of this function $q : \mathbb{S} \to \mathbb{R}^n$ is defined as the following:

where $\dot{C}(t)$ is the first derivative of the function $C(t)$.

A full review of other methods for shape analysis can be found in Srivastava and Klassen [2016b]

Sliding landmarks by Bookstein.

Section 1.2.4 in FSU disserttion is a good description of this.

Shape analysis in general.

Partial Shape matching.

2. Missing Data

Missing data problems arise in many statistical settings due to survey non-response (CITE), designed missingness in surveys Little and Rhemtulla [2013], Wood and Harel [2018], measurement error Guo et al. [2010], and latent classes Hagenaars and McCutcheon [2002], to name just a few examples. There are numerous methods proposed in the literaure describing methods for handling missing data in a statistically principled manner. For example, (Demspter Laird and Rubin) describes the EM algorithm for finding maximum likelihood estimates in the presence of missing data. Another method, of particular interest in this setting, for handling mising data is multiple imputation Little and Rubin [1987], Schafer [2000]. The idea behind multiple imputation is to contruct a model for the missing values conditional on the observed values, and then fill in the missing values by randomly drawing from this distribution. This process is repeated M times to create M completed data sets. An analysis of interest is performed on each of the M completed data sets (e.g. calculating a mean, multiple regression, clustering, etc.) and then the results are combined across the completed data sets using appropriate combining rules Little and Rubin [1987]. For a full review of the missing data literature, see Horton and Kleinman [2007], Murary [2018]

3. Missing Data and Shapes

Work such as Brown and Jackson [2012] examines missing data in the morphology of crocodile skulls. however, their data set consists of linear morphometric measurements of the skulls rather than dealing directly with the shape of th skull through landmarks or semi-landmarks. Other research on missing data in shape analysis has focused on landmark based data such as Gunz [2009, 2004], Neeser [2009], Couette and White [2009]. By defining landmarks on each shape, the problem can be framed in a more traditional statistical setting where each landmark can be thought of as a covariate and more traditional missing data techniques, such as the EM algorithm and multiple imputation, can be used. Arbour and Brown [2014] look at four different methods for dealing with missing landmark data: Bayesian Principal Component Analysis (BPCA), least-squares regression, thin-plate splines (TPS), and mean substitution.

The work most closely related to this current work can be found in Robinson [2012]

Partial shape matching paper at Florida State.

3.1. **13.1 - Incomple e Data.** Papers mentioned:

- Albers and Gower (2010) - Missing data in procrustes
- Mitchelson (2013) - handling occluded landmarks.
- Gunz et. al. (2009) - missing data in studies of biological evolution of bone surfaces

4. Shapes of closed curves1

Our interest is in the completion of shapes of partially observed closed curves, and their classification. This first requires us to adopt a suitable representation for the shape of a fully observed curve. We adopt a parametric representation of a closed curve by representing as an absolutely continuous function[1] $C : \mathbb{S} \to \mathbb{R}^2$, thus automatically ensuring that the curve is closed. The notion of a shape of such a curve requires invariances to transformations that represent nuisance information. Specifically, if $\Gamma := \{\gamma : S \to S$ is an orientation-preserving diffeomorphism$\}$ and $SO(2)$ is the rotation group in $\mathbb{R}^3$, the shape of a parameterized curve $C : D \to \mathbb{R}^2$ is defined to be the equivalence class

$$[C] := \Big\{\sigma OC(\gamma(t)) + a, \gamma \in \Gamma, O \in SO(2), a \in \mathbb{R}^2, \sigma > 0\Big\}.$$

Thus $[C]$ is the set of all possible curves that can be obtained through a translation $(C + a)$, rotation (OC), scale change (σC), a reparametrization $(C(\gamma))$, or a combination of the transformations, of the curve C. In words, the shape of a curve C is what is left once variations due scale, translation, rotation and reparametrisation has been accounted for.

In this instance, it is also of interest to preserve the size and shape of the curve. This simply requires removing any scale changes from the definition of the equivalence class. This size-and-shape of a parameterized $C : D \to \mathbb{R}^2$ is defined to be the equivalence class:

$$[C^\star] := \Big\{\sigma OC(\gamma(t)) + a, \gamma \in \Gamma, O \in SO(2), a \in \mathbb{R}^2, \Big\}.$$

Thus $[C^\star]$ is the set of all possible curves that can be obtained through a translation $(C + a)$, rotation (OC), a reparameterization $(C(\gamma))$, or any combination of the transofrmations of curve C, notably with scaling not included as one of the possible transformations.

A key ingredient in several classification methods (e.g. linear/quadratic discriminant analysis; kernel-based methods) for functional data is the notion of similarity or distance between curves. A popular choice is the distance induced by the $\mathbb{L}^2$ norm of a Hilbert space of square-integrable functions. However, it is well known that the $\mathbb{L}^2$ is unsuitable for comparing curves in the presence of parameterisation variability Kurtek et al. [2012], Srivastava and Klassen [2016a]. To this end we employ a suitable representation (transformation) of a curve C that allows us to compute distances easily while accounting for the necessary invariances.

4.1. **Notation.** A closed planar curve C is always an absolutely continuous mapping $C : \mathbb{S} \to \mathbb{R}^2$, and the set of planar closed curves will be denoted by $\mathcal{C}$. The set $\mathcal{C}$ is equipped with the norm $\|x\|_{\mathbb{L}^2} := [\int_{\mathbb{S}} \|x(t)\|_2 dt]^{1/2}$ where

[1]Upon identification of $\mathbb{S}$ with $[0, 1] \cong \mathbb{R}/2\pi\mathbb{Z}$, a curve $C : [0, 1] \to \mathbb{R}^2$ is absolutely continuous if and only if there exists an integrable function $g : [0, 1] \to \mathbb{R}^2$ such that $C(t) - C(0) = \int_0^t g(u)du, \forall t \in [0, 1]$.

$\|\cdot\|_2$ is the Euclidean norm in $\mathbb{R}^2$. $SO(2)$ is the special orthogonal group of rotation matrices of $\mathbb{R}^2$, and Γ is group of orientation-preserving diffeomorphisms of $\mathbb{S}$. The set Γ_I will denote the group $\{\gamma : [0,1] \to [0,1], \gamma' > 0, \gamma(0) = 0, \gamma(1) = 1\}$.

4.2. **Square-root velocity transform.** For a detailed introduction to the transform, its properties and advantages we refer to the reader to Chapter 6 of the book by Srivastava and Klassen [2016a]. Here we briefly outline the important concepts required for our purposes.

For an absolutely continuous curve $C : \mathbb{S} \to \mathbb{R}^2$ consider the transformation

$$C \mapsto \frac{C'}{\|C'\|_{\mathbb{L}^2}} =: Q_C,$$

where C' is the (vector) derivative of $C(t)$ with respect to t, and $\|\cdot\|$ is the usual Euclidean norm in $\mathbb{R}^2$.

The unique (up to translations) inverse of a square-root transformed curve Q_C is $\int_o^t Q_C(s)\|Q_C(s)\|_2 ds$. To ensure that the curves are closed we need to impose an additional constraint: $\int_{\mathbb{S}} \|C'(t)\|_2 dt = \int_{\mathbb{S}} Q_C(t)\|Q_C(t)\|_2 = 0$.

For a curve C, by taking its derivative and dividing by its length $\|C'\|_{\mathbb{L}^2}$, the transform accounts for translation and scale variabilities. Thus the image of the set of absolutely continuous closed curves with fixed lengths l, $\{C : \mathbb{S} \to \mathbb{R}^2 : \int_{\mathbb{S}} \|C'(t)\|_2 dt = l\}$, under the square-root transform map $C \mapsto Q_C$ is the set

$$\mathcal{Q} := \Big\{ Q_C : \int_{\mathbb{S}} \|Q_C\|_2^2 = 1, \int_{\mathbb{S}} Q_C(t)\|Q_C(t)\|_2 dt = 0 \Big\},$$

since $\|Q_C\|_{\mathbb{L}^2} = \|C'\|_{\mathbb{L}^2}^{1/2}$. Thus the set $\mathcal{Q}$ is a subset of $\mathbb{L}^2(\mathbb{S}, \mathbb{R}^2) := \{Q : \mathbb{S} \to \mathbb{R}^2 : \int_{\mathbb{S}} \|Q(t)\|_2^2 < \infty\}$. It is referred to as the pre-shape space corresponding to the curves, since variations due to rotation and parameterization are yet to be accounted for. It is not a linear space and is a manifold [Srivastava and Klassen, 2016a].

