INTRINSIC INFERENCE ON THE MEAN GEODESIC OF PLANAR SHAPES AND TREE DISCRIMINATION BY LEAF GROWTH

BY STEPHAN F. HUCKEMANN[1]

Georg-August-Universität Göttingen

Dedicated to the Memory of Herbert Ziezold 1942–2008

For planar landmark based shapes, taking into account the non-Euclidean geometry of the shape space, a statistical test for a common mean first geodesic principal component (GPC) is devised which rests on one of two asymptotic scenarios. For both scenarios, strong consistency and central limit theorems are established, along with an algorithm for the computation of a Ziezold mean geodesic. In application, this allows to verify the geodesic hypothesis for leaf growth of Canadian black poplars and to discriminate genetically different trees by observations of leaf shape growth over brief time intervals. With a test based on Procrustes tangent space coordinates, not involving the shape space's curvature, neither can be achieved.

1. Introduction. In this paper, the novel statistical problem of developing asymptotics for the estimation of the mean geodesic on a shape space is considered. It is the generalization to a non-Euclidean geometry of the asymptotics for the estimation of a straight first principal component line from multivariate data in the Euclidean geometry. Due to curvature involved, however, methods from linear algebra as employed in the Euclidean geometry cannot be used, and a new approach has to be developed. The task at hand is more involved, yet somehow comparable to the situation of generalizing the concept of the mean for multivariate data to a mean for manifold valued data. For such manifold valued means pioneering work for definitions, existence, uniqueness, algorithms and asymptotics has been done by Gower (1975), Ziezold (1977), Kendall (1990), Goodall (1991), Hendriks and Landsman (1996), Hendriks and Landsman (1998), Le (2001), Bhattacharya and Patrangenaru (2003), Bhattacharya and Patrangenaru (2005) and many others. In this work, definitions for a mean geodesic, an algorithm and asymptotics are proposed and developed for data on Kendall's space of planar shapes. In particular, the following two different statistical scenarios are considered: asymptotics with

respect to underlying shapes—the *mean geodesic of shapes*—and asymptotics with respect to underlying sampled geodesics—the *mean geodesic of geodesics*.

The study of geodesics on shape spaces as the simplest model for a path of temporal evolution of shape is of high interest in shape analysis, in particular, in biological studies comparing growth patterns.

Unlike previous attempts in the literature [e.g., Jupp and Kent (1987), Kent et al. (2001), Kume, Dryden and Le (2007)] building on a Euclidean tangent space linearization of the shape space, the mean geodesic of geodesics defined here builds on a Euclidean tangent space linearization of the *space of geodesics* of the shape space which has been introduced in Huckemann and Hotz (2009). Hence, as a new and abstract concept, we treat here geodesics as data points.

In application, in a joint research study on leaf growth with the Institute for Forest Biometry and Informatics at the University of Göttingen, it turns out that it is precisely this subtle difference of linearizing the space of geodesics and not the shape space that successfully allows to discriminate genetically different Canadian black poplars by observation of leaf shape growth during a short time interval of the growing period. The research study presented here is fundamental for model building of leaf shape growth as well as for designing effective subsequent studies to investigate multiple endogenous and exogenous factors in leaf shape growth: for example, since the beginning of the last century it has been well known that the leaf shape of (genetically) identical trees varies along a climate gradient [e.g., Brenner (1902), Bailey and Sinnott (1915), Royer et al. (2009)]. Since Wolfe (1978) this relationship has been successfully exploited for paleoclimate reconstruction resulting in the "Climate Leaf Analysis Multivariate Program" [CLAMP, Wolfe (1993)]. Naturally, the underlying studies have been based on the shape of mature leaves; little is known about the temporal evolution of shape along a climate gradient. The research presented here indicates that a study involving only very few measurements of growing leaves may allow for a fairly good reconstruction and analysis of growth patterns, further elucidating the relationship of climate and leaf shape.

This paper is organized in a theoretical and an applied part.

The theoretical first part consisting of the following two sections establishes the statistical theory for the two types of means. In Section 2, after a brief review of Kendall's space of planar shapes, the concept of a Fréchet mean is extended to the space of geodesics while the underlying random deviates assume values in the shape space. Strong consistency in the sense of Ziezold (1977) a well as in the sense of Bhattacharya and Patrangenaru (2003) are established. In the Appendix, it shown that the original arguments can be extended nearly one-to-one to the general case considered here. In order to apply the central limit theorem (CLT) of Huckemann (2011c), smoothness in geodesics of the square of the canonical distance between shapes and geodesics for geodesics close to the data is established. Then in Section 3, smoothness is shown for the square of a metric of Ziezold type [cf. Huckemann (2011c)] for the space of geodesics leading to the other CLT. Finally, after establishing an explicit method for optimal positioning a fast algorithm

for the computation of a mean geodesic of geodesics is derived. An algorithm for the mean geodesic of shapes has been derived earlier [Huckemann and Hotz (2009)].

The applied second part introduces the leaf shape data considered, the driving questions from forest biometry, statistical tests and some answers through the data analysis. In Section 4, the problem of discrimination by short growth observations is discussed. In particular, the relevance of the geodesic hypothesis from Le and Kume (2000) is noted for the devising of statistical tests in Section 5. These are evaluated in Section 6 showing that only the *test for common geodesics* can establish the validity of the geodesic hypothesis and the discrimination of genetically different trees on the basis of observations of brief leaf shape growth. Section 7 concludes with a discussion and gives an outlook.

2. The first geodesic principal component for planar shape spaces. Throughout this work, $\mathbb{E}(Y)$ denotes the classical expectation of a random variable Y in a Euclidean space $\mathbb{R}^D$, $D \in \mathbb{N}$. A *distance* δ on a topological space Γ is a continuous mapping $\delta : \Gamma \times \Gamma \to [0, \infty)$ that vanishes on the diagonal $\{(\gamma, \gamma) : \gamma \in \Gamma\}$; in contrast to a metric, δ is neither required to be nonzero off the diagonal, to be symmetric nor to satisfy the triangle inequality.

Kendall's planar shape spaces. In the statistical analysis of similarity shapes based on landmark configurations, geometrical m-dimensional objects (usually $m = 2, 3$) are studied by placing $k > m$ *landmarks* at specific locations of each object, cf. Figure 1 on page 1109. Each object is then described by a matrix in the space $M(m, k)$ of $m \times k$ matrices, each of the k columns denoting an m-dimensional landmark vector. The usual inner product is denoted by $\langle x, y \rangle :=$ tr(xy^T) giving the norm $\|x\| = \sqrt{\langle x, x \rangle}$. For convenience and without loss of generality for the considerations below, only *centered* configurations are considered. Centering can be achieved by multiplying with a sub-Helmert matrix from the right, yielding a configuration in $M(m, k - 1)$. For this and other centering methods cf. Dryden and Mardia (1998), Chapter 2. Excluding also all configurations with all landmarks coinciding gives the space of *configurations*

$$F_m^k := M(m, k - 1) \setminus \{0\}.$$

Since only the similarity shape is of concern, in particular we are not interested in size, we may assume that all configurations are contained in the *pre-shape sphere* $S_m^k := \{x \in M(m, k - 1) : \|x\| = 1\}$. Then, all normalized configurations that are related by a rotation from the special orthogonal group SO(m) form the equivalence class of a *shape*

$$[x] = \{gx : g \in \text{SO}(m)\}$$

and the canonical quotient is *Kendall's shape space*

$$\Sigma_m^k := S_m^k / \text{SO}(m) = \{[x] : x \in S_m^k\} \qquad \text{with canonical projection } \mathfrak{p} : S_m^k \to \Sigma_m^k.$$

In this paper, we restrict ourselves to planar configurations, that is, to the case of $m = 2$. Then, complex notation comes in handy. For a detailed discussion of the following setup, cf. Kendall (1984), Kendall (1989) as well as Kendall et al. (1999). We take the notation from Huckemann and Hotz (2009). Identify F_2^k with $\mathbb{C}^{k-1} \setminus \{0\}$ such that every landmark column corresponds to a complex number. This means in particular that $z \in \mathbb{C}^{k-1}$ is a complex row(!)-vector. With the Hermitian conjugate, $a^* = (\bar{a}_{kj})$ of a complex matrix $a = (a_{jk})$ the pre-shape sphere S_2^k is identified with $\{z \in \mathbb{C}^{k-1} : zz^* = 1\}$ on which SO(2) identified with $S^1 = \{\lambda \in \mathbb{C} : |\lambda| = 1\}$ acts by complex scalar multiplication. Then the well-known Hopf–Fibration mapping to complex projective space gives $\Sigma_2^k = S_2^k / S^1 = \mathbb{C}P^{k-2}$.

