Shape spaces for pre-aligned star-shaped objects – studying the growth of plants by principal components analysis

T. Hotz,[*] S. Huckemann,[†] A. Munk[‡]
Institute for Mathematical Stochastics,
University of Göttingen, Germany

D. Gaffrey and B. Sloboda
Institute for Forest Biometry and Informatics,
University of Göttingen, Germany

June 2009

Abstract

We analyse the shapes of star-shaped objects which are pre-aligned. This is motivated from two examples studying the growth of leaves, and the temporal evolution of tree rings. In the latter case measurements were taken at fixed angles while in the former case the angles were free. Subsequently, this leads to different shape spaces, related to different concepts of size, for the analysis. While several shape spaces already existed in the literature when the angles are fixed, a new shape space for free angles, called *spherical shape space*, needed to be introduced. We compare these different shape spaces both regarding their mathematical properties, and in their adequacy to the data at hand; we then apply suitably defined principal component analysis on these. In both examples we find the shapes to evolve mainly along the first principal component during growth; this is the "geodesic hypothesis" formulated by Le, H. and Kume, A. (Detection of Shape Changes in Biological Features, Journal of Microscopy, 2000 (200), 140–147). Moreover, we were able to link change points of this evolution to significant changes in environmental conditions.
Keywords: Shape analysis; Shape space; Principal components analysis; Log-linear; Growth; Trees; Star-Shaped; Contours.

1 Introduction

The scientific study of the growth and development of plants has a long history, at least dating back to Theophrastus of Eresus in Lesbos (ca. 371–287 BCE)[1]

[1]He was originally named Tyrtamos, but later called Theophrastus by Aristotle, with whom he worked, for the divinity of his style $\tau\grave{o}$ $\tau\hat{\eta}\varsigma$ $\varphi\rho\acute{\alpha}\sigma\epsilon\omega\varsigma$ (p. xxiii ibid.).

and his comprehensive περί φυτών αἴτιών *(On the Causes of Plants)*. In fact, he even discusses some views of Democritus (ca. 460–370 BCE) on the relationship of a tree's shape and its speed of growth, see (Theophrastus, 1976, I.8.2 and II.11.7). Another important contribution was made in the 9th century by Abū Ḥanīfa ad-Dīnawarī (ca. 815–895 CE) who, gathering the knowledge of his time, described the many phases in plants' lives from birth to death in his *kitāb an-nabāt (Book of Plants)*, cf. Bauer (1988); unfortunately, that particular chapter, *General properties of the plants*, has been lost (p.58 ibid.). Quantitative relationships between the size and the shape of an organism apparently began to interest biologists in the 19th century, opening the field which is nowadays known as *allometry*, the phrase having been coined by Huxley and Teissier (1936), cf. Gayon (2000); Niklas (1994) provides an overview over the subject.

In this article, we are going to analyse the evolution of plants' shape over time. We will consider two examples: in the first one the same leaf has been repeatedly photographed over one growing period, in the second one the tree rings (*annuli*) of a stem disk have been determined which allows one to analyse the (lateral) growth of the stem. In both cases, we are interested in determining the development of the shape, in describing it parsimoniously, and in understanding its course. Naturally, our specimens' *size* will increase over time; beyond that, much can be learned from the change of their geometry, i.e. their *shape*: this is what we set out for in this research.

Before continuing, we have to clarify what we mean by *size* and *shape*, as both our results as well as their interpretation hinge on these definitions, cf. Bookstein (1989). A *size variable*, e.g. the square root of a polygon's area, determines the size of an object in such a way that rescaling the object rescales the size variable by the same factor. Kendall (1977, 1986) then defines *shape* as "the geometrical information left, when filtering out *size*, *location* and *rotation*"; this will also be our viewpoint here: we are concerned with *similarity shape*, i.e. two objects feature the same shape iff they are *similar* in the sense of Euclidean geometry.

We stress that our interest lies in the study of the evolution of shape over time, not so much in its relationship to size, i.e. allometry. This is not to say that there will be no allometries, i.e. correlations between size and shape, but the way we defined shape, size and shape can only be correlated through time: as time progresses the specimen under consideration grows in size and simultaneously will vary its shape.

In order to be able to analyse these shapes statistically then, we need to represent shapes in some metrical space such that we can speak of the distance of two shapes. We call this a *shape space*. Obviously different ways of "filtering out" the similarity group lead to different shape spaces and thus to different notions of shape, each with specific advantages and disadvantages over the other concepts. Having determined a shape space, statistical analysis of shape appears within reach. Most shape spaces, however, are non-Euclidean manifolds, even worse, some are only quotients of non-Euclidean manifolds with unbound curvature, and worst, some are non-metrical spaces only (cf. e.g. Schmidt et al. (2007)). This requires either sophisticated methods respecting the non-Euclidean nature of the space, or to work by approximation: usually, shape spaces are approximated locally by suitable linear Euclidean spaces in which standard multivariate analysis can be carried out. Small (1996) as well as Dryden and Mardia (1998) give a broad overview over such methods for
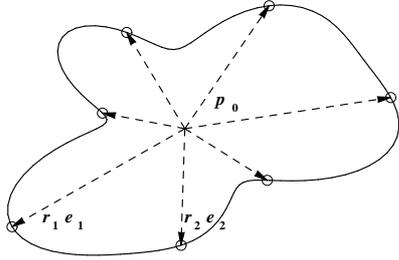
Figure 1: A two-dimensional contour that is star-shaped with respect to a central point p_0. By choosing seven pre-specified directions e_1, e_2, etc., along the contour seven intersection points $r_1 e_1, r_2 e_2$, etc. are determined.

landmark-based shape spaces; with suitable modifications, these linearisations are also employed in the statistical analysis based on more recently developed shape space models, cf. e.g. Krim and Yezzi (2006) for an overview. Meanwhile, methods of intrinsic shape analysis have been proposed, based solely or in part on the non-Euclidean structure, cf. e.g. Le (2001), Fletcher and Joshi (2004), Klassen et al. (2004) as well as Huckemann et al. (2009). In contrast to standard statistical analysis within linear spaces, intrinsic methods can be computationally costly and difficult to analyse theoretically.

Such involved non-Euclidean structures, however, are by no means inevitable: the data analyst might want to use his freedom in the choice of his shape space such to obtain a space which he can easily work with. In fact, for our two applications it is possible to define appropriate shape spaces that are either Euclidean vector spaces or spheres which constitute in a way the most elementary non-Euclidean spaces. Thus, instead of linearising some complicated differential structure, we want to start immediately with a structure which is as simple as possible, allowing us to work intrinsically without much effort. The landmark-based model of Bookstein (1986) is one popular approach where the first two landmarks of a planar object are mapped to pre-specified points by an similarity transformation, thereby specifying translation, rotation and scaling. The drawback of this model, however, is that its shape representation depends on the order in which the landmarks have been numbered, namely the first two play a special rôle. In general, it might not be possible to come up with simple shape spaces which do not depend on some artificial, i.e. subjective, ordering of the landmarks, but in our two examples the data structure is such that it is possible, as we are going to demonstrate.