Before defining the shape space, we discuss the actions of groups Γ and $SO(2)$ on the set $\mathcal{Q}$. The set Γ of reparameterisations (or warp maps) of $\mathbb{S}$ is group with group action given by composition. Its action on $\mathcal{Q}$ is defined by $(Q_C, \gamma) \mapsto Q_C(\gamma)\sqrt{|\gamma'|}$, where γ' is the derivative of γ (see Chapters 5 and 6 Srivastava and Klassen [2016a] for more details). The derivative of $\gamma : \mathbb{S} \to \mathbb{S}$ needs to be viewed as a derivative of $\gamma : [0,1] \to [0,1]$ based on the identification $\mathbb{S} \cong \mathbb{R}/2\pi\mathbb{Z}$, and hence $|z|$ is just the absolute value of the real number z. The action of the rotation group $SO(2)$ is defined in the usual way as the map $SO(2) \times \mathcal{Q} \to \mathcal{Q}$ with $(O, Q_C) \mapsto \{OQ_C(t) : t \in \mathbb{S}\}$.

Two important ramifications of the described framework, motivating its use in our work for analyzing shapes of curves, are the following. Under the square-root velocity framework:

1. The actions of $SO(2)$ and Γ on $\mathcal{Q}$ commute, i.e. they can be applied to a curve in any order. This ensures that their combined action is given by the product group $\Gamma \times SO(2)$.
2. The action of $\Gamma \times SO(2)$ on $\mathcal{Q}$ is by isometries: Given two curves C_1 and C_2, we have $\|OQ_{C_1}(\gamma)\sqrt{\gamma'} - OQ_{C_2}(\gamma)\sqrt{\gamma'}\|_{\mathbb{L}^2} = \|Q_{C_1} - Q_{C_2}\|_{\mathbb{L}^2}$, for every $(\gamma, O) \in \Gamma \times SO(2)$. This ensures that if two square-root velocity transformed curves are rotated and reparameterized the same way, their distance remains unchanged.

Starting with a curve C we can now define its shape to be the equivalence class or its orbit of its corresponding square-root transform:

$$[Q_C] = \text{closure}\{OQ_C(\gamma)\sqrt{\gamma'} : (\gamma, O) \in \Gamma \times SO(2)\},$$

where the closure is with respect to the norm $\|\cdot\|_{\mathbb{L}^2}$ on Q. The shape space consequently is defined as $\mathcal{Q}_s := \{[Q_C] : Q_C \in \mathcal{Q}\}$. Property (2) ensures that the metric induced by the norm $\|\cdot\|_{\mathbb{L}^2}$ on $\mathcal{Q}$ descends onto a metric d on the shape space (quotient space) $\mathcal{Q}_s$ in a natural way. Given two curves C_1 and C_2, the shape distance between them is defined as

$$\begin{aligned}(1)\quad d(C_1, C_2) &:= \inf_{(\gamma,O)\in\Gamma\times SO(2)} \|Q_{C_1} - OQ_{C_2}(\gamma)\sqrt{\gamma'}\|_{\mathbb{L}^2} \\ &= \inf_{(\gamma,O)\in\Gamma\times SO(2)} \Big[\int_{\mathbb{S}} \Big\|Q_{C_1}(t) - OQ_{C_2}(\gamma(t))\sqrt{\gamma'(t)}\Big\|_2^2 dt\Big]^{1/2} \\ &= \inf_{(\gamma,O)\in\Gamma\times SO(2)} \|OQ_{C_1}(\gamma)\sqrt{\gamma'} - Q_{C_2}\|_{\mathbb{L}^2}.\end{aligned}$$

The symmetry with respect to the action either on Q_{C_1} or Q_{C_2} is an attractive feature and will be used profitably in the sequel.

5. Preserving scale

For an absolutely continuous curve $C : \mathbb{S} \to \mathbb{R}^2$ consider the transformation

$$C \mapsto C' =: Q^\star_C = Q_C\|C'\|_{\mathbb{L}^2}$$

where C' is the (vector) derivative of $C(t)$ with respect to t.

In this case, the unique (again, up to translations) inverse of this curve $Q^\star_C$ is $\int_0^t Q^\star_C(s)ds$. To ensure that the curves are closed we need to impose an additional constraint: $\int_{\mathbb{S}} Q^\star_C(t) = 0$. Thus the image of the set of absolutely continuous closed curves , $\{C : \mathbb{S} \to \mathbb{R}^2 : \int_{\mathbb{S}} \|C'(t)\|_2 dt = l\}$, under the scale preserving transform map $C \mapsto Q^\star_C$ is the set

$$\mathcal{Q} := \Big\{Q^\star_C : \int_{\mathbb{S}} Q^\star_C(t)dt = 0\Big\},$$

Starting with a curve C we can now define its size and shape to be the equivalence class or its orbit of its corresponding velocity transform:

$$[Q^\star_C] = \text{closure}\{OQ^\star_C(\gamma)\sqrt{\gamma'} : (\gamma, O) \in \Gamma \times SO(2)\},$$

where the closure is with respect to the norm $\|\cdot\|_{\mathbb{L}^2}$ on $Q^\star$.

In the square-root velocity framework, dividing the velocity function by its norm gives us a function that is of norm one. This implies that it is a point on the unit sphere. In the finite-dimensional vecotr set-up this amounts to keeping just the direction of the vectors and discarding all other information and the distance between two curves can then be viewed as the angle between these two vectors. However, when preserving scale this defintion of distance will need to be modified as both the angle AND magnitude of the vectors are both of interest.

Therefore, a modified defintion of distance that preserves scaling between two curves will be defined:

$$
\begin{aligned}
(2)\quad d^{\star}(C_1, C_2) &:= \inf_{(\gamma, O)\in\Gamma\times SO(2)} \|Q^{\star}_{C_1} - OQ^{\star}_{C_2}(\gamma)\sqrt{\gamma'}\|_{\mathbb{L}^2} \\
&= \inf_{(\gamma, O)\in\Gamma\times SO(2)} \Big[\int_{\mathbb{S}} \Big\|Q^{\star}_{C_1}(t) - OQ^{\star}_{C_2}(\gamma(t))\sqrt{\gamma'(t)}\Big\|_2^2\, dt\Big]^{1/2} \\
&= \inf_{(\gamma, O)\in\Gamma\times SO(2)} \|OQ^{\star}_{C_1}(\gamma)\sqrt{\gamma'} - Q^{\star}_{C_2}\|_{\mathbb{L}^2}.
\end{aligned}
$$

6. Missingness Function

Now consider a function, R, defined over the set $\mathbb{S}$, where $R(t) = 1$ if the shape function is observed at $t \in \mathbb{S}$ and $R(t) = 0$ otherwise.

Specifically,

$$R : \mathbb{S} \to \{0, 1\}^2$$

In $\mathbb{R}^2$, for a given point t we have $C(t) = (x_t, y_t)$ and the defintion of the function R allows these two be missing individually (e.g. x_t could be observed and y_t is missing or y_t could be observed and x_t is missing), in our setting both the x_t and y_t values are either both observed or both missing and so specifically in this case the function is of the following form:

$$R : \mathbb{S} \to \{\{0, 0\}, \{1, 1\}\}$$

Now for a specific curve C in the equivalence class $[C]$ (or $[C^{\star}]$), we can define the function R over $\mathbb{S}$. However, for every reparameterization of C in the equivalence class $[C]$ (or $[C^{\star}]$), there is a corresponding reparameterization of R. Also, note that $R^{-1}(1)$ is the subset of domain $\mathbb{S}$ such that the coordinates of the shape are observed for every point in this set. Further, here we only the case where $R^{-1}(1)$ be a connected subset of $\mathbb{S}$ is considered, though in general, this does not have to be the case. This prevents us from considering partially observed shapes with multiple "gaps" (though this is an interesting idea for future work).