The spaces of geodesics. Note that every geodesic can be parametrized by unit speed, which we assume in the following. Every great circle $\gamma(t) = x \cos t + v \sin t$, $x, v \in S_2^k$, $\langle x, v \rangle = 0$ is a geodesic on S_2^k, the space of geodesics is denoted by $\Gamma(S_2^k)$. A great circle is called a *horizontal great circle* if additionally $\langle ix, v \rangle = 0$, the space of *horizontal great circles* is denoted by $\Gamma^H(S_2^k)$. It is well known [e.g., Kendall et al. (1999), Huckemann and Hotz (2009)] that this space projects to the space $\Gamma(\Sigma_2^k)$ of geodesics of the shape space via

$$\Gamma(\Sigma_2^k) = \{\mathfrak{p} \circ \gamma : \gamma \in \Gamma^H(S_2^k)\}.$$

Then, with the *Stiefel manifold* (giving all great circles)

$$O_2(2, k - 1) = \{(x, v) \in F_2^k \times F_2^k : \langle x, x \rangle = 1 = \langle v, v \rangle, \langle x, v \rangle = 0\}$$

every tuple in the implicitly defined submanifold (additionally requiring horizontality)

$$O_2^H(2, k - 1) := \{(x, v) \in O_2(2, k - 1) : \langle x, iv \rangle = 0\}$$

corresponds to an element in $\Gamma^H(S_2^k)$. Several tuples, however, may determine the same geodesic. To this end, consider the action of the orthogonal group $O(2)$ and S^1 from right and left, respectively, by $(x, v) \mapsto h_t(x, v)g_{\phi,\varepsilon}$ with

$$g_{\phi,\varepsilon} = \begin{pmatrix} \cos\phi & -\varepsilon \sin\phi \\ \sin\phi & \varepsilon \cos\phi \end{pmatrix} \in O(2), \qquad h_t = e^{it} \in S^1,$$

for $\phi, t \in [0, 2\pi)$ and $\varepsilon = \pm 1$ defined by

$$(x, v)g_{\phi,\varepsilon} = (x \cos\phi + v \sin\phi, v\varepsilon \cos\phi - x\varepsilon \sin\phi),$$

$$h_t(x, v) = (e^{it}x, e^{it}v).$$

For a manifold M with a group K acting from the right and a group G acting from the left, denote by

$$K \setminus M / G = \{[P], P \in M\} \qquad \text{where } [P] = \{gPk : g \in G, k \in K\}$$

the canonical double quotient [e.g., Terras (1988)]. With this notation, we take the following from Huckemann and Hotz (2009).

THEOREM 2.1.   *The space of point sets of all geodesics on planar shape space can be given the canonical structure*

$$\Gamma(\Sigma_2^k) \cong O(2) \setminus O_2^H(2, k-1)/S^1$$

of a compact manifold of dimension $4k - 10$.

In analogy to the naming of the *pre-shape sphere* S_2^k, call $O_2^H(2, k-1)$ the space of *pre-geodesics*.

A simpler argument yields $\Gamma(S_2^k)$ as the compact manifold

$$(1) \qquad\qquad \Gamma(S_2^k) \cong O(2) \setminus O_2(2, k-1).$$

Distance from shapes to geodesics and between geodesics.   The spherical distance $r(p, \gamma) = \arccos \sqrt{\langle p, x \rangle^2 + \langle p, v \rangle^2}$ of a point p to the geodesic γ defined by $(x, v) \in O_2^H(2, k-1)$ naturally defines a distance

$$\rho([p], \mathfrak{p} \circ \gamma) = \min_{e^{it} \in S^1} r(e^{it} p, \gamma)$$

of the shape $[p]$ to the geodesic $\mathfrak{p} \circ \gamma$ in the shape space. For a shape $[p] \in \Sigma_2^k$ denote by

$$\Gamma_{[p]}^{\pi/4} = \left\{ \gamma \in \Gamma(\Sigma_2^k) : \rho([p], \gamma) < \frac{\pi}{4} \right\}$$

the open set of geodesics closer to $[p]$ than $\pi/4$. The proof of the following Theorem 2.2 is deferred to Appendix B.

THEOREM 2.2.   *For fixed $p \in S_2^k$ the function*

$$\gamma \mapsto \rho([p], \gamma)^2$$

is smooth on $\Gamma_{[p]}^{\pi/4}$.

In order to measure the distance between geodesics equip $O(2) \setminus O_2^H(2, k-1)$ with a suitable Riemannian structure—two of such structures are straightforward, cf. Edelman, Arias and Smith (1998), or more simply, embed $O_2^H(2, k-1)$ in a Euclidean space and consider the quotient distance w.r.t. to the corresponding extrinsic metric. More precisely, we generalize a setup introduced by Ziezold (1994) on the quotient Σ_2^k.

DEFINITION 2.3.   The Euclidean distance

$$d(P, Q) := \sqrt{\|x - y\|^2 + \|v - w\|^2}$$

for $P = (x, v)$, $Q = (y, w) \in O_2^H(2, k-1) \subset F_2^k \times F_2^k$ defines the canonical quotient distance

$$\delta([P], [Q]) := \min_{\substack{h, h' \in O(2), \\ g, g' \in S^1}} d(g Ph, g' Qh')$$

for $[P][Q] \in \Gamma(\Sigma_2^k)$ called the *Ziezold distance* on $\Gamma(\Sigma_2^k)$.

The mean geodesic of shapes. In earlier work [Huckemann, Hotz and Munk (2010b)] establishing a general framework for geodesic principal component analysis, the mean geodesic of shapes has been called a *first geodesic principal component*.

DEFINITION 2.4. Suppose that $X, X_1, \ldots, X_n$ are i.i.d. random pre-shapes mapping from an abstract probability space $(\Omega, \mathcal{A}, \mathcal{P})$ to S_2^k equipped with its Borel σ-algebra. For $\omega \in \Omega$ call the geodesics $\gamma_n(\omega), \gamma^* \in \Gamma(\Sigma_2^k)$ the *first sample and population geodesic principal component* (GPC) of the sample $[X_1(\omega)], \ldots, [X_n(\omega)]$ and $[X]$, respectively, if

$$\sum_{j=1}^{n} \rho([X_j(\omega)], \gamma_n(\omega))^2 = \min_{\gamma \in \Gamma(\Sigma_2^k)} \sum_{j=1}^{n} \rho([X_j(\omega)], \gamma)^2 \qquad \text{for all } \omega \in \Omega,$$

$$\mathbb{E}(\rho([X], \gamma^*)^2) = \min_{\gamma \in \Gamma(\Sigma_2^k)} \mathbb{E}(\rho([X], \gamma)^2).$$

The random set of all sample GPCs is denoted by $E_n^{(\rho)}(\omega)$, $E^{(\rho)}([X])$ is the set of all population GPCs.

THEOREM 2.5 (Asymptotics for the mean geodesic of shapes). *For i.i.d. random pre-shapes $X, X_1, \ldots, X_n$, the set of first sample GPCs $E_n^{(\rho)}(\omega)$ is a uniformly strongly consistent estimator of the set of first population GPCs $E^{(\rho)}([X])$ in the sense that for every $\varepsilon > 0$ and a.s. for every $\omega \in \Omega$ there is a number $n(\varepsilon, \omega) \in \mathbb{N}$ such that*

$$\bigcup_{j=n}^{\infty} E_j^{(\rho)}(\omega) \subset \{\gamma \in \Gamma(\Sigma_2^k) : \delta(\gamma, E^{(\rho)}([X])) \le \varepsilon\}.$$

Moreover, if $E^{(\rho)}([X])$ contains a unique element γ^* contained in

$$\bigcap_{p \in \mathrm{Supp}(X)} \Gamma_{[p]}^{\pi/4}$$

with the support $\mathrm{Supp}(X)$ of X, if $\gamma_n \in E_n^{(\rho)}(\omega)$ is a measurable selection and $x = \phi(\gamma) \in \mathbb{R}^{4k-10}$ are local coordinates near γ^* with $\phi(\gamma^*) = 0$, then

$$A\sqrt{n}\phi(\gamma_n) \to \mathcal{N}(0, \Sigma) \qquad \text{in distribution}$$

with the $(4k - 10)$-*dimensional normal distribution* $\mathcal{N}(0, \Sigma)$ *with zero mean and covariance matrix* $\Sigma = \text{COV}(\text{grad}_x \rho(X, \gamma^*)^2)$ *and* $A = \mathbb{E}(H_x \rho(X, \gamma^*)^2))$. *Here, grad_x and H_x denote the gradient and Hessian of ρ^2, respectively, w.r.t. the coordinate x.*

PROOF. The assertion of strong consistency is a consequence of the general Theorem A.3 and Theorem A.4 in the Appendix. Since $\rho([X], \gamma)^2$ is smooth in γ as long as γ is closer to $[X]$ than $\pi/4$, by Theorem 2.2, the assertion of the central limit theorem (CLT) follows from the CLT of Huckemann [(2011c), Theorem A.1] since $\Gamma(\Sigma_2^k)$ is compact. □

REMARK 2.6. For practical applications of Theorem 2.5, in a given chart A and Σ could be estimated by classical numerical and multivariate methods. Alternatively, estimates can be obtained simply from the data's covariance in a chart around a sample mean. In particular, in case of nonsingular A, that covariance tends asymptotically to $A^{-1}\Sigma(A^{-1})^T$.