Tracking the growth of a single specimen, one can often view the growth of some part as originating from a point. A leaf naturally starts growing from its stem, forming the leaf blade, cf. left image of Figure 2. Also the tree rings (*annuli*) of a stem disk capture the evolution of the tree's stem at that particular height from the central pith outwards in our second example, see Figure 8. Marking specific points at the leaf's boundary or at a tree ring, it appears natural to view the polygon they form as a star-shaped domain with the centre being the starting point of growth. Hence we assume that the contours of the m-dimensional geometrical objects being studied are *star-shaped* w.r.t. a distinguished point $p_0 \in \mathbb{R}^m$, i.e. every ray $t \mapsto p_0 + tv$ ($t > 0, v \in \mathbb{R}^m$) originating from p_0 intersects the contour at a unique point. If there is a collection of distinct unit vectors $e_1, \ldots, e_k \in \mathbb{R}^m$, $k \geq 2$ which has been fixed in advanced this leads to a *radii-tuple* $(r_1, \ldots, r_k) \subset (0, \infty)^k$ where
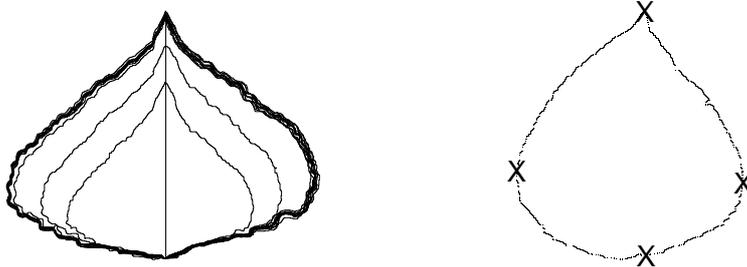
Figure 2: Left: Original contours of a single Canadian black poplar leaf during a growth period. Right: Describing a typical leaf contour with four landmarks: stem, tip and maximal extensions orthogonal.

$p_0 + r_1 e_1, \ldots, p_0 + r_k e_k$ are the unique points of intersection with the contour, cf. Figure 1. Usually, in applications, $m \in \{2, 3\}$ whereas k will be much larger. We note this still differs from the situation in classical morphometrics where the data are given by measurements of certain distances on an object: the radii-tuple contains all the information about the landmarks whereas the shape may not be completely determined when only some distances of landmarks have been measured, cf. the discussion of (Dryden and Mardia, 1998, p. 7).

In our example with the leaves, no angles, i.e. no unit vectors, have been fixed in advance. For this situation we will derive the new *spherical shape space* in Section 2. For the tree rings, on the other hand, data have been collected along fixed rays emanating from the pith such that our analysis will be based on the radii-tuple. We will show how one can obtain a Euclidean shape space in that scenario within Section 3. We will analyse both examples by first reducing the data dimension through principal component analysis, and afterwards tracking their scores over time. This leads to fresh insights into the growth of leaves and tree stems: while both occur mainly unidirectional, we find change points in the growth direction for the latter which we can link to environmental changes. Finally, we will discuss both our methodology and our findings in Section 4.

2 Spherical shape spaces for modelling poplar leaves

2.1 Shapes of poplar leaves

To study the growth of leaves we selected two Canadian black poplar leaves from a dataset collected at the University of Göttingen's Institute for Forest Biometry and Informatics, cf. Table 1.

These leaves have been repeatedly photographed over their growth period from June 2007 to September 2007. The first six measurements have been taken in June with only a few days between, subsequent measurements with increasing time intervals followed until the beginning of September. These leaves' contours have subsequently been digitised as well as translated and rotated such that the starting point of growth from the stem p_0 was placed at the origin and the main

Table 1: Features of leaves considered.

leaf	recorded time period	number of contours
f2b7	June – September	12
f2b9	June – July	7

leaf vein pointed to the positive vertical axis, see Figure 2. Then 4 anatomical landmarks were placed at each leaf's contour: the first at the base, i.e. at the start of the main leaf vein, another one at the end of the main leaf vein, and two more landmarks at the largest extents of the leaf, orthogonal to the dominating direction of the main leaf vein. Thus we have marked the bottom (p_0), right (x_1), top (x_2), and left (x_3) "end" of the leaf, thereby obtaining a pre-aligned, quadrangular representation of the leaf, see again right image of Figure 2; cf. also (Thompson, 1942, p. 1041 et seqq.).

To analyse these quadrangles, we adapt Kendall's shape space model accordingly. In the Kendall (1984) formulation, on every m-dimensional geometrical object studied, k *landmarks* at certain locations $x_1, \ldots, x_k$ are specified. The locations are arbitrary but should correspond to each other on different objects in a meaningful way. Each landmark x_j is an m-dimensional column vector and the $m \times k$ - matrix $X := (x_1, \ldots, x_k)$ is the *configuration matrix*.

In the general approach, location information is filtered out by *Helmertising*, i.e. by multiplying X from the right with a k-sub-Helmert matrix. This yields a $m \times (k-1)$ -matrix which can be viewed as containing $k-1$ landmarks only. The very objective of Helmertising is to filter out location information in a uniform way thus ensuring that shape distances are independent of the order the landmarks are numbered. In view of our applications for pre-aligned objects with a specified central location p_0, two aspects have to be considered. Obviously, the location p_0 is also a landmark. Hence, one might want to Helmertise the enlarged configuration matrix $(x_1, \ldots, x_k, p_0)$. This way, however, the information of "pre-alignment" encoded in p_0 is lost but the starting point of growth is clearly of importance in our applications. For this reason, if $x_1, \ldots, x_k$ are landmarks on the contour of a pre-aligned object studied with a central location p_0, we instead remove location information by subtracting p_0 from every column of the configuration matrix, i.e. by placing p_0 at the origin:

$$x_j^* := x_j - p_0 \ (j = 1, \ldots, k), \quad X^* := (x_1^*, \ldots, x_k^*).$$

Following Kendall (1984), we divide X^* by its *size* which is taken as the Euclidean norm, $\|X^*\| = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{k-1} x_{ij}^{*\,2}}$, and the *spherical shape space* is obtained as

$$S_m^k := \left\{ Z \in \mathbb{R}^m \times \mathbb{R}^{k-1} : \|Z\| = 1 \right\},$$

the unit-sphere in the space of $m \times (k-1)$ - matrices. At this point we note that trivial configurations $x_1 = \ldots = x_k$ with zero size, i.e. where all landmarks except, possibly the centre are on top of each other, are excluded from our considerations. With this restriction, the spherical shape distance

$$d(X, Y) = \arccos \frac{\text{trace}\left(X^*(Y^*)^T\right)}{\|X^*\| \, \|Y^*\|}$$

5

of two configurations X, Y, is well defined and indeed invariant under common relabelling of landmarks and under a common translation. Note that this distance arises naturally from the sphere's non-Euclidean geometry, given by its natural embedding into Euclidean space. In classical Kendall shape analysis, S_m^k is called the *pre-shape sphere* since the Kendall shape spaces are then obtained by additionally filtering out rotation information. As we are concerned with pre-aligned data where the rotation has been fixed in advance, e.g. by placing the main leaf vein tangential to the positive vertical axis as described above, we can immediately use the spherical representation $\|X^*\|^{-1} X^*$ in S_m^k.