For a given shape, there are many possible reparameterizations of the missingness function, and an equivalence class can be defined for the missingness function across reparametertizations as follows:

$$[\mathcal{R}] = \{R(\gamma(t)), \gamma \in \Gamma\}$$

.

Note that for every curve C and corresponding R, in order to maintain the correct relationship, if C is reparameterized by $\gamma \in \Gamma$ then R must be reparameterized using the same γ that acted on C.

Given $\mathcal{R}$, we can break $[C^\star]$ into the observed and missing part of the size and shape as follows:

$$[C^\star]_{obs} = \{OC(\gamma(t)) + a, R(\gamma(t)), \gamma \in \Gamma, O \in SO(2), a \in \mathbb{R}^2 | R(\gamma(t)) = 1\}$$

and

$$[C^\star]_{mis} = \{OC(\gamma(t)) + a, R(\gamma(t)), \gamma \in \Gamma, O \in SO(2), a \in \mathbb{R}^2 | R(\gamma(t)) = 0\}$$

.

Likewise for $[C]$ we can do the same thing, but also consider scaling:

$$[C]_{obs} = \{\sigma OC(\gamma(t)) + a, R(\gamma(t)), \gamma \in \Gamma, O \in SO(2), a \in \mathbb{R}^2, \sigma 0 | R(\gamma(t)) = 1\}$$

and

$$[C]_{mis} = \{\sigma OC(\gamma(t)) + a, R(\gamma(t)), \gamma \in \Gamma, O \in SO(2), a \in \mathbb{R}^2, \sigma 0 | R(\gamma(t)) = 0\}$$

.

6.1. **Completing the curve.** Suppose we are given closed planar curves $C_j^p : \mathbb{S} \to \mathbb{R}^2, j = 1 \ldots, n_p$ each of which has been observed only on a region $\{t : R(\gamma(t) = 1\} \subset \mathbb{S}$, where $\mathbb{S}$ is the unit circle in $\mathbb{R}^2$. We assume that the curves are absolutely continuous. Additionally, a training sample $\{(y_i, C_i^f), 1, \ldots, n_f\}$ consisting of class labels $y_i \in \{0, 1, \cdots G\}$ and $C_i^f : \mathbb{S} \to \mathbb{R}^2$, fully observed on a common domain $\mathbb{S}$, is provided. The set of fully observed curves are elements of the set $\mathcal{C}^f$; denote by $\mathcal{C}^p$ the set of all partially observed curves.

The problem at hand is to model the shape (and size) of and complete each partially observed curve C_j^p and assign it to one of G groups. We consider two approaches: one based on a variational formulation, and the other using kernel-based classifier. Before describing our approaches, some comments on the set Γ are in order.

6.2. **The set of reparameterizations Γ.** Elements of the group Γ of diffeomorphisms of $\mathbb{S}$ can be viewed in the following manner. The unit circle $\mathbb{S}$ can be identified with the quotient group $\mathbb{R}/2\pi\mathbb{Z} \cong [0, 1]$. Through this identification, every continuous mapping $\beta : \mathbb{R} \to \mathbb{R}$ induces a continuous mapping of $\mathbb{S}$ onto itself such that $\beta(t + 1) = \beta(t) + 1$ for all $t \in \mathbb{R}$. If β is monotone increasing, we say that the induced map on $\mathbb{S}$ is orientation-preserving (based on a choice of clockwise or anti-clockwise orientation).

Consider now the set

$$\Gamma_{\mathbb{R}} := \{\beta : \mathbb{R} \to \mathbb{R} : \beta(t + 1) = \beta(t) + 1, \text{ continuous and increasing}\}$$

.

Each member β of $W_{\mathbb{R}}$ induces a warp map $\tilde{\beta} : \mathbb{S} \to \mathbb{S}$ with $\tilde{\beta}(e^{2\pi it}) = e^{2\pi i\beta(t)}$, where β is referred to as the lift of $\tilde{\beta}$. This β satisfies $\beta(t+1) = \beta(t)+1$ for all $t \in [0,1]$, and consequently we have, for $t \in [0,1]$, $\beta(t) = \gamma(t) + c$, where γ is a warp map of $[0,1]$ and $c \in (0,1]$ (through the identification of $[0,1]$ with $\mathbb{R}/2\pi\mathbb{Z}$). This procedure can be viewed as one that produces a warp map of $\mathbb{S}$ by 'unwrapping' $\mathbb{S}$ at a chosen point s and generating a warp map of $[0,1]$. If $\Gamma_I := \{\gamma : [0,1] \to [0,1] : \gamma' > 0, \gamma(0) = 0, \gamma(1) = 1\}$ is the group of diffeomorphisms of $[0,1]$, the map $\Gamma \mapsto \mathbb{S} \times \Gamma_I$ is a bijection [2]. We will hence employ the product group $\mathbb{S} \times \Gamma_I$ in place of Γ. This ensures that the domain of each curve C_j^p and C_i^f can be identified with $[0,1]$ upon unwrapping the circle.

6.3. **Curve Completion1.** Out goal is to complete $[C]_{obs}$ or $([C^\star]_{obs})$ by randomly drawing from the distribution on the missing part of the shape conditional on the observed part. Specifically, for a fixed translation, rotation, scaling, and parameterization, we seek to approximate the distribution of $C_{mis}|C_{obs}$ and use random draws from this distribution to complete to curve. Here this distribution will be approximated non-parametrically and hot-deck type multiple imputation Andridge and Little [2010] will be used similar in spirit to the idea of predictive mean matching (PMM) (Little [1988]). The basic idea of PMM is to to take partially observed records (termed the donees) and find fully observed records that are close to a particular donee (termed donors). Random draws from the set of donors are then used to complete the partially observed donees. This is repeated M times to create M completed data sets.

The same type of procedure will be performed here, but instead of traditional statistical records, here shapes are considered. The spirit of the procedure proposed here is similar to PMM in that first the partially observed shape (the donee) is measured against fully observed shapes to find potential donors. Once the set of potential donors is defined, one of these shapes is randomly drawn and used to complete the partially observed donee. This process is crepeated M times to create M completed shapes.

The observed region $\mathcal{R}_j$ associated with a partially observed curve C_j^p is the subinterval $[0, t_j]$ with $t_j < 1$ for all $j = 1, \ldots, n_p$. Suppose that $C_j^p(0) = \mathbf{a}_j := (a_{1j}, a_{2j})^T$ and $C_j^p(t_j) = \mathbf{b}_j := (b_{1j}, b_{2j})^T$.Then the set of curves that can possibly comprise the missing segment of curve C_j^p is

$$\mathcal{X}_j := \{X : [t_j, 1] \to \mathbb{R}^2 : X(t_j) = \mathbf{b}, X(1) = \mathbf{a}\}$$

.

[2]Technically, this is not a bijection since for a $\gamma \in \Gamma$, the corresponding $\beta \in \Gamma_I$ has a jump discontinuity at the point $t_c \in [0,1]$ where $\beta(t_c) + c = 1$. This can be circumvented by assuming that the members of Γ and Γ_I are absolutely continuous (as opposed to diffeomorphims); then the map between to the sets is bijective a.e.

For a partially observed $C_j^p : [0, t_j]$ and an $X \in \mathcal{X}_j$ with $j = 1, \ldots, n$, define its completion to be the concatenated closed curve

$$C_j^p \circ X(t) := C_j^p(t)\mathbb{I}_{t\in[0,t_j]} + X(t)\mathbb{I}_{(t_j,1]}$$

.