Uniqueness and location of the first GPC. The hypothesis of a unique first GPC is essential for the following framework. Clearly, that hypothesis translates to an anisotropy condition on the random shape. For example, on the basis of the geodesic hypothesis for biological growth as detailed in Section 4, we may assume uniqueness in the application in Section 6. The development of a test for specific anisotropy would certainly be of merit for other potential applications. By definition, every first GPC will be close to the support of $[X]$. Data analysis and numerical simulations show that the intrinsic mean is usually very close to the first GPC [e.g., Huckemann and Hotz (2009), Huckemann, Hotz and Munk (2010b)]. Moreover, for sufficient concentration, the intrinsic mean is unique and contained in a ball around the support of radius $\pi/4$ [cf. Afsari (2011), Kendall (1990), Le (2001)]. Certainly, further research is necessary to tackle questions of uniqueness and location.

3. The Ziezold mean of a random first GPC. We are now in the situation of having samples of first sample GPCs and to determine their Fréchet mean w.r.t. to some distance. In order to apply a CLT, we are aiming for a mean in a smooth sense. It turns out that the comparatively simple Ziezold distance features the desired smoothness except at cut points.

Recall that on a Riemannian manifold the *cut locus* $C(p)$ of p comprises all points q such that the extension of a length minimizing geodesic joining p with q is no longer minimizing beyond q. If $q \in C(p)$ or $p \in C(q)$, then p and q are *cut points*. For example, on a sphere, antipodals are cut points.

THEOREM 3.1. *The following hold*:

  (i) *the action of S^1 and $O(2)$ is isometric with respect to d, that is, $d(P, Q) = d(gPh, gQh)$ for all $P, Q \in O_2^H(2, k-1)$ and all $g \in S^1, h \in O(2)$;*

(ii) δ *is a metric on* $\Gamma(\Sigma_2^k)$ *and* δ^2 *is smooth except at cut points*.

PROOF.   Property (i) is easily verified, in fact, the left-action of S^1 and right-action of $O(2)$ are even isometric on the ambient $\mathbb{C}^{k-1} \times \mathbb{C}^{k-1}$. Moreover on $F_2^k \times F_2^k$, the isotropy groups are $\{(g_{0,1}, h_0), (g_{\pi,1}, h_\pi)\}$. For this reason, in consequence of the Principal Orbit theorem [e.g., Bredon (1972), Chapter IV.3], δ extends to the natural geodesic quotient metric on the manifold $O(2) \backslash (F_2^k \times F_2^k)/S^1$. Hence, in particular, δ^2 is smooth on the embedded submanifold $\Gamma(\Sigma_2^k)$ unless evaluated at cut points. As another consequence, since the extension of δ is a metric, so is δ, yielding (ii). $\square$

In view of the application in Section 6, we now consider samples of independent random geodesics obtained from not necessarily independent shapes as typically occur during observation of growth. In particular, the test for common geodesics devised in Section 5 relies on the following Theorem 3.3.

DEFINITION 3.2 (The mean geodesic of geodesics).   Call $\gamma^* \in \Gamma(\Sigma_2^k)$:

a population *Ziezold mean geodesic* of a random geodesic Ξ if

$$\mathbb{E}(\delta(\Xi, \gamma^*)^2) = \min_{\gamma \in \Gamma(\Sigma_2^k)} \mathbb{E}(\delta(\Xi, \gamma)^2),$$

a sample *Ziezold mean geodesic* of random geodesics $\Xi_1, \ldots, \Xi_n$ if

$$\sum_{j=1}^n \delta(\Xi_j(\omega), \gamma^*)^2 = \min_{\gamma \in \Gamma(\Sigma_2^k)} \sum_{j=1}^n \delta(\Xi_j(\omega), \gamma)^2.$$

The sets of population and sample Ziezold mean geodesics are denoted by $E^{(\delta)}(\Xi)$ and $E_n^{(\delta)}(\omega)$, respectively.

THEOREM 3.3 (Asymptotics for the mean geodesic of geodesics).   *For i.i.d. random geodesics* $\Xi, \Xi_1, \ldots, \Xi_n$ *the set of sample Ziezold mean geodesics* $E_n^{(\delta)}(\omega)$ *is a uniformly strongly consistent estimator of the set of population Ziezold mean geodesics* $E^{(\delta)}(\Xi)$ *in the sense that for every* $\varepsilon > 0$ *and a.s. for every* $\omega \in \Omega$ *there is a number* $n(\varepsilon, \omega) \in \mathbb{N}$ *such that*

$$\bigcup_{j=n}^\infty E_j^{(\delta)}(\omega) \subset \{\gamma \in \Gamma(\Sigma_2^k) : \delta(\gamma, E^{(\delta)}(\Xi)) \leq \varepsilon\}.$$

If $E^{(\delta)}(\Xi)$ *contains a unique element* γ^*, $\gamma_n \in E_n^{(\delta)}(\omega)$ *is a measurable selection and* $x = \phi(\gamma) \in \mathbb{R}^{4k-10}$ *are local coordinates near* γ^* *with* $\phi(\gamma^*) = 0$, *and if*

$$\mathrm{supp}(\Xi) \subset \{\gamma \in \Gamma(\Sigma_2^k) : \delta(\gamma, C(\gamma^*)) > \varepsilon\}$$

for some $\varepsilon > 0$ with the support $\mathrm{supp}(\Xi)$ *of* Ξ *in case of* $\varnothing \neq C(\gamma^*) \subset \Gamma(\Sigma_2^k)$, *then*

$$A\sqrt{n}\phi(\gamma_n) \to \mathcal{N}(0, \Sigma) \qquad \text{in distribution}$$

with the $(4k - 10)$*-dimensional normal distribution* $\mathcal{N}(0, \Sigma)$ *with zero mean and covariance matrix* $\Sigma = \mathrm{COV}(\mathrm{grad}_x \delta(\Xi, \gamma^*)^2)$ *and* $A = \mathbb{E}(H_x \delta(\Xi, \gamma^*)^2))$. *Here*, grad_x *and* H_x *denote the gradient and Hessian of* δ^2, *respectively, w.r.t. the coordinate* x.

PROOF.   Since δ is a metric by Theorem 3.1, the assertion on strong consistency is a consequence of Ziezold (1977) as Bhattacharya and Patrangenaru [(2003), Remark 2.5] teach. Since δ is neither an intrinsic nor an extrinsic metric, the CLT of Bhattacharya and Patrangenaru (2005) cannot be applied. Rather, The assertion of the CLT follows from the more general Huckemann [(2011c), Theorem A.1], since by Theorem 3.1, δ^2 is smooth near γ^* by hypothesis and $\Gamma(\Sigma_2^k)$ is compact.   □

For practical applications of Theorem 3.3, proceed as detailed in Remark 2.6.

For $P, Q \in O_2^H(2, k - 1)$, $g \in S^1$ and $h \in O(2)$ call gQh is in *optimal position* to P, if $d(P, gQh) = \delta([P], [Q])$. Since both groups $O(2)$ and S^1 are compact, given $P \in O_2(2, k - 1)$, every $Q \in O_2(2, k - 1)$, can be placed into optimal position to P. Moreover, if $[P^*]$ is the unique Ziezold mean geodesic of sampled geodesics $[P_1], \ldots, [P_n]$ then P^* is the extrinsic mean of the $g_j P_j h_j$ placed into optimal position to P^*, $g_j \in S^1$, $h_j \in O(2)$, $j = 1, \ldots, n$, that is,

$$P^* = \underset{P \in O_2(2, k-1)}{\arg\min} \sum_{j=1}^{n} \min_{\substack{h_j \in O(2), \\ g_j \in S^1}} d(P, g_j P_j h_j),$$

cf. Huckemann (2011b). The extrinsic mean then is the orthogonal projection to $O_2^H(2, k - 1)$ of the classical Euclidean mean in ambient $M(2, k - 1) \times M(2, k - 1)$, cf. Hendriks and Landsman (1998), Bhattacharya and Patrangenaru (2003).

In the first step, we solve the problem of optimally positioning analytically, in the second step we compute the orthogonal projection. Based on the two, the algorithm of Ziezold (1994) is adapted, to compute the Ziezold mean geodesic.

THEOREM 3.4.   *Let* $P = (x, v)$, $Q = (y, w) \in O_2^H(2, k - 1)$ *and define*

$$A := \langle x, y \rangle + \varepsilon \langle v, w \rangle, \qquad B := \langle x, w \rangle - \varepsilon \langle v, y \rangle,$$

$$C := \langle x, iy \rangle + \varepsilon \langle v, iw \rangle, \qquad D := \langle x, iw \rangle - \varepsilon \langle v, iy \rangle.$$

Then, for $g_{\phi, \varepsilon}, h_t$ *putting* Q *into optimal position* $g_t Q h_{\phi, \varepsilon}$ *to* P, *it is necessary that*

$$\tan \phi = \frac{B + D \tan t}{A + C \tan t}$$

and that t satisfies:

  (i)

$$\tan t = \alpha \pm \sqrt{\alpha^2 + 1} \qquad with \; \alpha = \frac{C^2 + D^2 - A^2 - B^2}{2(AC + BD)}$$

in case of $AC + BD \neq 0$;

  (ii) $t = 0$ *in case of* $AC + BD = 0 \neq A^2 + B^2 - C^2 + D^2$.