Alternatively, one might consider the *vertex transformation vectors* of Hobolth et al. (2002). They view planar star-shaped objects with k landmarks as deformations of a regular k-sided polygon. More precisely, they represent every landmark as a complex number $z_j \in \mathbb{C}$ and then remove translation by centreing – which we replace by subtracting z_0, the complex number representing the central location $p_0 \in \mathbb{R}^2$, to obtain $z_j^* = z_j - z_0$. Then, denoting the corresponding vertices of the regular k-sided polygon by $\omega_j = \exp(2\pi i j/k)$, they define the vertex transformation vectors

$$d_j = z_j^*/\omega_j. \tag{1}$$

Finally, they normalise scaling and rotation with $\alpha \in \mathbb{C}$ such that

$$\frac{1}{k} \sum_{j=1}^{k} \alpha d_j = 1. \tag{2}$$

In our case, we do not want remove the rotation as the data are already pre-aligned, so instead one might want to divide through size. Then, however, one obtains a sphere as the shape space, instead of a representation with linear side conditions that give rise to a Euclidean *vertex transformation space* as originally intended by the authors.

2.2 Principal component analysis

Let us first recall the main ingredients for principal component analysis (PCA) in the "classical" setting: suppose that $R^{(1)}, \ldots, R^{(n)}$ are independent realisations of a multivariate random variable R taking values in a Euclidean space $\mathbb{R}^k$. Then consider the empirical covariance matrix ZZ^T, where $Z = \big((R^{(1)} - \overline{R})^T, \ldots, (R^{(1)} - \overline{R})^T\big)$ are the centred realisations with respect to the mean $\overline{R} = \frac{1}{n} \sum_{i=1}^n R^{(i)}$. The eigenvectors of ZZ^T are the *principal components* (PCs) and the eigenvalues (ordered to be non-increasing) give the *variance explained* by the respective principal components, i.e. the (univariate) variance of R projected on the respective PC. As the PCs form a basis of $\mathbb{R}^k$, these variances explain all (multivariate) variation in the data; their sum is the *total variance*. The first PC thus gives the direction of largest variation, and so on. The scalar product of a realisation with a particular PC is called its *score* on that PC.

For non-Euclidean shape spaces, the situation is more difficult. Therefore, we have to generalise the concepts of mean, variance and principal components to multivariate random variables that take values in the spherical shape space; for this we follow the more general methodology of Huckemann et al. (2009). Recall that a geodesic (i.e. the path minimising the distance) in Euclidean space is a straight line whereas on a sphere it is a great circle.

Given independent realisations $Y^{(1)}, \ldots, Y^{(n)}$ of a random variable Y on the spherical space S_m^k, we are concerned with the minimisation of the two quantities:

$$\sum_{i=1}^{n} d(Y^{(i)}, \mu)^2 \quad \text{and} \tag{3}$$

$$\sum_{i=1}^{n} d(Y^{(i)}, \delta)^2 \tag{4}$$

for $\mu \in S_m^k$, and a geodesic $\delta : t \to \delta(t)$ on S_m^k. Note that since S_m^k is compact, both quantities are finite.

A point $\hat{\mu}_I \in S_m^k$ minimising (3) is called an *intrinsic mean* (IM) with *total (intrinsic) variance*

$$V_{\text{int}} := \sum_{i=1}^{n} d(Y^{(i)}, \hat{\mu}_I)^2.$$

A geodesic δ_1 on S_m^k minimising (4) is called a *first spherical principal component* (SPC). A geodesic δ_2 on S_m^k that minimises (4) over all geodesics δ on S_m^k that have at least one point in common with δ_1 and that are orthogonal to δ_1 at all points in common with δ_1 is called a *second SPC*.

Every point $\hat{\mu}_P$ that minimises (3) over all common points μ of δ_1 and δ_2 is called a *principal component mean* (PM). Given a first and a second SPC δ_1 and δ_2 with PM $\hat{\mu}_P$, a geodesic δ_3 is a *third SPC* if it minimises (4) over all geodesics that meet δ_1 and δ_2 orthogonally at $\hat{\mu}_P$. Analogously, SPCs of higher order are defined. Figure 3 illustrates SPCs and means for a sample of three points on a two-sphere. Each of these points corresponds to a triangular, two-dimensional configuration.

Given a SPC δ denote by $Y_{(\delta)}^{(i)}$ the orthogonal projection of $Y^{(i)}$ onto δ. We accordingly call the signed distance of $Y_{(\delta)}^{(i)}$ to $\hat{\mu}_P$ the *geodesic score* of $Y^{(i)}$ on δ; the sign orients the geodesic. By Theorem 2.6 of Huckemann et al. (2009), geodesic scores are uniquely defined outside a null set on S_m^k.

In Euclidean geometry these definitions yield the mean and the principal components as introduced above. In contrast to Euclidean geometry however, $\hat{\mu}_I \neq \hat{\mu}_P$, in general, cf. Theorem 4.1 of Huckemann and Ziezold (2006), and Figure 3 above.

Variance in Euclidean space can be obtained equivalently either by considering projections or by considering residuals. In non-Euclidean geometry the two approaches yield different results. We consider here projection only: suppose we are given SPCs $\delta_1, \delta_2, \ldots$ with PM $\hat{\mu}_P$. Then, define the *geodesic variance* explained by the s-th SPC, $1 \leq s \leq m(k-1)$, by

$$V_{\text{proj}}^{(s)} \quad := \quad \sum_{i=1}^{n} d(Y_{(\delta_s)}^{(i)}, \hat{\mu}_P)^2,$$

leading to cumulative variances

$$V_{\text{proj}}^{[l]} \quad := \quad \sum_{s=1}^{l} V_{\text{proj}}^{(s)}, \quad l = 1, \ldots, m,$$
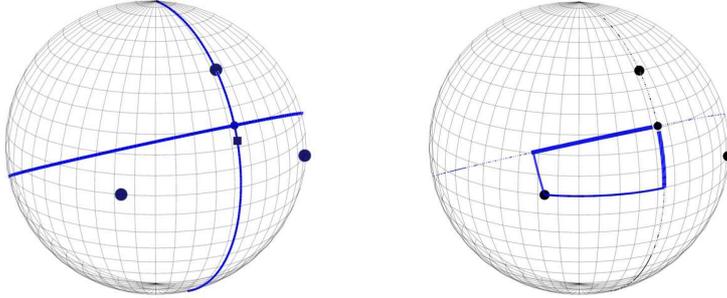
Figure 3: A sample of three points (large dots) on a two-sphere. Left: the first SPC (the thick line close to the equator, i.e. the great circle approximating the sample best with respect to squared spherical distances) intersects the second SPC (thinner meridional line) at the *principal component mean* (small dot). Below is the *intrinsic mean* (the small square), which does not lie on the first SPC. Right: *geodesic scores* of the first data point on the first two SPCs (thick lines) with corresponding residuals (thin lines). Due to spherical curvature, Pythagoras' Theorem does not hold.