$X(t)$ should be chosen from the distribution $C_{mis}|C_{obs}$. In this case, this distribution will be approximated empirically. To begin, for a particular partially observed shape, $C_j^p \in \mathbb{C}^p$, the distance between C_j^p and each of the completed shapes in $C_i^f \in \mathbb{C}^f$ is computed. For a fixed $C_i^f \in \mathcal{C}^f$, for each $i = 1, \ldots, n_f$ define the cost functional $\Phi_{\theta_j} : \mathcal{C}^p \times \mathcal{C}^f \to \mathbb{R}$ by

$$\begin{aligned}\Phi^p_{\theta_j}(C_j^p, C_i^f) &:= d^2(C_j^p, \sigma O C_i^f(\gamma)|_s), \quad \theta_j \in \Theta_j \\ &= \inf_{(s,\gamma,\sigma,O)\in\mathbb{S}\times\Gamma_I\times\mathcal{R}^2\times SO(2)} \|Q_{C_j^p}, \sigma O Q_{C_i^f}(\gamma)|_s\sqrt{\gamma'}\|^2_{\mathbb{L}^2}.\end{aligned}$$

$$\Phi^{p,\star}_{\theta_{ij}}(C_j^p, C_i^f) = d^\star(C_j^p, C_i^f|_s) = \inf_{(s,\gamma,O)\in\mathbb{S}\times\Gamma_I\times SO(2)} \|Q_{C_j^p}, O Q_{C_i^f}(\gamma)|_s\sqrt{\gamma'}\|^2_{\mathbb{L}^2}$$

Let $\theta^*_{ij} = (s^*_{ij}, \gamma^*_{ij}, O^*_{ij}) = \text{argmin}_{\theta_{ij}\in\Theta_{ij}} \Phi_{\theta_{ij}}(C_j^p, C_i^f)$. These values in θ^*_{ij} will be used again when completing tooth C_j^p.

Now for a fixed j, after computing $\Phi^\star_{\theta_{ij}}(C_j^p, C_i^f)$ for all $i = 1, 2, \ldots, n_f$, the indices of the donor set are defined to be $D_j^\star = \{i : \Phi^\star_{\theta_j}(C_j^p, C_i^f) \leq \Phi^\star_{\theta_{ij}}(C_j^p, C_i^f)_{(k)}\}$. $\Phi^\star_{\theta_{ij}}(C_j^p, C_i^f)_{(k)}$ is the k-th smallest value of $\Phi^\star_{\theta_{ij}}(C_j^p, C_i^f)_{(k)}$ over all $i = 1, 2, \ldots, n_f$ (e.g. $\Phi^\star_{\theta_{ij}}(C_j^p, C_i^f)_{(1)}$ and $\Phi^\star_{\theta_{ij}}(C_j^p, C_i^f)_{(n_f)}$ are the minimum and the maximum, respectively).

M values are drawn, with replacement, from $D_j^\star$ and the chosen indices are then used as donors to fill in the missing part of C_j^p with X_j and the completed shape is written as $C_j^p \circ X_j$. Denote by $Q_{C_j^p\circ X}$ the square-root transform of $C_j^p \circ X$. For each $j = 1, \ldots, n_p$, let $\Theta_{ij} := \mathbb{S} \times \Gamma_I \times SO(2) \times \mathcal{X}_j$, and recall that $\mathcal{C}^f$ and $\mathcal{C}^p$ denote the sets of fully and partially observed curves, respectively. For a fixed $C_i^f \in \mathcal{C}^f$, for each $i = 1, \ldots, n_f$ define the cost functional $\Phi_{\theta_{ij}} : \mathcal{C}^p \times \mathcal{C}^f \to \mathbb{R}$ by

$$X_j^{\star\star} = \underset{X_j\in\mathcal{X}_j}{\text{argmin}} \|Q_{C_j^p\circ X_j}, O^*_{ij} Q_{C_i^f}(\gamma^*_{ij})\sqrt{\gamma'^*_{ij}}\|^2_{\mathbb{L}^2}$$

or if scaling is desired

$$X_j^{*} = \underset{X_j\in\mathcal{X}_j}{\text{argmin}} \|Q_{C_j^p\circ X_j}, O^*_{ij} Q_{C_i^f}(\gamma^*_{ij})\sqrt{\gamma'^*_{ij}}\|^2_{\mathbb{L}^2}$$

.

X^*_{ij} (or $X^{\star\star}_{ij}$) can be viewed as draws from the distribution of $C_{mis}|C_{obs}$ where $C_{mis}|C_{obs}$ is approximated non-parametrically.

The optimal shape completion of a partially observed curve $Q_{C_j^p}, j = 1, \ldots, n_p$ is $Q_{C_j^p \circ X_j^*}$, where X_j^* is obtained from the solution set of:

$$\theta_j^* := (s_j^*, \gamma_j^*, O_j^*, X_j^*) = \underset{\theta_j \in \Theta_j}{\operatorname{argmin}} \Phi_{\theta_j}(C_j^p, C_i^f). \tag{3}$$

In the expression above s_j^* corresponds to the optimal point at which $\mathbb{S}$ was unwrapped in order to identify Γ with $\mathbb{S} \times \Gamma_I$; $\gamma_j^* : [0,1] \to [0,1]$ and O_j^* represents the optimal reparameterization of the curve $C_j^p \circ X_j^*$. The use of a valid distance on the quotient shape space $\mathcal{Q}_s$ allows us to apply the shape transformations on $Q_{C_i^p}, i = 1 \ldots, n_p$ in the variational problem with introducing any arbitrariness. The (product) group structure of $\mathbb{S} \times \Gamma_I \times SO(2)$, and its action on $\mathcal{Q}$, ensures that if the transformations were to have been applied to $Q_{C_j^p \circ X^*}$, the resulting solution set would instead contain the corresponding group inverses. (INSERT ILLUSTRATIVE FIGURE).

6.4. **Classification.**

7. Simulation Study

This simulation study is based on the the shapes of the occlusal surface of upper and lower extant bovid molars. Specifically, teeth from seven bovid tribes (Alcelaphini, Antilopini, Bovini, Hippotragini, Neotragini, Reduncini, and Tragelaphini) and twenty species (arundium, buselaphus, caffer, campestris, capreolus, dorcas, ellipsiprymnus, equinus, fulvorufulva, gazella, gnou, leche, marsupialis, niger, oreotragus, oryx, ourebi, scriptus, strepsiceros, taurinus). Six tooth types were considered here lower/mandibular molars 1, 2, and 3 (LM1, LM2, LM3) and upper/maxillary molars 1, 2, 3 (UM1, UM2, UM3). Specific counts of the sample sizes of each tooth type and tribe can be found in table 1.

For each tooth specimen in a specific tooth type (each tooth type was considered separately), the raw representation of each shape was comprised of 60 points traced around the occlusal surface of the tooth. For each tooth, these 60 points were split into two groups roughly divided by a line connecting the mesostyle to the entostyle (metastylid to the ectostylid) of mandibular teeth (maxillary teeth) (Full deatils in appenidix). An example of this can be seen in Figure 1. This type of cut was chosen for this simulation study as this is a break point commonly observed by biological anthropologists in fossilized bovid teeth.

Then, for each half of a tooth specimen, the donee was matched against all the remaining complete teeth (partial teeth were not matched against the full tooth that the partial tooth was derived from) in that specific tooth type to find a set of k suitable donor teeth. The donee was then completed M times by randomly drawing M teeth with replacement from the set of k donor teeth. The selected donor teeth were then used to to complete the tooth as described in section XXX.