$-\pi/2 \le t < \pi/2$ *may be arbitrary in case of* $AC + BD = 0 = A^2 + B^2 - C^2 + D^2$.

  PROOF.

$$d(P, g_t Q h_{\phi,\varepsilon})^2 = 4 - 2\big(\cos\phi(\langle x, e^{it}y\rangle + \varepsilon\langle v, e^{it}w\rangle)$$
$$+ \sin\phi(\langle x, e^{it}w\rangle - \varepsilon\langle v, e^{it}y\rangle)\big)$$

gives

$$\frac{4 - d(P, g_t Q h_{\phi,\varepsilon})^2}{2} = A\cos\phi\cos t + B\sin\phi\cos t + C\cos\phi\sin t$$
$$+ D\sin\phi\sin t.$$

For fixed ϕ, a necessary condition for $t = t(\phi)$ to maximize the above r.h.s. is that

$$\tan t = \frac{C\cos\phi + D\sin\phi}{A\cos\phi + B\sin\phi} = \frac{C + D\tan\phi}{A + B\tan\phi}.$$

Similarly, a necessary condition for $\phi = \phi(t)$ is that

$$\tan\phi = \frac{B\cos t + D\sin t}{A\cos t + C\sin t} = \frac{B + D\tan t}{A + C\tan t}.$$

Letting $\zeta = \tan t$, $\eta = \tan\phi$, we obtain

$$\zeta = \frac{C + D\eta}{A + B\eta} = \frac{C(A + C\zeta) + D(B + D\zeta)}{A(A + C\zeta) + B(B + D\zeta)}$$

and, equivalently

$$(AC + BD)\zeta^2 + (A^2 + B^2)\zeta = (C^2 + D^2)\zeta + (AC + BD),$$

yielding the assertion. □

  THEOREM 3.5.   *Suppose that* $P = (x, v) \in F_2^k \times F_2^k$, *then* $(\zeta, \eta) \in O_2^H(2, k - 1)$ *is the orthogonal projection of P to* $O_2^H(2, k - 1)$ *if and only if*

$$\zeta = \frac{1}{\langle x, \zeta\rangle}(x - \langle x, \eta\rangle\eta - \langle x, i\eta\rangle i\eta),$$

$$\eta = \frac{1}{\langle v, \eta\rangle}(v - \langle v, \zeta\rangle\zeta - \langle v, i\zeta\rangle i\zeta),$$

ζ *is arbitrary in case of* $\langle x, \zeta\rangle = 0$, *and* η *is arbitrary in case of* $\langle v, \eta\rangle = 0$.

PROOF. Apply Lagrange minimization to $\|x - \zeta\|^2 + \|v - \eta\|^2$ for $\zeta, \eta \in F_2^k$ under the constraining condition $\Phi(\zeta, \eta) = 0$ for

$$\Phi(x, v) = \begin{pmatrix} 1 - \langle x, x \rangle \\ 1 - \langle v, v \rangle \\ 2\langle x, v \rangle \\ 2\langle x, iv \rangle \end{pmatrix}. \qquad \square$$

Algorithm to obtain a pre-geodesic of a Ziezold mean geodesic. Let $P_1, \ldots, P_J$ be a sample of pre-geodesics. Starting with an initial value $(x^{(0)}, v^{(0)}) = P^{(0)} := P_1$, say, obtain $P^{(n+1)} = (x^{(n+1)}, v^{(n+1)})$ from $P^{(n)} = (x^{(n)}, v^{(n)})$ for $n = 0, 1, \ldots$ by putting all $P_j (j = 1, \ldots, J)$ in optimal position $P_j^* = (y_j^*, w_j^*)$ to $P^{(n)}$ by computing the corresponding $\phi_j, t_j, \varepsilon_j$ from Theorem 3.4. Then set

$$(x, v) := \frac{1}{J} \left(\sum_{j=1}^J y_j^*, \sum_{j=1}^J w_j^* \right)$$

and let $P^{(n+1)}$ be the orthogonal projection of (x, v) to $O_2^H(2, k-1)$ from Theorem 3.5.

4. Leaf growth data and problem statement.

From leaf data to shape descriptors. We consider leaf shape data collected from two clones and a reference tree of black Canadian poplars at an experimental site at the University of Göttingen. These data are similar but different from the data reported on in Huckemann, Hotz and Munk (2010a) and Huckemann (2011a). They consist of the shapes of 21 leaves from clone 1 and of 11 leaves from clone 2 as well as of the shapes of 12 leaves from the reference tree, all of which have been recorded nondestructively over several days during a major portion of their growing period of approximately one month (the maximal number of observations is 17, the minimal 2 with a median of 13). The top row in Figure 1 shows typical contours of leaf growth. At each time point, from each leaf contour a quadrangular *landmark based configuration* has been extracted by placing one landmark at the petiole (where the stalk enters the leaf blade), one at the leaf tip (the endpoint of the main leaf vein) and two, each at the maximal extensions orthogonal to the line connecting petiole and tip, cf. the bottom image of Figure 1. These four landmarks encode in particular the information of length, width and vertical and horizontal assymetry. As detailed in Section 2, these landmarks additionally convey a correspondence of leaves with *shapes*, that is, points in the *shape space* Σ_2^4. This space is a non-Euclidean manifold and a special case of *Kendall's landmark based shape spaces*.
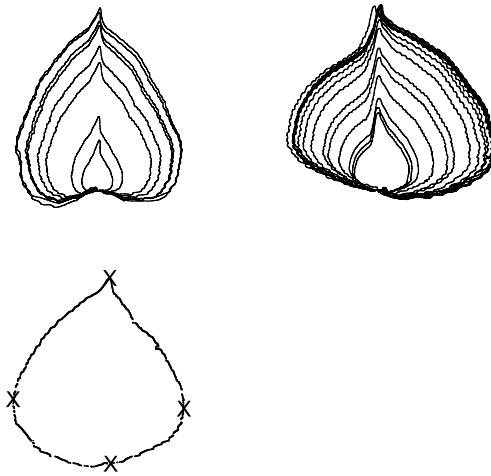
FIG. 1.   Top row: *typical leaf growth over a growing period of a reference tree* (left) *and one of the two clones* (right). Bottom left: *typical digitized leaf contour and landmarks of the corresponding quadrangular configuration at petiole*, *tip*, *and largest extensions orthogonal to the connecting line.*

The geodesic and parallel hypotheses for biological growth.   Investigating landmark based configurations of rat skulls, Le and Kume (2000) observed that:

> the shape change due to biological growth mainly follows a geodesic in Kendall's shape space.

In a research modeling, the growth of tree-stem disks as well as leaf growth this *geodesic hypothesis* has been corroborated by Hotz et al. (2010). Jupp and Kent (1987), Kent et al. (2001), Kenobi, Dryden and Le (2010), Kume, Dryden and Le (2007) have proposed more subtle models for shape growth essentially building on polynomials in Procrustes residuals (cf. Section 5 below).

Additionally, Morris et al. (2000) observed parallel growth patterns and coined the *parallel hypothesis*, stating that Procrustes residuals of related biological objects follow curves parallel in the Euclidean geometry of the tangent space at a Procrustes mean. In view of the geodesic hypothesis, we restrict those curves to straight lines, generally however, not mapping to geodesics (cf. Figure 2).

A brief discussion of the geodesic hypothesis.   In D'Arcy Thompson's seminal work [Thompson (1917)], biological form and growth of form has been explained by the invocation of the mathematical concept of force. More recently, the relationship between growth and energy minimization has been explored by Bookstein (1978). These works have led Le and Kume (2000) to the above hypothesis, being aware that, firstly, geodesics depend on a specific geometry of a shape space, and secondly, even though many paths of growth seem to follow geodesics, there are examples where the geodesic fit is rather poor [e.g., Kenobi, Dryden and Le

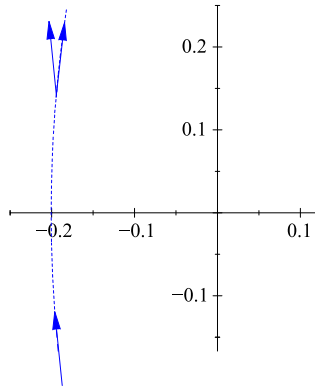FIG. 2. *Tangent space of the two-dimensional* Σ_2^3 (*obtained from* Σ_2^4 *by leaving out one landmark*) *under the inverse Riemann exponential at the intrinsic mean corresponding to the extent of the overall data* (*clones* $1+2$ *and reference tree*). *The left top vector is the affine parallel transport of the bottom vector in the Euclidean tangent space, the right top vector is its intrinsic parallel transplant along a geodesic* (*dashed*).