Table 2: Cumulative variances of the first three SPCs (spherical shape space) as percentages of total variance obtained by projection and total intrinsic variances for each of the two data-sets of leaf-contours.

leaf	SPC1	SPC2	SPC3	total variance
f2b7	86.1	97.2	99.1	0.0134
f2b9	96.7	99.2	99.6	0.0303

and *total variance obtained by projection*

$$V_{\text{proj}} \quad := \quad \sum_{s=1}^{m} V_{\text{proj}}^{(s)}.$$

In Euclidean geometry $V_{\text{proj}} = V_{\text{int}}$ by Pythagoras' Theorem which is no longer true in spherical geometry, cf. the right part of Figure 3.

Spherical principal components can be calculated iteratively. For the following computations we have used an implementation based on the algorithms provided in Huckemann and Ziezold (2006).

2.3 Growth of poplar leaves

We computed spherical principal components (SPCs) for the quadrangles representing the leaves as defined above. Note that this shape space is 5-dimensional. However, most of the variation over time is explained by the first principal component, see Table 2.

Figure 4 shows the evolution of shape of the first leave over time by as captured by the first PC, as well as the change in size $\|X^*\|$. Not surprisingly, the

latter appears to show the typical logistic growth pattern, cf. Niklas (1994), as does the first PC; similarly for the second leaf (not shown). And indeed, we see a strong linear relationship between size and shape for both leaves in Figure 5. Note however, that each leaf has its own PC, hence follows its own path through shape space, i.e. the allometries we observe here are *intra-subject* allometries whereas most allometric studies analyse allometries within populations. Hence, the allometry we obtain here can be explained by two coincidental events observations: firstly, growth in terms of size and shape happens on the same time-scale, and secondly, the leaf's shape appears to develop straight, i.e. along a geodesic in shape space, towards the shape of the full-grown leaf. This supports the "biological-geodesic hypothesis" stating that biological growth mainly follows the first principal component in shape space, see Le and Kume (2000). We cannot conclude, however, that there is an allometry, i.e. a correlation, between size and shape of the full-grown leaves in a population of leaves.

We note the need to distinguish this evolution of shape along a geodesic in shape space from the shapes of some objects, e.g. spicules, whose growth is forced along geodesics because they are confined to some curved surface, e.g. the cell wall, as described by (Thompson, 1942, p. 675 et seqq., and Ch. X). The latter is the result of a physical constraint to stay within a hollow structure, the former is related to the mathematical definition of shape space. Indeed, if physical constraints restrain the growth we expect an evolution of shape along a curve in shape space which is not a geodesic but features additional curvature.
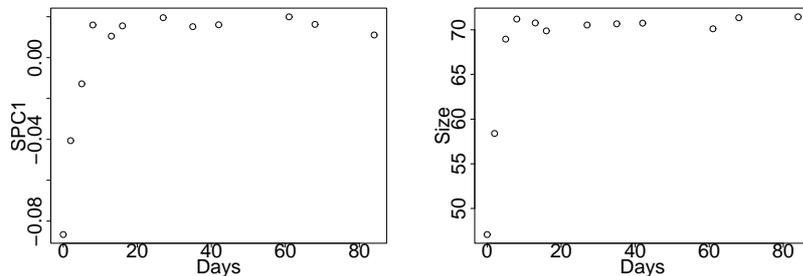
Figure 4: First spherical PC measured in arc length (left) and size measured in mm (right) over time (day zero corresponds to June 5, 2007) for the contours of leaf f2b7.

3 Analysing tree rings in log-shape space

3.1 Shapes of tree rings

In forest biometry, the temporal development of entire tree populations and the development of single tree stems are of great interest. Modelling and understanding these temporal evolutions is of high importance for biological research and forest economical planning as well as for the study of exterior effects such as environmental and climate-based impacts. In this study, we will address the latter, the development of single tree stems. The evolution of such a stem has two aspects: growth not only affects the volume and thus the yield of a tree
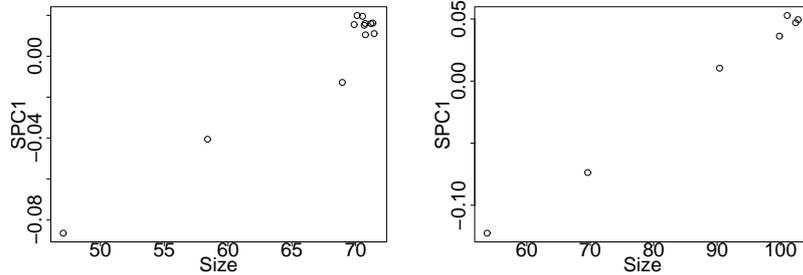
Figure 5: First spherical PC measured in arc length vs. size measured in mm (right). Left: leaf f2b7. Right: leaf f2b9.

Table 3: Features of disks considered.

tree	data-set	at height (m)	number of rings
d104	bottom-disk	0.4	63
	middle-disk	13.1	46
d177	bottom-disk	0.4	62
	middle-disk	15.5	41

but also the tree's shape. The growth of trees in regard to their yield which is closely related to their *size* has been extensively studied for more than two hundred years due to its direct economical impact, cf. e.g. Vanclay (2003). For instance growth change induced by external stress has been of specific interest in recent years, cf. e.g. Gaffrey and Sloboda (2004). Obviously, yield is also related to the shape of tree stems: from an economical point of view, nearly cylindrical pine tree stems are desirable, however, thus leading to studies on how the *shape* of a tree stem develops.

We present here a case study of two Douglas fir stems labelled "d104" and "d177" (Gaffrey and Sloboda, 2001), studying their annual evolution of shape. These trees from an experimental site in the Netherlands have been cut by the end of 1997 at the age of about 65 years. Each of the trees has been cut in several horizontal disks and on every disk, the radii of all rings have been recorded at $k = 36$ evenly spaced angles. Because of the large vertical distances between the disks, we decided to analyse each disk individually, showing us the temporal evolution of the stem's shape at the corresponding heights. For this case study we have selected 2 representative data-sets for each tree which we call *bottom-disk* and *middle-disk*, cf. Table 3. Figure 8 (upper-left) displays the bottom-disk of tree d177.