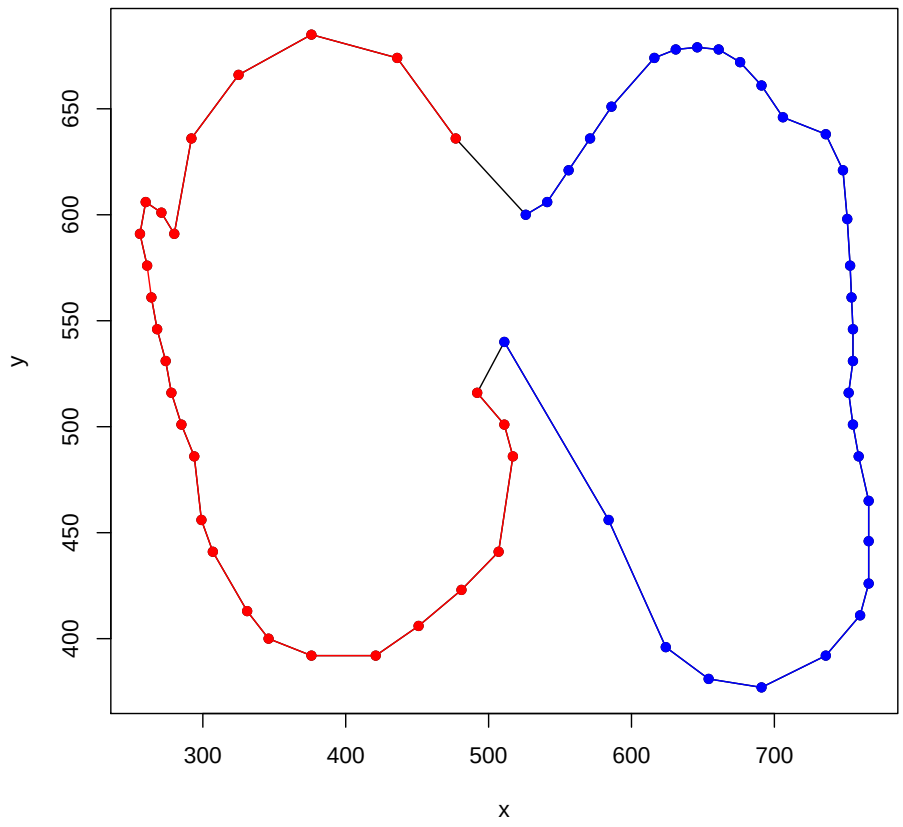


FIGURE 1. Example cuts for an Alcelaphini LM1 (DSCN2871)

The resulting imputed teeth were then used to evaluate the performance of several different types of analyses. Classification is of major interest to biological anthropologists, and a k-nearest neighbors approach to classification was evaluated as well as a (kernel based approach using Frechet means was evaluated maybe?). For k-nearest neighbors, both partial matching and matching after imputation was checked.

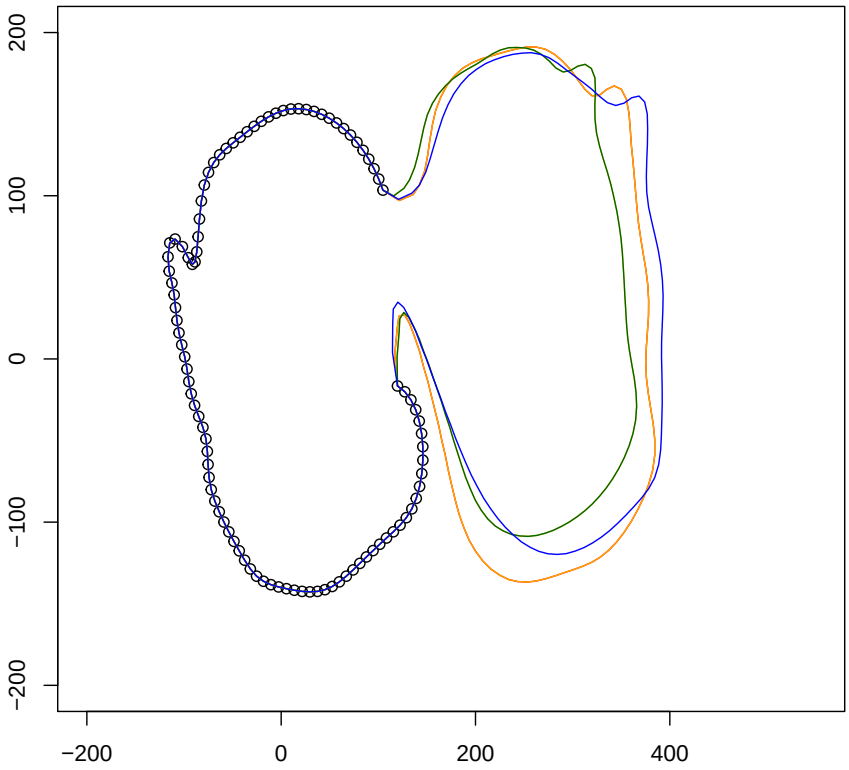
In additional, the total area of the tooth is of interest to biological anthropologists. The accuracy of total area estimation using imputations was also evaluated to test if the total area of a partially observed tooth can be accurately estiamted using this imputation procedure.

We checked different values of M and k and included scaling vs not scaling.

```
k = 5
M = 5
```

```
  side <- 1
  tooth <- "LM1"
load(paste0("/Users/gregorymatthews/Dropbox/shapeanalysisgit/results/results20190610_

plot(t(results_list[[1]]$imputed_partial$imputed[[1]]),col = "red", type = "l", xlim
points(t(results_list[[1]]$imputed_partial$partial_obs))
points(t(results_list[[1]]$imputed_partial$imputed[[2]]),col = "orange", type = "l")
points(t(results_list[[1]]$imputed_partial$imputed[[3]]),col = "gold", type = "l")
points(t(results_list[[1]]$imputed_partial$imputed[[4]]),col = "darkgreen", type = "l
points(t(results_list[[1]]$imputed_partial$imputed[[5]]),col = "blue", type = "l")
```



LM1 - 355 LM2 - 407 LM3 - 411 UM1 - 405 UM2 - 456 UM3 - 416
DSCN2601 example
@

8. Results

Describe the simulation study.

	Alcelaphini	Antilopini	Bovini	Hippotragini	Neotragini	Reduncini	Tragelaphini	Total
LM1	106	27	22	26	45	76	53	355
LM2	117	30	25	37	55	90	53	407
LM3	117	30	19	34	53	96	62	411
UM1	117	30	23	37	43	80	75	405
UM2	118	30	23	41	53	97	94	456
UM3	71	30	17	58	52	110	78	416