(2010)]. For example, for the space of planar triangles, the hyperbolic geometry of the complex upper half plane introduced by Bookstein (1986) seems just as natural as the spherical geometry of the complex projective space in one complex dimension [Kendall's shape space for planar triangles; cf. Kendall (1984)]. However, considering geodesics in Kendall's planar shape space as a rough working hypothesis, in particular for the leaf shapes in question, seems like a promising starting point for statistical investigation in the same way that approximate linearity has served statisticians well since the time of Gauss (or even earlier).

Problem statement.   As visible in Figure 1, the shape of leaves of the clones can usually be well discriminated from the shape of leaves of the reference tree by visual inspection. Following the geodesic hypothesis, the shape change under growth could be predicted from initial observations, ideally two initial observations would suffice. Since for the data at hand, the evolution of leaf contours have been followed elaborately along several time points, then the effort for future research could be cut down considerably. This leads to the following fundamental problem.

> Can leaf growth of genetically identical trees be predicted and discriminated from growth of genetically different trees on the basis of few initial measurements?

In this research, we restrict ourselves to measuring shape by four landmarks as detailed above.

5. Tests for shape dynamics.   The precise definitions used in this section can be found in Sections 2 and 3. For every leaf considered a *first geodesic principal*

component (GPC—a generalization to manifolds of a first principal component direction) is computed either from its first two shapes (then the GPC is just the geodesic connecting the two, the corresponding group is called "young") or from the rest of the shapes (that group is called "old") over the growing period [for an algorithm, see Huckemann and Hotz (2009)]. For two groups (young vs. young, young vs. old and old vs. old of different trees) to be tested for a common mean first GPC, the procedure detailed below produces data in a Euclidean space, namely in the tangent space of the space of geodesics at a mean geodesic. For comparison, a classical test for equality of mean shape as well as a test for equality of mean direction based on classical methodology described below, similarly produce data in a Euclidean space, namely the tangent space of the shape space at a mean shape. For all three procedures, within the respective Euclidean space, the corresponding tests then test for a common mean via the classical Hotelling T^2-test. Note that no two groups from the same tree are tested because of statistical dependence.

Robustness under nonnormality.   Recall that the *Hotelling T^2*-distribution is a generalization of a *student t*-distribution. As in the univariate case, the corresponding test statistic follows this distribution, if the coordinate data follow a multivariate normal distribution. Since the shape spaces considered are compact, obviously, we may never assume this central hypothesis. It is well known, however, that the corresponding statistic is robust to some extent under nonnormality, one condition being finite higher order moments. Clearly, this condition is met on a compact space. Even more, it is known that asymptotically the distribution of the corresponding statistic is unchanged under unequal change of covariances if the ratio of sample sizes tends to 1, for example, Lehmann (1997), page 462. The simulation in Figure 3 illustrates robustness for fairly small sample sizes: under nonnormality and unequal sample sizes (bottom right display), and under nonnormality and unequal covariances with equal sample sizes (top row). The test is not robust under nonnormality, however, if sample sizes and covariances differ considerably (bottom left display).

The test for common geodesics.   Every first GPC computed above determines a unique element in the space of geodesics $\Gamma(\Sigma_2^4)$ of Σ_2^4. Using the embedding of the space of pre-geodesics $O_2^H(2,3)$ into $\mathbb{C}^3 \times \mathbb{C}^3$ as detailed in Section 3, these elements are orthogonally projected to the tangent space of the two group's Ziezold mean [an element of the space of geodesics $\Gamma(\Sigma_2^4)$] thus giving data in a Euclidean space. The corresponding null hypothesis is then

> the temporal evolution of shape for every group follows a common geodesic.

In other words, if γ_i are the first GPCs of leaf growth of groups $i = 1, 2$ to be specified later, then the null hypothesis states $H_0 : \gamma_1 = \gamma_2$. This is a hypothesis on the mean geodesic of geodesics which can be tested by use of Theorem 3.3.
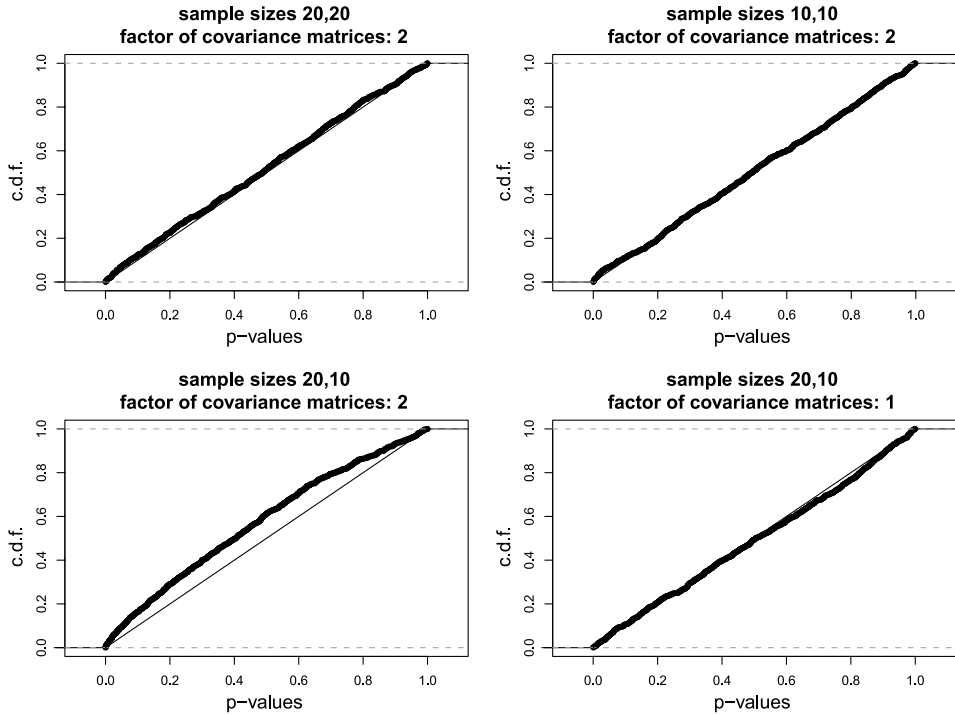
FIG. 3. *Hotelling's T^2-test under nonnormality. The cumulative distributions of the empirical test statistics have been generated with 1,000 repetitions of two groups of sizes as displayed in the respective headers, with three-dimensional deviates uniform in a 3D cube centered at the origin, whose covariance matrices (which are determined by the dimensions and rotation of the interval) are constant multiples of one another up to a random rotation (fixed for every display). The respective headers give the corresponding constant factors.*

Tests for common means. Following the classical scheme [e.g., Dryden and Mardia (1998), Chapter 7], all shapes of the two groups considered are projected to the tangent space of their overall Procrustes mean giving Procrustes residuals with the null hypothesis,

> the Procrustes tangent space coordinates of the temporal evolution of shape for every leaf have the same Euclidean mean.

A theoretical note on related tests. Taking instead the Procrustes residuals at the common Procrustes mean of the Procrustes means of every temporal evolution of shape would give a different test. If all leaves considered have equally many individual shapes, then this second test is very closely approximated by the first test. Otherwise, choosing suitable weights will give a close approximation. The goodness of the approximation can be numerically confirmed, it also follows from the fact that only one effect is tested; cf. Huckemann, Hotz and Munk (2010b),

pages 2–4. Similar tests for static shape utilize intrinsic or Ziezold means, respectively. In fact, for hypotheses on nondegenerate three-dimensional shapes one could only test hypotheses building on intrinsic or Ziezold means [because Procrustes means may lie outside the manifold part such that the CLT is not applicable, cf. Huckemann (2011b)].

Tests for common directions. Returning to the classical scheme, following Morris et al. (2000), we compute the Euclidean first principal component of each set of Procrustes residuals corresponding to the shapes of a single leaf's evolution pointing into the direction of growth. For the analysis of the unit length directions, we proceed as described above. Within the Procrustes paradigm, the residual tangent space coordinates of these directions at their common residual mean closer to the data are projected orthogonally to the Euclidean space of suitable dimension [cf. Huckemann (2011b)]. The null hypothesis is then

> the Procrustes tangent space coordinates of the temporal evolution of shape for every leaf share the same first Euclidean PC.

This is the version of the *parallel hypothesis* for this paper. If the static mean shapes of the two groups considered are different, then the common Procrustes mean depends on the ratio of the two sample sizes. Moreover, even for common static shape, the null hypothesis incorporates effects of curvature; cf. Figure 2.

6. Discriminating canadian black poplars by partial observation of leaf growth. For the following tests, as the first groups called "young" in the following, the first two initial shapes from every leaf considered have been taken and the unique geodesic joining the two has been computed. For the second groups, called "old" in the following, for every leaf considered the first GPC of the rest of the shapes has been computed, if the number of the rest of shapes exceeded 3. For this reason, the number of GPCs in some of these groups is possibly smaller then the number in corresponding groups of "young." To the respective groups, the three tests introduced in Section 5 have been applied. The results are reported in Table 1 in different order, however, since the tests for common geodesics and for common directions test for closely related concepts.