For each disk the radial distances of each ring from the pith (which is roughly the centre of the innermost ring) are recorded for $k = 36$ angles beginning at $0°$ (which points north) in steps of $10°$. Every tree ring can thus be described by either a *radii tuple*

$$r \quad := \quad (r_1, \ldots, r_k) \in (0, \infty)^k$$

or a *configuration matrix*

$$X \quad := \quad (x_1, \ldots, x_k) \in (\mathbb{R}^2 \times \mathbb{R}^k) \setminus \{0\}$$

10

the columns of which are the *landmarks*:

$$x_j^T = \left(r_j \cos \frac{2(j-1)\pi}{k}, r_j \sin \frac{2(j-1)\pi}{k} \right)^T, j = 1, \dots, k.$$

Obviously, the radii no longer contain any information about location and rotation, thus in order to attain the shape, only *size* has to be filtered out. Size has been viewed as area (2D) or volume (3D), cf. e.g. Small (1996), or as a certain mean mutual distance of contour points, cf. e.g. Kendall (1984) and Bookstein (1986). In 2D for our specific data format, the two views are closely related: the area bounded by a two-dimensional star-shaped contour can be approximated by a multiple of the *arithmetic mean* of squared radii (cf. our choice of size in Section 2.1),

$$\hat{A} = \frac{\pi}{k} \sum_{j=1}^{k} r_j^2;$$

and for the square root of the sum of squared mutual distances we have

$$\sqrt{\sum_{1 \le i < j \le k} \|r_i e_i - r_j e_j\|^2} = \sqrt{k \sum_{j=1}^{k} r_j^2} = k \sqrt{\frac{1}{\pi} \hat{A}},$$

if $\sum_{i=j}^{n} r_j e_j = 0$, i.e. if p_0 is the mean of $p_0 + r_1 e_1, \dots, p_0 + r_k e_k$.

In the case of a general location of p_0, dividing the radii-tuple by its Euclidean length $\|r\|$ its shape is obtained as a point on the unit hyper-sphere of $\mathbb{R}^k$ as suggested by Dryden (2005). This directly leads to the sphere S^{k-1} of normed radii tuples as a shape space.

Note that the approach of Hobolth et al. (2002) leads to the mean radius as size, cf. our discussion further below on p. 12.

Alternatively in a mathematically simpler approach, following Mosimann (1970), Darroch and Mosimann (1985) and Dryden and Gattone (2001) define *size* by instead using the *geometric mean*

$$\hat{S} := \left(\prod_{j=1}^{k} r_j \right)^{\frac{1}{k}} \tag{5}$$

and consider the logarithms of the resized radii-tuple:

$$R := (R_1, \dots, R_k) \quad \text{with} \quad R_j := \log \frac{r_j}{\hat{S}}.$$

Then, these data come to lie in a hyperplane through the origin of $\mathbb{R}^k$ which we call the *log shape space*:

$$R \in \Lambda^k := \left\{ (x_1, \dots, x_k) \in \mathbb{R}^k : \sum_{j=1}^{k} x_j = 0 \right\}.$$

As the log-radii shape space is a linear subspace of Euclidean $\mathbb{R}^k$, we can carry out "classical" multivariate statistical analysis: mean values, principal components, etc. will come to lie in Λ^k.

11

Besides the mathematical elegance, there is also a biological reasoning for taking logarithms: the *allometric* equation

$$y = bx^{\alpha}$$

relates two morphological measurements x and y for the common situation of relative growth, see Huxley and Teissier (1936). Taking logarithms then transforms this into the linear relationship $\log y = \log b + \alpha \log x$, cf. Jolicoeur (1963). Defining size as the geometric mean (5) is then called *isometric size*. While choosing isometric size is debatable in general morphological studies where lengths of different parts are collected, cf. Mosimann (1970), in our setting all measurements are distances to the contour, taken at equi-distant angles, and therefore a priori comparable, rendering isometric size appropriate.

The spherical shape space introduced in Section 2.1 is also applicable in the present situation: fixing the landmarks along prespecified unit length vectors $e_1, \ldots, e_k$ leads to all geodesics between such landmarks to run through the restricted space as well. Indeed, any such geodesic is given by normalised linear combinations of the corresponding shapes, i.e. as an orthogonal projection of the straight line connecting the two in the embedding space; any point on the geodesic can thus be represented as a configuration along these prespecified unit length vectors. In particular means and principal components as introduced above in Section 2.2 preserve the radii representation. We note that this does not hold in general Kendall shape spaces because of the optimal positioning that is necessary to remove the rotation, for a related discussion cf. Lele (1993).

In the specific situation of fixed $e_1, \ldots, e_k$ the vertex transformation shape space of Hobolth et al. (2002) as introduced in Section 2.1 can be adapted to the case of pre-aligned configurations preserving linearity. Indeed, the vertex transformation vectors d_j in (1) have fixed arguments, differing only in their modulus for differing shapes, in the situation of landmarks fixed along $e_1, \ldots, e_k$. The condition for resizing in (2) then is linear in the radii which act as the coefficients of the unit length vectors $d_j/|d_j|$. Thereby size is implicitly defined as the mean radius. We conclude that the *vertex transformation shape space* also qualifies in the present situation.

For all of the above spaces we know by now how to calculate principal components, as we shall do in the following section. We note that Jolicoeur and Mosimann (1960) probably were first to perform PCA on morphological measurements, and Burnaby (1966) proposed PCA on the subspace orthogonal to a general growth vector. Cadima and Jolliffe (1996) discuss PCA on Λ^k, using the geometric mean of morphological distance measures as isometric size just as we do.

In conclusion we note that Krepela (2002) analyses tree stem variation among different trees by describing each horizontal disk by its height and a single radius. Thus modelling with Kendall's classical space for 2D-shapes and employing as usual Procrustes analysis to locally approximate these spaces by Euclidean spaces he finds that a large amount of shape variation is explained by the first PC alone and most of the shape variation by the first two PCs.