TABLE 1. Sample sizes for the simulation study by tooth type and tribe

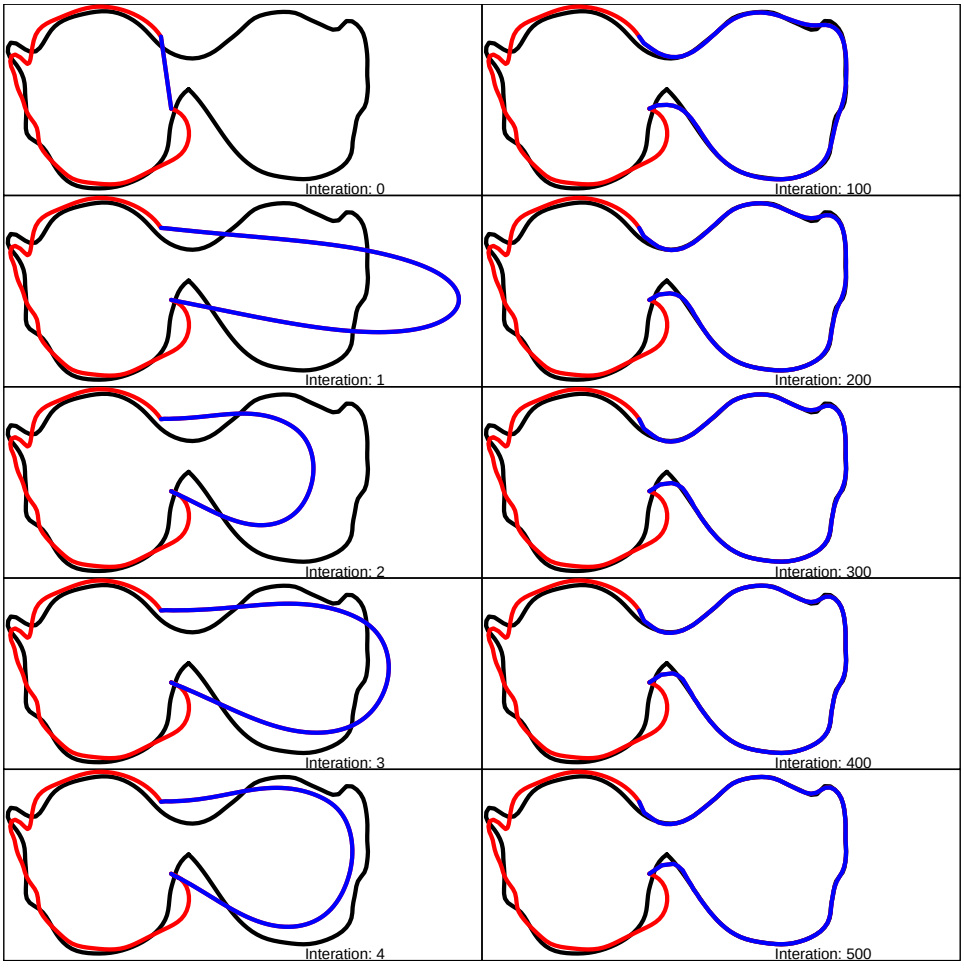


FIGURE 2. An example of how the donor tooth is used to impute the donee

GRADY'S DESCRIPTION: Required scripts and data are loaded into the R environment from external files. R package fdasrvf is used for requisite

shape distance functions, namely for resampling points around a curve and finding the distance between two closed shapes.

A loop begins for both the individual teeth and their left and right halves. The two halves are generated at this step via finding the two points which are separated with the smallest Euclidean distance, provided that these two points are roughly twenty indices away in either direction, ensuring that teeth will be split roughly at the midpoint, where the X would be on a figure-eight.

Teeth halves are stored as "partial shape" alongside an indicator for whether it is the first or second half of the tooth in question, and the complete shape list is amended to not have the tooth in question present so that it would not be matched with itself.

Using the partial shape distance function, distances between the given half of the given tooth and the remaining completed teeth are taken. From the K smallest distances, a sample of size M with replacement is taken. For each of unique teeth in M, an imputed version of the tooth is then created: First, the partial shape is closed via a series of tightly-spaced points connecting the starting and ending point. Both the partial and complete shape are then resampled with splinesfor the partial shape, the straight segment connecting the start and end is done separately from this, in order to avoid spline-based issues where the shape is predicted to curve inwards in a manner not reflective of the true shape.

The partial and whole shape are then compared in order to find the portion of the complete shape which the observed partial shape most adequately corresponds to. This is done by comparing a series of starting segments of the completed shape to the observed partial shape. For each of the subsets, the partial complete shape is scaled and centered, then a best rotation onto the observed partial shape is made. From this, an error term is found, and the segment with the minimum error term is selected as where the closed partial shape will then be overlaid. From here, the straight line segment from earlier is then gradually morphed so that it more and more adequately fits the curvature of the complete shape. After NUMBER iterations, a final version of the imputed partial tooth is obtained.

Classification is done using KNN, where the distances between the imputed tooth and the whole teeth are measured, and the most common class designation amongst the TEN ? teeth with the smallest distance measure is used to predict the class designation of the imputed tooth. A total of M of these predictions is then obtained based on the sampling from the K smallest distances earlier. Further, KNN classification based off the earlier, partial distances is used a means of assessing the performance of classifying based upon imputation.

8.1. **Completing the curves.**

8.2. **Numerical Results.**

8.3. **Tribe Classification.** Results of classification accuracy for tribe are presented using 1-nearest neighbor classification for values of k (5, 10, and 20) and M (5, 10, and 20) along with results when no imputation is used for classification. Six different teeth are considered here (LM1, LM2, LM3, UM1, UM2, and UM3) along with two sides (1 (left) and 2 (right)).

		k=5			k=10			k=20			
Tooth	Side	M=5	M=10	M=20	M=5	M=10	M=20	M=5	M=10	M=20	Partial
LM1	1	0.848	0.854	0.867	0.811	0.834	0.834	0.758	0.777	0.792	0.885
LM1	2	0.854	0.862	0.878	0.842	0.845	0.854	0.831	0.854	0.856	0.896
LM2	1	0.897	0.899	0.922	0.872	0.88	0.892	0.857	0.862	0.867	0.907
LM2	2	0.877	0.892	0.899	0.904	0.897	0.924	0.877	0.897	0.867	0.882
LM3	1	0.876	0.873	0.87	0.842	0.856	0.856	0.818	0.81	0.83	0.886
LM3	2	0.854	0.873	0.879	0.856	0.881	0.869	0.822	0.847	0.859	0.893
UM1	1	0.849	0.872	0.892	0.857	0.867	0.877	0.817	0.844	0.852	0.926
UM1	2	0.872	0.891	0.894	0.862	0.874	0.886	0.815	0.857	0.854	0.921
UM2	1	0.84	0.831	0.828	0.789	0.816	0.822	0.794	0.805	0.818	0.884
UM2	2	0.835	0.849	0.831	0.822	0.864	0.838	0.798	0.816	0.851	0.883
UM3	1	0.769	0.75	0.805	0.74	0.774	0.805	0.798	0.816	0.851	0.853
UM3	2	0.692	0.743	0.675	0.688	0.726	0.74	0.637	0.671	0.851	0.856

TABLE 2. Tribe Classification of size and shape using nearest neighbor classification (i.e. k = 1)

		k=5			k=10			k=20			
Tooth	Side	M=5	M=10	M=20	M=5	M=10	M=20	M=5	M=10	M=20	Partial
LM1	1	0.724	0.752	0.744	0.744	0.749	0.766	0.766	0.766	0.746	0.78
LM1	2	0.71	0.744	0.758	0.758	0.749	0.744	0.744	0.744	0.738	0.806
LM2	1	0.71	0.744	0.769	0.769	0.764	0.774	0.774	0.774	0.73	0.806
LM2	2	0.735	0.742	0.769	0.769	0.762	0.789	0.789	0.789	0.776	0.779
LM3	1	0.725	0.701	0.737	0.737	0.72	0.735	0.735	0.735	0.732	0.737
LM3	2	0.742	0.781	0.779	0.779	0.754	0.796	0.796	0.796	0.803	0.793
UM1	1	0.652	0.674	0.679	0.679	0.691	0.691	0.691	0.691	0.701	0.783
UM1	2	0.681	0.696	0.743	0.743	0.716	0.751	0.751	0.751	0.714	0.842
UM2	1	0.579	0.614	0.607	0.607	0.645	0.638	0.638	0.638	0.636	0.776
UM2	2	0.586	0.618	0.651	0.651	0.614	0.64	0.64	0.64	0.612	0.844
UM3	1	0.476	0.486	0.51	0.51	0.488	0.5	0.5	0.5	0.488	0.62
UM3	2	0.502	0.505	0.5	0.5	0.495	0.51	0.51	0.51	0.507	0.714

TABLE 3. Tribe Classification using shape and nearest neighbor classification (i.e. k = 1)