As visible in Table 1, the test for common geodesics (first data column) allows to discriminate very well genetically different trees from genetically identical trees by the observation of growth over restricted time intervals only. Due to curvature (cf. Figure 2) in the first box for genetically identical trees, whenever over these restricted time intervals the means differ (ultimate data column, that is, when the growth of young leaves is compared to the growth of old leaves), then the directions also differ highly significantly (middle data column). For genetically different trees (third box), all of the group means differ highly significantly (ultimate data column) and most of the directions as well. Note that genetically different young

Displaying p-values for several tests for the discrimination of clones from the reference tree via leaf growth ("young" denotes the dataset comprising the first initial two observations and "old" the dataset comprising the rest of the observations). For convenience, the sample size (number of different leaves followed over their growing period) of the corresponding data set is reported in parentheses

Dataset 1	Dataset 2	Geodesics	Directions	Means
Clone 1 young (21)	Clone 2 old (11)	0.75	1.6e–04	4.7e–08
Clone 1 young (21)	Clone 2 young (11)	0.97	0.82	0.71
Clone 2 young (11)	Clone 1 old (20)	0.066	0.0021	5.5e–07
Clone 1 old (20)	Clone 2 old (11)	0.17	0.21	0.71
Correct classification of clones				
At 95%-level		100.00%	50.00%	50.00%
At 99%-level		100.00%	50.00%	50.00%
Clone 1 young (21)	Reference young (12)	0.0012	0.65	6.5e–06
Clone 2 young (11)	Reference young (12)	0.0043	0.79	0.0015
Clone 1 young (21)	Reference old (9)	0.0067	0.0077	1.9e–07
Clone 2 young (11)	Reference old (9)	0.026	0.023	1.0e–05
Clone 1 old (21)	Reference young (12)	0.00022	0.0013	2.4e–06
Clone 2 old (11)	Reference young (12)	0.0092	0.014	0.0046
Clone 1 old (21)	Reference old (9)	0.087	0.023	0.0014
Clone 2 old (11)	Reference old (9)	0.021	0.018	0.0068
Correct classification of the reference tree				
At 95%-level		87.50%	75.00%	100.00%
At 99%-level		62.50%	25.00%	100.00%
Correct classification: clones vs. reference tree				
At 95%-level		93.75%	62.50%	75.00%
At 99%-level		81.25%	37.50%	75.00%

leaves cannot be discriminated by their directions. This can be explained by comparison with the bottom right display of Figure 4: leaf shapes of young leaves tend to be comparatively close to each other.

Figure 4 depicts the first two dominating coordinates (explaining between 80% and 90% of the total variation) of a tangent space projection of the overall dataset (young and old) for clones and the reference tree. In contrast to the test for common geodesics (cf. top display), almost all of the different groups (young vs. old of clone 1, clone 2 and the reference tree) can be discriminated by the directional data (bottom left display, e.g., the red and black crosses tend to lie on the left-hand side, the red and black circles on the right-hand side). In all of the first two displays (top and bottom left), for the clones, the variance of the "young" data appears slightly larger than the variance of the "old" data. For the reference tree,
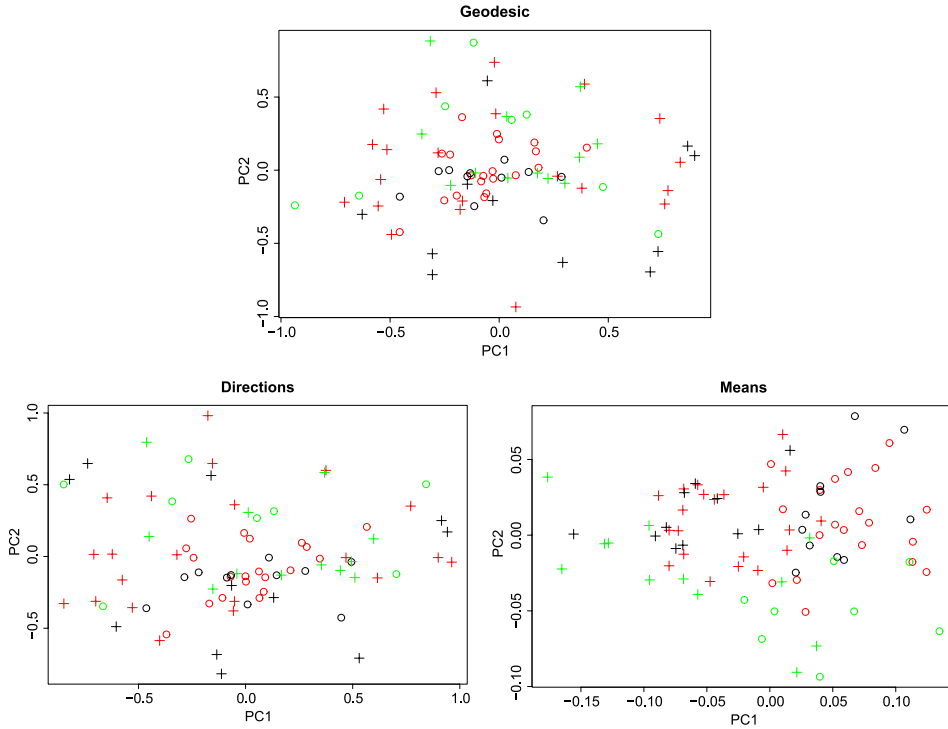
FIG. 4. *Projection of leaf shape growth from two clones* (*red for clone* 1 *and black for clone* 2) *and a reference tree* (*green*) *to the dominating coordinates of the corresponding tangent space at the overall mean as detailed in Section* 5. *Each cross represents a single leaf's initial shape evolution over two observations* (*young*), *each circle represents the rest of the leaf's shape evolution during its observed growing period* (*old*). Top: *GPCs projected to the tangent space at the Ziezold mean geodesic*. Bottom row: *all shapes have been projected to the tangent space of the overall Procrustes mean*. Bottom left: *unit directions of first Euclidean PCs*. Bottom right: *Euclidean means*.

however, this is not the case. This effect is not visible when observing means only (bottom right display). As discussed in Section 5, unequal covariances may be troublesome w.r.t. to the validity of the Hotelling T^2-test employed if sample sizes are not approximately equal. Considering in Table 1 only the samples of similar sizes 9, 11 and 12, however, comparable classification results are obtained.

Conclusion. From this study, we conclude the following.

(a) Clone and reference tree can be discriminated by partial observations of leaf shape growth not necessarily covering the same interval of the growing period via the test for common geodesics. This is not possible via a test for common means (due to temporal change of shape) or common directions (due to curvature).

(b) The "geodesic hypothesis" has been validated by use of the test for common geodesics, notably it could not have been validated using the test for common directions.

(c) For a statistical prediction of future leaf shape growth, two initial observations suffice.

Application. Let us elaborate on one consequence of conclusion (a). Suppose that we have several leaf shape growth data of clones and the reference tree of short but arbitrary time intervals (young, old, intermediate, etc.). Then most likely the test for common means will not be able to identify the clone from the reference tree, because the mean shapes of the different time intervals will most likely be different even for the same leaf considered. Similarly, due to curvature the test for common directions will fail, unless all data are jointly highly concentrated. It is only the test for common geodesics that may furnish the desired discrimination.

7. Discussion. The *geodesic hypothesis* of Le and Kume (2000)—its scope and limitations have been discussed in Section 4—has been corroborated in many scenarios of biological growth. In this paper, a *test for common geodesics* for this hypothesis has been devised and successfully applied to the problem of discriminating poplar leaf growth based on two initial observations only. For computational feasibility, not the concept of an intrinsic mean geodesic but rather that of a Ziezold mean geodesic has been employed. One can thus call the test devised a *semi-intrinsic test*. The semi-intrinsic test for common geodesics has been compared to a test for common directions building on directions in the space of Procrustes residuals [following Morris et al. (2000)]. This is essentially a *nonintrinsic test* because it linearizes the shape space and not the space of shape descriptors tested for. It turned out that for the discrimination task at hand, curvature present rendered this test ineffective. The author is not aware of any other test for the geodesic hypothesis in the literature.

In this work, we have considered two types of mean first GPCs, one defined by a sample of GPCs with underlying samples of random shapes, which—like growth patterns—are obviously dependent. For independent sampling, the other mean first GPC has been defined directly by the shape data. Since ρ (defining the latter) is different from δ (defining the former, cf. Section 2)—as a manifestation of "inconsistency" [cf. Kent and Mardia (1997), Huckemann (2011c)]—the limit of mean random geodesic of geodesics and the population geodesic of shapes may be different as well. Studying their relationship, however, may provide further insight.