3.2 Evolution of tree rings over time

For the two data-sets *bottom-disk* and *middle-disk* for each of the two trees *d104* and *d177* introduced in Section 3.1, principal components in log-shape
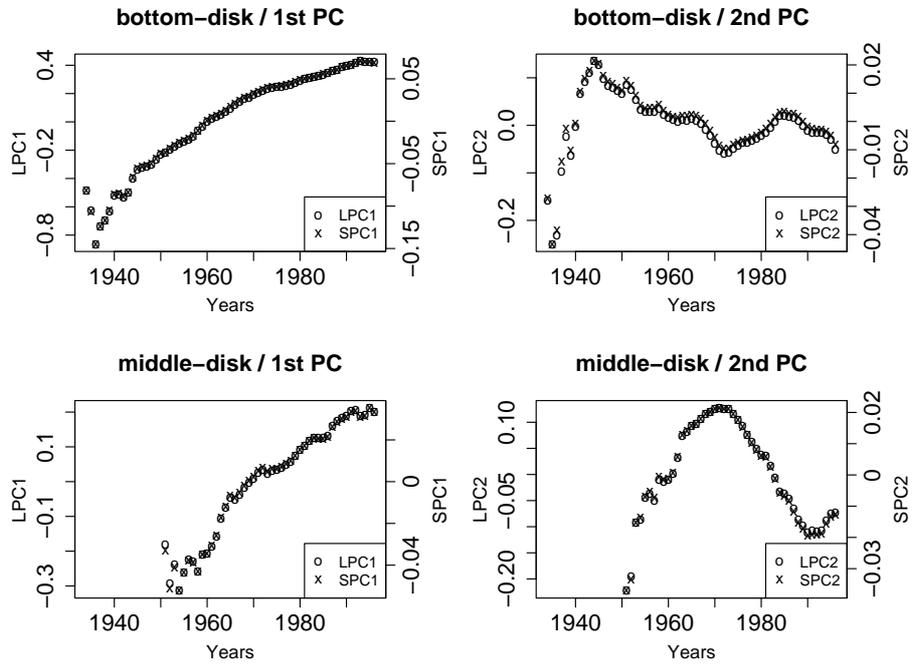
(a) The first two LPCs and SPCs for the bottom-disk and middle-disk of tree d104.
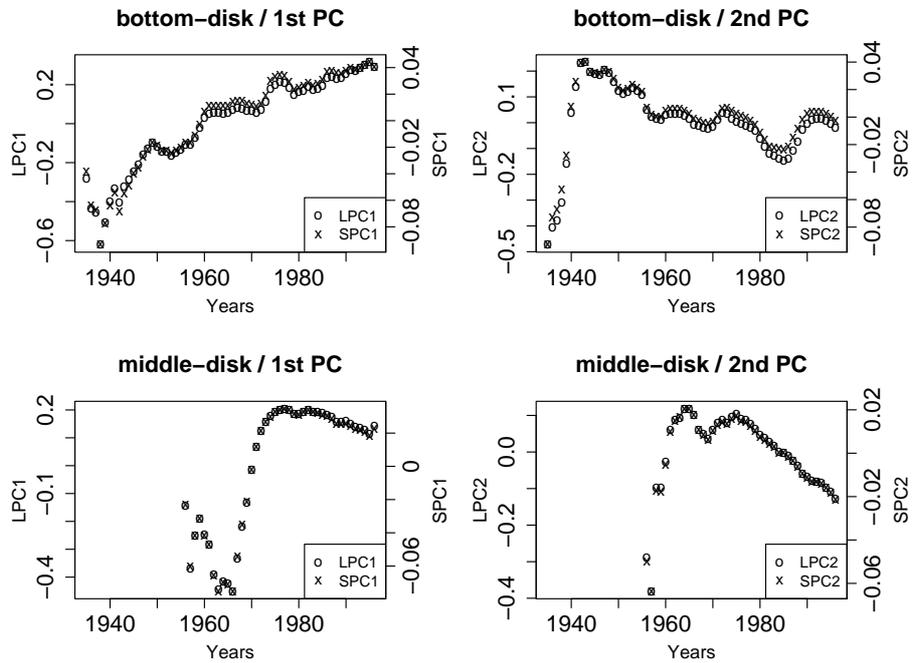
(b) The first two LPCs and SPCs for the bottom-disk and middle-disk of tree d177

Figure 6: The respective PCs for each of the two data-sets of trees d104 and d177 plotted against time; note the differing scales on the vertical axes caused by the different metrics on the respective shape spaces.

Table 4: Cumulative variances of the first five LPCs (log-shape space), SPCs (spherical shape space) and VPCs (vertex transformation shape space) as percentages of total variance obtained by projection and total intrinsic variances for each tree and each of the two data-sets.

tree	data-set	LPC1	LPC2	LPC3	LPC4	LPC5	total variance
d104	middle-disk	67.2	89.3	94.7	96.9	98.3	0.038
	bottom-disk	91.1	94.5	96.3	97.4	98.2	0.138
d177	middle-disk	71.1	85.8	92.2	96.8	98.2	0.078
	bottom-disk	59.5	81.4	90.2	93	95	0.091
		SPC1	SPC2	SPC3	SPC4	SPC5	
d104	middle-disk	66.6	89.3	94.6	97.0	98.3	0.001
	bottom-disk	91.0	94.5	96.2	97.4	98.1	0.005
d177	middle-disk	71.2	86.2	92.3	96.7	98.2	0.002
	bottom-disk	53.9	80	90.2	92.9	94.8	0.004
		VPC1	VPC2	VPC3	VPC4	VPC5	
d104	middle-disk	66.6	89.3	94.6	97.0	98.3	0.039
	bottom-disk	91.1	94.5	96.2	97.4	98.1	0.139
d177	middle-disk	71.2	86.2	92.3	96.7	98.2	0.078
	bottom-disk	53.9	80	90.2	93	94.8	0.089
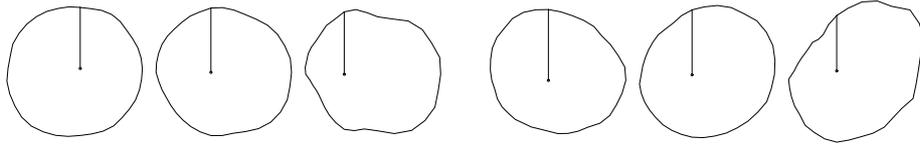
(a) Tree d104   (b) Tree d177

Figure 7: Along first LPC for tree d104 (left) and tree d177 (right): movement from the minimum data-score (left images) to the maximum data-score (centre images) and further beyond by the same distance (right images) along the first PC for the bottom-disks. As before, the vertical radius depicted points north from the pith location. Thus the left images correspond to early shapes, the middle images to the shapes at the time of cutting and the right images to some never observed shapes around 100 years into the future.
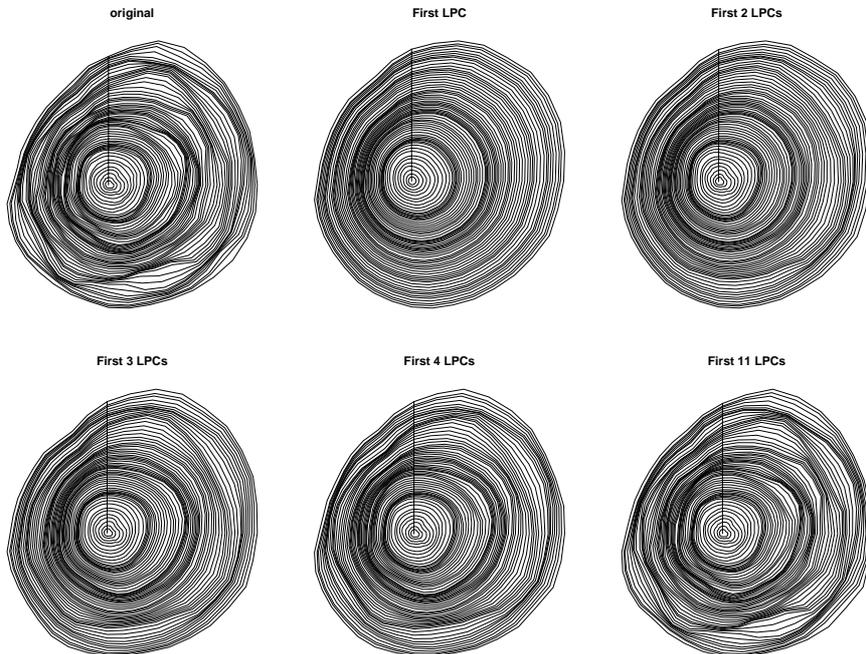
Figure 8: Bottom-disk of tree d177; top-left: original disk, in the following images the rings corresponding to size and projections onto the respective log-linear principal components are depicted. The vertical radius from the pith points north.