Tribe classification, Scaled results:
Tribe classification unscaled accuracy plot using knn = 1

```
#    #tooth <- "LM1"
# #side <- 1
# library(ggplot2)
# acc_res_list <- list()
# for (tooth in c("LM1","LM2","LM3","UM1","UM2","UM3")){
#   for (side in 1:2){#print(c(tooth, side))
# load(paste0("/Users/gregorymatthews/Dropbox/shapeanalysisgit/results/summaries/resu
# acc_part_k5_M5 <- acc_part
# acc_imputed_k5_M5 <- acc_imputed
# rm(acc_part)
# rm(acc_imputed)
#
# load(paste0("/Users/gregorymatthews/Dropbox/shapeanalysisgit/results/summaries/resu
# acc_part_k5_M10 <- acc_part
# acc_imputed_k5_M10 <- acc_imputed
# rm(acc_part)
# rm(acc_imputed)
#
# load(paste0("/Users/gregorymatthews/Dropbox/shapeanalysisgit/results/summaries/resu
# acc_part_k5_M20 <- acc_part
# acc_imputed_k5_M20 <- acc_imputed
# rm(acc_part)
# rm(acc_imputed)
#
# load(paste0("/Users/gregorymatthews/Dropbox/shapeanalysisgit/results/summaries/resu
# acc_part_k10_M5 <- acc_part
# acc_imputed_k10_M5 <- acc_imputed
# rm(acc_part)
# rm(acc_imputed)
#
# load(paste0("/Users/gregorymatthews/Dropbox/shapeanalysisgit/results/summaries/resu
# acc_part_k10_M10 <- acc_part
# acc_imputed_k10_M10 <- acc_imputed
# rm(acc_part)
# rm(acc_imputed)
#
# load(paste0("/Users/gregorymatthews/Dropbox/shapeanalysisgit/results/summaries/resu
# acc_part_k10_M20 <- acc_part
# acc_imputed_k10_M20 <- acc_imputed
# rm(acc_part)
# rm(acc_imputed)
#
# load(paste0("/Users/gregorymatthews/Dropbox/shapeanalysisgit/results/summaries/resu
# acc_part_k20_M5 <- acc_part
```

```
# acc_imputed_k20_M5 <- acc_imputed
# rm(acc_part)
# rm(acc_imputed)
#
# load(paste0("/Users/gregorymatthews/Dropbox/shapeanalysisgit/results/summaries/resu
# acc_part_k20_M10 <- acc_part
# acc_imputed_k20_M10 <- acc_imputed
# rm(acc_part)
# rm(acc_imputed)
#
# load(paste0("/Users/gregorymatthews/Dropbox/shapeanalysisgit/results/summaries/resu
# acc_part_k20_M20 <- acc_part
# acc_imputed_k20_M20 <- acc_imputed
# rm(acc_part)
# rm(acc_imputed)
#
# # dat <- data.frame(knn = rep(1:60,4), acc = c(acc_part_5,acc_imputed_5,acc_imputed
# # library(ggplot2)
# # g <- ggplot(aes(x = knn, y = acc, colour = type), data = dat) + geom_point() + ge
# # print(g)
#
# #print(c(mean(acc_imputed_20[1:20]),mean(acc_imputed_10[1:20]),mean(acc_imputed_5[1
#
# # acc_res_list[[paste0(tooth,"_",side)]] <- data.frame(
# #   accuracy  = c(mean(acc_imputed_k20_M20[1:20]),mean(acc_imputed_k20_M10[1:20]),m
# #   tooth_type = rep(paste0(tooth,"_",side),10),  k = c(rep(20,3),rep(10,3),rep(5,3
#
# #knn = 1
# acc_res_list[[paste0(tooth,"_",side)]] <- data.frame(
#   accuracy  = c(mean(acc_imputed_k20_M20[1]),mean(acc_imputed_k20_M10[1]),mean(acc_
#   tooth_type = rep(paste0(tooth,"_",side),10),  k = c(rep(20,3),rep(10,3),rep(5,3),
#
#
#   }}
#
# dat <- do.call(rbind, acc_res_list)
# dat$k <- factor(dat$k, levels = c("20", "10", "5", "No Imp"))
# dat$M <- factor(dat$M, levels = c("20", "10", "5", "No Imp"))
#
# png(paste0("/Users/gregorymatthews/Dropbox/shapeanalysisgit/manuscript/fig/acc_by_t
# ggplot(aes(x = tooth_type, y = accuracy, col = k, shape = M), data = dat) + geom_po
# dev.off()
#
#
```

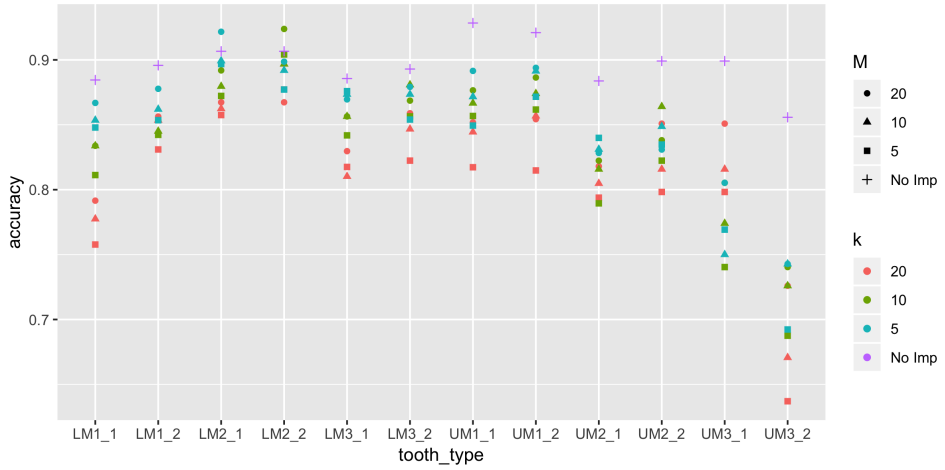


FIGURE 3. Accuracy of tribe classification unscaled

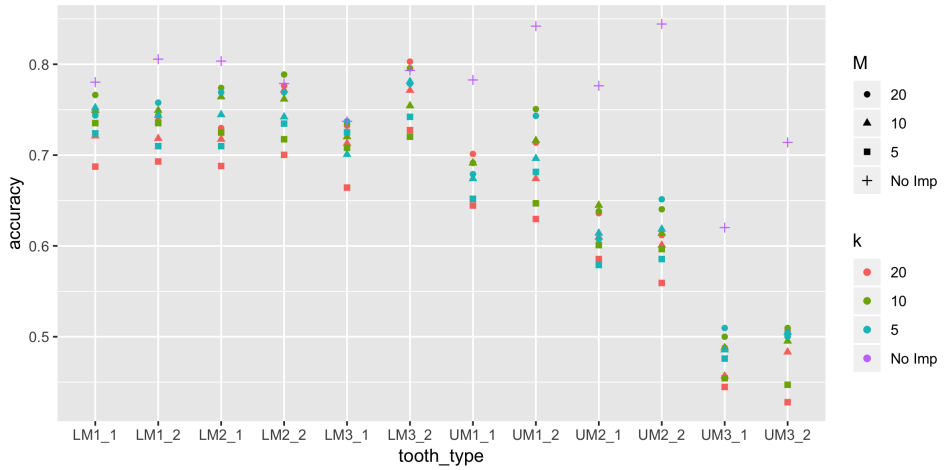


FIGURE 4. Accuracy of tribe classification scaled

Accuracy By Tooth Type

REFERENCES

Growth curve models for correlated traingular shapes. In *Proceedings of the 21st Symposium on the Interface between Computing Science and Statistics. (eds. Berk, K. and Malone, L.)*, pages 445–454. Interface Foundation, Fairfax Station, 1989.

R.R. Andridge and R. J. A. Little. A review of hot deck imputation for survey non-response. *International statistical review*, 78(1):40–64, 2010.

J. Arbour and C. Brown. Incomplete specimens in geometricmorphometric analyses. *Methods in Ecology and Evolution*, 5:16–26, 2014.