In conclusion, let us ponder on extensions and generalizations of this research. One may view all shape descriptors as generalized Fréchet means on suitable spaces. For geodesics on Kendall's planar shape spaces, we have provided an explicit framework using a Ziezold mean geodesics which can be computed fairly easy. Straightforward but considerably more complicated is the use of intrinsic mean geodesics. At this point, we note that the space of generalized geodesics

on Kendall' shape space for dimension $m \geq 3$ ceases to be a manifold. Like a shape space, it can be viewed as the quotient of a Riemannian manifold modulo a Lie group action. In contrast to Kendall's shape spaces, the top space itself (a submanifold of a Grassmannian) admits two canonical geometries [cf. Edelman, Arias and Smith (1998)]. For the embedding underlying the Ziezold mean geodesic, we have used the simpler of the two. Possibly, this framework extends to other one-dimensional shape descriptors, such as arbitrary circles on spheres [cf. Jung, Foskey and Marron (2010)], or more generally, the family of constant curvature curves, as well as to higher-dimensional shape descriptors [for geodesic descriptors cf. Huckemann, Hotz and Munk (2010b), for nongeodesic descriptors cf. Jung, Foskey and Marron (2011)]. Thus, (semi)-intrinsic inference on any of such descriptors may be possible.

One final word of caution: even for fairly simple spaces such as the torus or the surface of an infinite cylinder, the canonical topology of the space of geodesics is non-Hausdorff [cf. Beem and Parker (1991)] and thus may not admit any meaningful statistical descriptors.

APPENDIX A: STRONG CONSISTENCY

For this section, suppose that $X, X_1, X_2, \ldots$ are i.i.d. random elements mapping from an abstract probability space $(\Omega, \mathcal{A}, \mathcal{P})$ to a topological space Q equipped with its Borel σ-field; (P, d) denotes a topological space with distance d.

DEFINITION A.1. For a continuous function $\rho : Q \times P \to [0, \infty)$, define the *set of population Fréchet ρ-means of X in P* by

$$E^{(\rho)}(X) = \arg\min_{\mu \in P} \mathbb{E}(\rho(X, \mu)^2).$$

For $\omega \in \Omega$, denote by

$$E_n^{(\rho)}(\omega) = \arg\min_{\mu \in P} \sum_{j=1}^{n} \rho(X_j(\omega), \mu)^2$$

the *set of sample Fréchet ρ-means*.

By continuity of ρ, the mean sets are closed random sets. For our purpose here, we rely on the definition of *random closed sets* as introduced and studied by Choquet (1954), Kendall (1974) and Matheron (1975). Since their original definition for $P = Q$, $\rho = d$ a metric by Fréchet (1948) such means have found much interest.

We will work with the following two definitions of *strong consistency*, each has been coined as such for metrical Fréchet means by the respective authors.

DEFINITION A.2. Let $E_n^{(\rho)}(\omega)$ be a random closed set and $E^{(\rho)}$ a deterministic closed set in a space with distance, (P, d). We then say that:

(ZC) $E_n^{(\rho)}(\omega)$ is a *strongly consistent estimator in the sense of* Ziezold (1977) of $E^{(\rho)}$ if for almost all $\omega \in \Omega$

$$\bigcap_{n=1}^{\infty} \overline{\bigcup_{k=n}^{\infty} E_k^{(\rho)}(\omega)} \subset E^{(\rho)}.$$

(BPC) $E_n^{(\rho)}(\omega)$ is a *strongly consistent estimator in the sense of* Bhattacharya and Patrangenaru (2003) of $E^{(\rho)}$ if $E^{(\rho)} \neq \varnothing$ and if for every $\varepsilon > 0$ and almost all $\omega \in \Omega$ there is a number $n = n(\varepsilon, \omega) > 0$ such that

$$\bigcup_{k=n}^{\infty} E_k^{(\rho)}(\omega) \subset \{p \in P : d(E^{(\rho)}, p) \leq \varepsilon\}.$$

For quasi-metrical means on *separable* (i.e., containing a dense countable subset) quasi-metrical spaces, Ziezold (1977) proved (ZC). For metrical means on spaces that enjoy the stronger *Heine–Borel property* (i.e., that every bounded closed set is compact), Bhattacharya and Patrangenaru (2003) proved (BPC). One may argue from a statistical point of view that for "consistency" one would want to have equality of points, and if not possible, at least an equality of sets [the set of resdiual means always contains at least two antipodal points, e.g., Huckemann (2011b)], rather than an inclusion only. Even though it would be interesting to construct an example with strict inclusion, it seems that this case has no relevance in applications.

In order to generalize Ziezold's and the Bhattacharya–Patrangenaru Strong Consistency theorem, we introduce two properties. A *continuity* property in the second argument *uniform* over the first argument—a consequence of the triangle inequality if ρ is a quasi-metric—and a version of *coercivity* in the second argument—again valid if ρ is a quasi-metric:

$$(2) \quad \left.\begin{array}{l} \text{for every } x \in Q, p \in P \text{ and } \varepsilon > 0 \text{ there is a } \delta = \delta(\varepsilon, p) > 0 \\ \text{such that } |\rho(x, p') - \rho(x, p)| < \varepsilon \text{ for all } p' \in P \text{ with } d(p, p') < \delta, \end{array}\right\}$$

$$(3) \quad \left.\begin{array}{l} \text{there are } p_0 \in P \text{ and } C > 0 \text{ such that } \mathcal{P}\{\rho(X, p_0) < C\} > 0 \text{ and} \\ \text{that such that for every sequence } p_n \in P \text{ with } d(p_0, p_n) \to \infty \\ \text{there is a sequence } M_n \to \infty \text{ with } \rho(x, p_n) > M_n \text{ for all } x \in Q \\ \text{with } \rho(x, p_0) < C. \text{ Moreover, if } p_n \in P \text{ with } d(p^*, p_n) \to \infty \\ \text{for some } p^* \in P, \text{ then } d(p_0, p_n) \to \infty. \end{array}\right\}$$

THEOREM A.3 (Ziezold's strong consistency).   *Let* $\rho : Q \times P \to [0, \infty)$ *be a continuous function on the product of a topological space with a separable space with distance* (P, d). *Then strong consistency holds in the Ziezold sense* (ZC) *for the set of Fréchet ρ-means on P if:*

   (i) *X has compact support, or if*

(ii) $\mathbb{E}(\rho(X, p)^2) < \infty$ for all $p \in P$ and ρ is uniformly continuous in the second argument in the sense of (2).

PROOF.   Obviously under (i) for every $p \in P$ and $\varepsilon > 0$ we may assume that there is $\delta = \delta(\varepsilon, p)$ such that $|\rho(X, p') - \rho(X, p)| < \varepsilon$ a.s. if $d(p, p') < \delta$. With this is mind, it suffices to prove the assertion under (ii).

For $\omega \in \Omega$, $p \in P$ set

(4)
$$
\left.
\begin{array}{ll}
F_n(p) = \dfrac{1}{n} \sum_{j=1}^{n} \rho(X_j(\omega), p)^2, & F(q) = \mathbb{E}(\rho(X, p)^2), \\[2mm]
\ell_n = \inf_{p \in P} F_n(q), & \ell = \inf_{p \in P} F(p), \\[2mm]
E_n = \{p \in P : F_n(p) = \ell\}, & E = \{p \in P : F(p) = \ell\}.
\end{array}
\right\}
$$

We now follow the steps laid out in Ziezold (1977). Let $p_1, p_2, \ldots$ be dense in P. From the usual Strong Law of Large Numbers in $\mathbb{R}$ we have sets $A_k \subset \Omega$, $\mathcal{P}(A_k) = 1$ such that $F_n(p_k) \overset{n \to \infty}{\to} F(p_k)$ for every $k = 1, 2, \ldots$ and $\omega \in A_k$. Setting $A := \bigcap_{k=1}^{\infty} A_k$ we have hence for all $p = p_k, k = 1, 2, \ldots,$

(5)
$$
F_n(p) \overset{n \to \infty}{\to} F(p) \qquad \text{for all } \omega \in A, \mathcal{P}(A) = 1.
$$

Next, let $p, p' \in P$. Setting $f(q, p', p) := \rho(q, p') - \rho(q, p)$ we have then

(6)
$$
|F_n(p') - F_n(p)| \leq \frac{1}{n} \sum_{j=1}^{n} (\rho(X_j, p') + \rho(X_j, p))|\rho(X_j, p') - \rho(X_j, p)|
$$
$$
=
\begin{cases}
\dfrac{1}{n} \sum_{j=1}^{n} (2\rho(X_j, p') + f(X_j, p, p'))|f(X_j, p, p')|, \\[3mm]
\dfrac{1}{n} \sum_{j=1}^{n} (2\rho(X_j, p) + f(X_j, p', p))|f(X_j, p, p')|.
\end{cases}
$$

W.l.o.g. we may suppose that $p_k \to p \in P$. In consequence from the top line of (6) for $p' = p_k$,

$$
\frac{1}{n} \sum_{j=1}^{n} \rho(X_j(\omega), p_k)^2 - \frac{1}{n} \sum_{j=1}^{n} (2\rho(X_j, p_k) + |f(X_j, p, p_k)|)|f(X_j, p, p_k)|
$$

$$
\leq \frac{1}{n} \sum_{j=1}^{n} \rho(X_j, p)^2)
$$

$$
\leq \frac{1}{n} \sum_{j=1}^{n} \rho(X_j, p_k)^2 + \frac{1}{n} \sum_{j=1}^{n} (2\rho(X_j, p_k) + |f(X_j, p, p_k)|)|f(Y_j, p, p_k)|.
$$

Taking the expected value, that is, employing (5) at p_k, by hypothesis (i) or (ii) as explained in the beginning of the proof, for arbitrary $\varepsilon > 0$ we may assume that there is a $\delta > 0$ such that for all $d(p_k, p) < \delta$

$$\mathbb{E}(\rho(X, p_k)^2) - \big(2\mathbb{E}(\rho(X, p_k)) + \varepsilon\big)\varepsilon$$

$$\leq \liminf_{n \to \infty} \frac{1}{n} \sum_{j=1}^n \rho(X_j, p)^2 \leq \limsup_{n \to \infty} \frac{1}{n} \sum_{j=1}^n \rho(X_j, p)^2$$

$$\leq \mathbb{E}(\rho(X, p_k)^2) + \big(2\mathbb{E}(\rho(X, p_k)) + \varepsilon\big)\varepsilon.$$

Letting $\varepsilon \to 0$ we can choose a subsequence $p_{k_i} \to p$, hence, this yields the validity of (5) for all $p \in P$.