space (LPCs), in spherical shape space (SPCs), and in vertex transformation space (VPCs) have been computed. We note that there are $\min(n, 35)$ LPCs as well as non-trivial VPCs, and $\min(n, 70)$ SPCs where n denotes the number of samples in the respective data-set. In Table 4 we report the percentages of total variance explained by the first five respective components. Rather strikingly we notice that the first PC alone explains a large amount of shape variation, the first two PCs together explain more than 80 %.

As a second result, we observe that spherical PCA for the spherical shape space, as well as classical PCA for the log-shape space and the vertex transformation shape space give similar results. These are particularly similar if the data are explained almost exclusively by the first PC. The equivalence of the eigenvectors becomes very much apparent in Figure 6 which shows scatterplots of the temporal shape evolution along the first two LPCs and SPCs, respectively. Even for the bottom disk of tree d177 (the amount of variance explained by the first PC differs by approx. 5 % between the three methods), the plots of the scores along the first PC against time of the three methods are almost identical. Figure 6(b)) displays the scores using the first two methods.

One reason for the similarity of the methods might be that the data are very concentrated as can be seen from the small scores. Hence the mapping of one shape space onto the other is in a good, first-order approximation linear and thus preserves the space spanned by the principal components. We note

15

that the similarity of the results also is reassuring when we are about to interpret them. Indeed, since there is no objective way to prefer any of these shape spaces over the others, our findings might depend on our subjective choice of the shape space; for this data, however, we are fortunate enough to get results whose interpretation remains the same whichever shape space we use. It would be interesting to analyse under which circumstances, i.e. under which mathematical assumptions, this holds. Because the results are so similar, we only give results for log-shape space in the following.

To visualise the shapes' evolution along their respective first PC, we depict the shapes that belong to different scores on the first PC in Figure 7, starting with the lowest score which represent the coming into being of the tree, next the highest score corresponding to the time of cutting the tree, and finally exaggerating this evolution for better visibility by showing the shape which corresponds to the maximal score plus the range of scores, i.e. this shape has the same distance to the middle one as has the first one but in opposite direction.

In a similar fashion, the backtransformation of the entire bottom disk of tree d177 is given in Figure 8 using one LPC, two LPCs etc.; there, the shapes were given the sizes of the corresponding original rings to obtain tree rings of the same scale, rendering them visually comparable with the original stem disk.

More common effects are apparent:

1. After a small number of initial years (between 10 and 15 years), the temporal shape evolution tends to be directed along the first PC, in particular for the bottom disks, see Figure 6. This directed motion is also very well visible in the second image of the first row in Figure 8 and in both parts of Figure 7: the first PC records an overall motion of the pith from east to west within the disk. This effect can be explained with the effort of the tree to counter the dominating wind force from the west at the experimental site by building up stem-mass east of the pith. This effect is not well visible in the middle disks, one may conjecture that here individual effects dominate, less aiming at overall tree-stability, rather then in search of light.

2. Along the second PC, mainly oscillatory patterns are visible, cf. Figures 6 and Figure 8. E.g. in the third image of the first row of the latter figure two bulges appear simultaneously at north and south-east a little after 1970 and disappear again in the early 1990's. Subsequent PCs incorporate more of such oscillatory shape change.

3. A change of orientation around 1970 visible in all second PCs (and mostly visible as changes of slopes in the first PCs).

4. A change point with similar features a little after 1990 well visible in the middle disk of tree d104, in the bottom disk of tree d177 and less visible in the bottom disk of tree d104.

5. One more change point with similar features around 1985 visible in the bottom disks of tree d104 and tree d177.

6. Some more changepoints about every five years before 1970 can be identified in some of the disks.

If the geodesic hypothesis holds, cf. Section 2.3, a change in the direction of growth might indicate a change in the trees' environmental conditions. And indeed, around the tree-age of 10 to 15 years the crowns of nearby trees met and thus tree competition intensified. This effect may explain the first change point in the bottom disk while the major two change points (approx. in 1973 and 1994) might be due to the fact that between the years 1972 and 1993, logging in that forest was halted, thus increasing tree competition. Most of the other change points noted above can be explained by thinning events and extended periods of non-thinning. Beginning from 1952, thinning occurred every three years, from 1961 every 5 years. From 1972 to 1994 thinning was completely halted. Both phenomena, random initial motion and change points certainly deserve further research.

Also, we observe again that the first PC highly correlates with time which in term correlates with size, cf. Section 2.3. Thus for the bottom disks we have a relative east-west motion of the pith correlating with size; again this amounts to an intra-subject effect, although one might expect from the data we've seen that smaller, i.e. younger, trees in this forest's population will again have a rather central pith as opposed to older trees which have already adapted to the dominating wind direction. Then, one should find a population level allometry between size and the shape of the outermost ring.

4 Discussion

It was the aim of this research to provide for simple shape spaces allowing to study how shapes of contours evolve that are

1. pre-aligned with respect to a pre-specified centre, and

2. star-shaped with respect to that centre.

In case of the tree-rings we assumed moreover that

(c) the landmarks are given on the contour at fixed angles.

For the general situation where the angles are free, we found that a spherical shape space was appropriate. This is an adaption of Kendall's shape space where neither rotation nor translation needs to be removed. When the angles are fixed, shape analysis can additionally be performed in log-shape space and vertex transformation space, both of which are Euclidean.

We emphasise that these three shape space not only feature a simple geometry but their geometries are also invariant under cyclical relabelling of the landmarks – excluding the centre, of course. This is a very desirable property of these spaces; indeed, we do not want our statistical analysis of the tree rings, for example, to depend on the order in which we labelled our landmarks, starting from north as we did or starting from south. On the contrary, Bookstein coordinates also lead to a linear space but they clearly let the centre and the first landmark play a special rôle, rendering subsequent analyses dependent on that subjective choice. On the other hand, our model cannot directly be used to study the shapes of general geometrical data, be they not star-shaped or not aligned. There, Bookstein's linear and hyperbolic models as well as the non-linear model of Kendall can still be successfully applied.