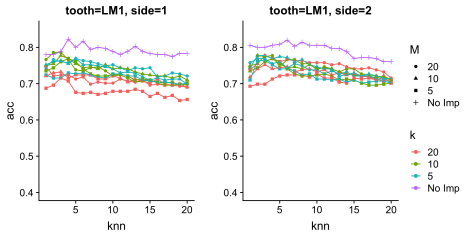


FIGURE 5. Accuracy of tribe classification scaled - LM1

FIGURE 6. Accuracy of tribe classification scaled- LM2

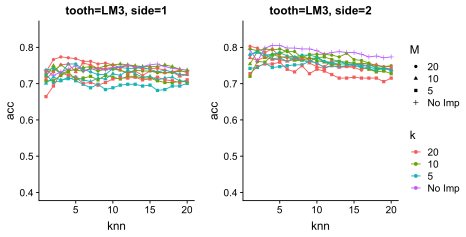


FIGURE 7. Accuracy of tribe classification scaled- LM3

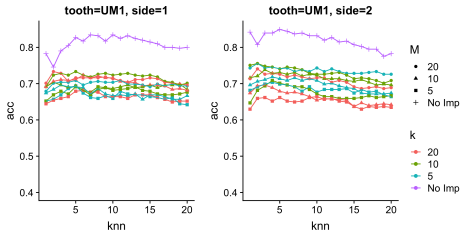


FIGURE 8. Accuracy of tribe classification scaled - UM1

Bookstein. Size and shape spaces for landmark data in two dimensions (with discussion). *Statistical Science*, 1:181–242, 1986.

F.L. Bookstein. Biometrics, biomathematics and the morphometric synthesis. *Bulletin of Mathematical Biology*, 58(2):313–365, 1996.

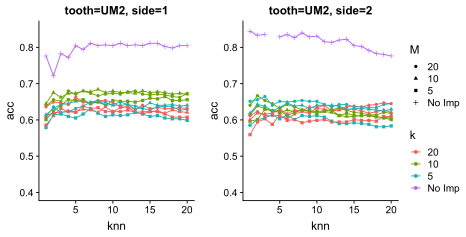


FIGURE 9. Accuracy of tribe classification scaled - UM2

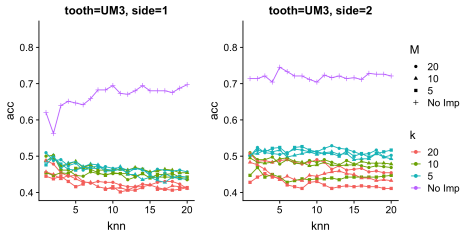


FIGURE 10. Accuracy of tribe classification scaled - UM3

F.L. Bookstein. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Medical Image Analysis*, 1 (3):225–243, 1997.

Dryde I.L. Gattone S.A. et. al. Brignell, C.J. Surface shape analysis, with an application to brain surface asymmetry in schizophrenia. *Biostatistics*, 11:609–630, 2010.

de Ruiter D.J. Athreya S. DeWitt T.J. Brophy, J.K. Quantitative morphological analysis of bovid teeth and its implications for paleoenvironmental reconstructions in south africa. *Journal of Archaeological Science*, 41: 376–388, 2014.

Arbour J. Brown, C. and D. Jackson. Testing of the effect of missing data estimation and distribution in morphometric multivariate data analyses. *Systematic Biology*, 61(6):941–954, 2012.

S. Couette and J. White. 3d geometric morphometrics and missing-data. can extant taxa give clues for the analysis of fossil primates? *General Palaeontology*, 9:423–433, 2009.

Dryden I.L. Czogiel, I. and C.J. Brignell. Bayesian matching of unlabeled marked point sets using random fields, with application to molecular alignment. *Annals of Applied Statistics*, 5:2603–2629, 2011.

Hirst J.D. Dryden, I.L. and J.L. Melville. Statistical analysis of unlabeled points sets: comparing molecules in chemoinformatics. *Biometrics*, 63(1): 237–251, 2007.

I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis*. John Wiley & Sons, Chichester, 1998.

C.R. Goodall and K.V. Mardia. A geometrical derivation of the shape density. *Advances in Applied Probability*, 23:496–514, 1991.

B.F. Green. The othogonal approximation of an oblique structure in factor analysis. *Psychometrika*, 17:492–440, 1952.

P.J. Green and K.V. Mardia. Bayesian alignment using hierarchical models, with applications in protein bioinformatics. *Biometrika*, 93(2):235–254, 2006.

W.D.K. Green. The thin-plate spline and images with curving features. In Gill C.A. Dryden I.L. Mardia, K.V., editor, *Image Fusion and Shape Variability*. University of Leeds Press, Leeds, United Kingdom, 1996.

Mitteroecker P. Bookstein F.L. Weber G.W. Gunz, P. Computer-aided reconstruction of incomplete human crania using statistical and geometrical estimation methods. In *Magistrat der Stadt Wien, Referat Kulturelles Erbe, Stadtarchologie Wien (Ed.), Enter the past: the e-way into the four dimensions of cultural heritage; CAA 2003; computer applications and quantitative methods in archaeology; proceedings of the 31st conference*, pages 92–94. Oxford: Archaeopress, 2004.

Mitteroecker P. Neubauer S. Weber G.W. Bookstein F.L. Gunz, P. Principles for the virtual reconstruction of hominin crania. *Journal of Human Evolution*, 57:48–62, 2009.

Y. Guo, O. Harel, and R.J. Little. How well quantified is the limit of quantification? *Epidemiology*, Suppl 4:S10–6, 2010.

J.A. Hagenaars and A.L. McCutcheon. *Applied Latent Class Analysis*. Cambridge University Press, Cambridge, United Kingdom, 2002.

Creasey A. Horgan, G.W. and B. Fenton. Superimposing two-dimensional gels to study genetics variation in malaria parasites. *Electrophoresis*, 13: 871–875, 1992.

N.J. Horton and K.P. Kleinman. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *American Statistician*, 61(1):79–90, 2007.

D.G. Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16(2):81–121, 1984.

S. Kurtek, A. Srivastava, E. Klassen, and Z. Ding. Statistical modeling of curves using shapes and related features. *Journal of the American Statistical Association*, 107(499):1152–1165, 2012.

H. Le and D.G. Kendall. The riemannian structure of euclidean shape spaces: A novel environment for statistics. *The Annals of Statistics*, 31 (3), 1993.

R.J.A. Little. Missing-data adjustments in large surveys. *Journal of Business and Economic Statistics*, 6:287–296, 1988.

R.J.A. Little and D.B. Rubin. *Statistical Analysis with Missing Data*. John Wiley & Sons, 1987.

T.D. Little and M. Rhemtulla. Planned missing data designs for developmental researchers. *Child Development Perspectives*, 7(4):199–204, 2013.

G.P. Lohman. Eigenshape analysis of micro fossils: a general morphometric procedure for describing changes in shape. *Mathematical Geology*, 15: 659–672, 1983.

K.V. Mardia and I.L. Dryden. Shape distributions for landmark data. *Advances in Applied Probability*, 21:742–755, 1989.

J.S. Murary. Multiple imputation: A review of practical and theoretical findings. *Statistical Science*, 33(2):142–159, 2018.

Ackermann R. R. Gain J. Neeser, R. Comparing the accuracy and precision of three techniques used for estimating missing landmarks when reconstructing fossil hominin crania. *American Journal of Physical Anthropology*, 140:1–18, 2009.

P. O'Higgins. *A morphometric study of cranial shape in Hominoidea*. PhD thesis, University of Leeds, 1989.

P. O'Higgins and I.L. Dryden. Sexual dimorphism in hominoids: further studies of craniofacial shape differences in *pan*, *gorilla*, *pongo*. *Journal of Human Evolution*, 25:183–205, 1993.

Daniel T. Robinson. *Functional Data Analysis and Partial Shape Matching in the Square Root Velocity Framework*. PhD thesis, Florida State University, 2012.

J. Schafer. *Analysis of Incomplete Multivariate Data*. Chapman and Hall/CRC, Boca Raton, 2000.

A. Srivastava and E. P. Klassen. *Functional and Shape Data Analysis*. Springer-Verlag, New York, 2016a.

A. Srivastava and E.P. Klassen. *Functional and Shape Data Analysis*. Springer, New York, NY, 2016b. ISBN 978-1-4939-4020-2.

Matthews G.J. Pellowskim J. Wood, J. and O. Harel. Comparing different planned missingness designs in longitudinal studies. *Sankhya B*, pages 1–25, 2018.