Our motivation lay in two applications, where we wanted to explore the growth of poplar leaves and of Douglas fir tree rings, respectively. In both cases we found shape evolution to happen unidirectionally as long as the physical conditions do not mandate a change of course, as was the case for the tree rings, cf. Section 3.2. Principal components analysis, performed intrinsically for the spherical shape space, allowed for a parsimonious description of the data, reducing it to no more than two dimensions relevant for the analysis. Our analyses where based on few leaves and stem disks, though. In the future, more extensive studies comprising larger populations need to be undertaken to statistically solidify the findings described here, e.g. by providing appropriate confidence bands for the amount of variance explained, or by allowing to test for the presence of change points.

Future applications in Forest Biometry lie in the estimation of the prospective growth of trees, cf. Figure 7. This is of value economically since the tree's shape determines how profitable the wood is which can be cut from it, e.g. long, straight boards give higher profits. From our results it appears possible to predict the future evolution of the shape from few data which have been obtained non-destructively, e.g. from horizontal drillings towards the pith.

References

Bauer, T. (1988). *Das Pflanzenbuch des Abū Ḥanīfa ad-Dīnawarī*. Wiesbaden: Otto Harrassowitz.

Bookstein, F. L. (1986). Size and shape spaces for landmark data in two dimensions (with discussion). *Statistical Science 1*(2), 181–222.

Bookstein, F. L. (1989). "Size and shape": A comment on semantics. *Systematic Zoology 38*(2), 173–180.

Burnaby, T. P. (1966, March). Growth-invariant discriminant functions and generalized distances. *Biometrics 22*(1), 96–110.

Cadima, J. F. and I. T. Jolliffe (1996, June). Size-and-shape-related principal component analysis. *Biometrics 52*(2), 710–716.

Darroch, J. N. and J. E. Mosimann (1985). Canonical and principal components of shape. *Biometrika 72*(2), 241–252.

Dryden, I. and S. Gattone (2001). Surface shape analysis from mr images. In *Proceedings in Functional and spatial data analysis*, University of Leeds Press, pp. 139–142.

Dryden, I. L. (2005). Statistical analysis on high-dimensional spheres and shape spaces. *Ann. Statist. 33*(4), 1643–1665.

Dryden, I. L. and K. V. Mardia (1998). *Statistical Shape Analysis*. Chichester: Wiley.

Fletcher, P. T. and S. C. Joshi (2004). Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. *ECCV Workshops CVAMIA and MMBIA*, 87–98.

Gaffrey, D. and B. Sloboda (2001). Tree mechanics, hydraulics and needle-mass distribution as a possible basis for explaining the dynamics of stem morphology. *Journal of Forest Science 47*(6), 241–254.

Gaffrey, D. and B. Sloboda (2004). Modifying the elastomechanics of the stem and the crown needle mass distribution to affect the diameter increment distribution: A field experiment on 20-year old abies grandis trees. *Journal of Forest Science 50*(5), 199–210.

Gayon, J. (2000, October). History of the concept of allometry. *American Zoologist 40*(5), 748–758.

Hobolth, A., J. Kent, and I. Dryden (2002). On the relation between edge and vertex modelling in shape analysis. *Scandinavian Journal of Statistics 29*(3), 355–374.

Huckemann, S., T. Hotz, and A. Munk (2009). Intrinsic shape analysis: Geodesic principal component analysis for Riemannian manifolds modulo Lie group actions (with discussion). *Statistica Sinica.* . To appear.

Huckemann, S. and H. Ziezold (2006). Principal component analysis for Riemannian manifolds with an application to triangular shape spaces. *Adv. Appl. Prob. (SGSA) 38*(2), 299–319.

Huxley, J. S. and G. Teissier (1936, May). Terminology of relative growth. *Nature 137*(3471), 780–781.

Jolicoeur, P. (1963, September). The multivariate generalization of the allometry equation. *Biometrics 19*(3), 497–499.

Jolicoeur, P. and J. E. Mosimann (1960, December). Size and shape variation in the painted turtle – a principal component analysis. *Growth 24*(4), 339–354.

Kendall, D. G. (1977). The diffusion of shape. *Adv. Appl. Prob. 9*, 428–430.

Kendall, D. G. (1984). Shape manifolds, procrustean metrics and complex projective spaces. *Bull. Lond. Math. Soc. 16*(2), 81–121.

Kendall, D. G. (1986). Comment on "size and shape spaces for landmark data in two dimensions" by Fred l. Bookstein. *Statistical Science 1*(2), 222–226.

Klassen, E., A. Srivastava, W. Mio, and S. Joshi (2004, March). Analysis on planar shapes using geodesic paths on shape spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence 26*(3), 372–383.

Krepela, M. (2002). Point distribution form model for spruce stems (picea abies [l.] karst.). *Journal of Forest Science 48*(4), 150–155.

Krim, H. and A. J. J. E. Yezzi (2006). *Statistics and Analysis of Shapes.* Modeling and Simulation in Science, Engineering and Technology. Boston: Birkhäuser.

Le, H. (2001, June). Locating Fréchet means with an application to shape spaces. *Adv. Appl. Prob. (SGSA) 33*(2), 324–338.

Le, H. and A. Kume (2000, November). Detection of shape changes in biological features. *Journal of Microscopy 200*(2), 140–147.

Lele, S. (1993, July). Euclidean distance matrix analysis (EDMA): estimation of mean form and mean form difference. *Math. Geol. 25*(5), 573–602.

Mosimann, J. E. (1970). Size allometry: Size and shape variables with characterizations of the lognormal and generalized gamma distributions. *Journal of the American Statistical Association 65*(330), 930–945.

Niklas, K. J. (1994). *Plant Allometry: the Scaling of Form and Process.* Chicago: The University of Chicago Press.

Schmidt, F. R., E. Töppe, D. Cremers, and Y. Boykov (2007). Intrinsic mean for semi-metrical shape retrieval via graph cuts. In F. A. Hamprecht, C. Schnörr, and B. Jähne (Eds.), *DAGM-Symposium*, Volume 4713 of *Lecture Notes in Computer Science*, pp. 446–455. Springer.

Small, C. G. (1996). *The Statistical Theory of Shape.* New York: Springer-Verlag.

Theophrastus (1976). *De Causis Plantarum, Volume I of III.* Number 471 in The Loeb classical library. London and Cambridge, MA: William Heinemann Ltd. and Harvard University Press. with an English translation by Benedict Einarson and George K. K. Link.

Thompson, D. W. (1942). *On Growth and Form* (new ed.). Cambridge, and New York: Cambridge University Press, and Macmillan Company.

Vanclay, J. K. (2003). Growth modelling and yield prediction for sustainable forest management. *The Malaysian Forester 66*(1), 58–69.