Let us now extend (5) to

$$(7) \quad F_n(p_n) \overset{n \to \infty}{\to} F(p) \qquad \text{for all sequences } p_n \to p \text{ and } \omega \in A, \mathcal{P}(A) = 1.$$

Utilizing the bottom line from (6) yields

$$|F_n(p_n) - F_n(p)| \leq \frac{1}{n} \sum_{j=1}^n \big(2\rho(X_j, p) + |f(X_j, p_n, p)|\big)|f(X_j, p, p_n)| \to 0.$$

Hence, as desired, $|F_n(p_n) - F(p)| \leq |F_n(p_n) - F_n(p)| + |F(p) - F_n(p)| \to 0$. Finally, let us show

$$(8) \qquad\qquad \text{if } \bigcap_{n=1}^{\infty} \overline{\bigcup_{k=n}^{\infty} E_n} \neq \varnothing \qquad \text{then } \ell_n \to \ell.$$

Note that, the assertion of the theorem is trivial in case of $\bigcap_{n=1}^{\infty} \overline{\bigcup_{k=n}^{\infty} E_n} = \varnothing$. Otherwise, for ease of notation let $B_n := \bigcup_{k=n}^{\infty} E_k$, $\overline{B_n} \searrow B := \bigcap_{n=1}^{\infty} \overline{B_n}$, $b \in B$. Then $b \in \overline{B_n}$ for all $n \in \mathbb{N}$. Hence, there is a sequence $b_n \in B_n$, $b_n \to b$. Moreover, there is a sequence k_n such that $b_n = p_{k_n} \in E_{k_n}$ for a suitable $k_n \geq n$. Then $\ell_{n_k} = F_{n_k}(p_{n_k}) \to F(b) \geq \ell$ by (7). On the other hand, by (5) for arbitrary fixed $p \in P$, there is a sequence $\varepsilon_n \to 0$ such that $F(p) \geq F_n(p) - \varepsilon_n \geq \ell_n - \varepsilon_n$. First, letting $n \to \infty$ and then considering the infimum over $p \in P$ yields

$$\ell \geq \limsup_{n \to \infty} \ell_n.$$

In consequence, $\ell_n \to \ell = F(b)$. In particular, we have shown that $b \in E \neq \varnothing$ thus completing the proof. $\square$

THEOREM A.4 (Bhattacharya–Patrangenaru's strong consistency). *Suppose that* (ZC) (*Ziezold's strong consistency*) *holds for a continuous function* $\rho : Q \times P \to [0, \infty)$ *on the product of a topological space with a space with distance* (P, d). *If additionally* $\varnothing \neq E^{(\rho)}$, $\bigcup_{n=1}^{\infty} E_n^{(\rho)}(\omega)$ *enjoys the Heine–Borel property for almost all* $\omega \in \Omega$ *and the coercivity condition* (3) *in the second argument is satisfied then the property of strong consistency* (BPC) *in the sense of Bhattacharya and Patrangenaru is valid.*

PROOF.   We use the notation of the previous proof. Consider a sequence $p_n \in E_n$ determined by

$$d(p_n, E) = \max_{p \in E_n} d(p, E) =: r_n.$$

If $r_n \not\to 0$, we can find a subsequence n_k such that $r_{n_k} \geq r_0 > 0$. In consequence, every accumulation point of p_{n_k} has positive distance to E, a contradiction to strong consistency. Hence either $r_n \to 0$ or there are no accumulation points.

We shall now rule out the case that there are no accumulation points. In view of the Heine–Borel property, this case can only occur for $r_n \to \infty$. Under condition (3) there is $p_0 \in P$, a subsequence $k(n)$ and for given n a subsequence $n_1, \ldots, n_{k(n)}$ of $1, 2, \ldots, n$ such that $\rho(X_{n_j}, p_0) < C$ for all $j = 1, \ldots, k(n)$, a.s., and

$$\frac{k(n)}{n} \to \mathcal{P}\{\rho(X, p_0) < C\} > 0.$$

Also, by hypothesis and condition (3), $d(p_0, p_n) \to \infty$. Moreover, by condition (3), there is a sequence $M_n \to \infty$ with

$$\ell_n = F_n(p_n) \geq \frac{1}{n} \sum_{j=1}^{k(n)} \rho(X_{n_j}(\omega), p_n)^2 > \frac{k(n)}{n} M_n^2 \to \infty \qquad \text{a.s.}$$

On the other hand, for any fixed $p \in E$, we have the Strong Law on $\mathbb{R}$, $\ell_n \leq F_n(p) \to F(p) = \ell$ a.s., yielding a contradiction. $\square$

APPENDIX B: PROOF OF THEOREM 2.2

Consider for fixed $p \in S_2^k$ the nonnegative $O(2)$-invariant smooth function on $O_2^H(2, k-1)$ defined by

$$f_p(x, v) := \left(\arccos\sqrt{\langle p, x\rangle^2 + \langle p, v\rangle^2}\right)^2.$$

Then the condition

$$f_p(\xi x, \xi v) = \min_{e^{it} \in S^1} f_p(e^{it}x, e^{it}v) = \arccos\left(\max_{e^{it} \in S^1}(\langle e^{it}p, x\rangle^2 + \langle e^{it}p, v\rangle^2)\right)$$

defines two smooth explicit branches $\pm\xi : O_2^H(2, k-1) \setminus M^0 \to S^1$ [cf. Huckemann and Hotz (2009), Theorem 4.3], outside the singularity set

$$M^0 = \{(x, v) \in O_2^H(2, k-1) : D(x, v) = 0 = A(x, v)^2 - B(x, v)^2\}$$

using the notation from Huckemann and Hotz (2009):

$$A(x, v)^2 = \langle x, p\rangle^2 + \langle v, p\rangle^2,$$
$$B(x, v)^2 = \langle x, ip\rangle^2 + \langle v, ip\rangle^2,$$
$$D(x, v) = 2(\langle x, p\rangle\langle x, ip\rangle + \langle v, p\rangle\langle v, ip\rangle).$$

One verifies that M^0 is bi-invariant, that is, invariant under the left action of S^1 and the right action of $O(2)$. Hence, on $O(2) \setminus (O_2^H(2, k-1) \setminus M^0)/S^1$, the square of the distance $\rho([p], [\gamma])$ agrees with the value of the bi-invariant function $(x, v) \mapsto f_p(\xi(x, v)x, \xi(x, v)v)$, hence

$$\mathfrak{p} \circ \gamma \mapsto \rho^2([p], \mathfrak{p} \circ \gamma)$$

is smooth on $O(2) \setminus (O_2^H(2, k-1) \setminus M^0)/S^1$.

Finally, we show that the geodesics in $\Gamma(\Sigma_2^k)$ determined by M^0 are at least $\pi/4$ away from $[p]$. Obviously, geodesics determined by $A(x, v)^2 = 0 = B(x, v)^2$ have distance $\pi/2$ to $[p]$. Suppose now that $(x, v) \in M^0$ with $A(x, v)^2 > 0$. W.l.o.g. assume that $\langle x, p \rangle \neq 0$. Then

$$B(x, v)^2 = \frac{\langle v, ip \rangle^2}{\langle x, p \rangle^2} A(x, v)^2,$$

which implies that $\langle v, ip \rangle^2 = \langle x, p \rangle^2$. In consequence, we have also $\langle v, p \rangle^2 = \langle x, ip \rangle^2$ and, in particular, $\text{sign}(\langle x, p \rangle \langle v, ip \rangle) = -\text{sign}(\langle x, ip \rangle \langle v, p \rangle) =: \varepsilon$. Then, we have for the shape distance ρ to $[p]$ of shapes along the geodesic γ through $[x]$ with initial velocity v at $[x]$ that

$$\cos \rho([p], \gamma(s)) = \max_{0 \leq t < 2\pi} \langle p \cos t + ip \sin t, x \cos s + v \sin s \rangle$$

$$= \max_{0 \leq t < 2\pi} (\langle x, p \rangle \cos(t - \varepsilon s) + \langle v, p \rangle \sin(t - \varepsilon s))$$

$$= \max(|\langle x, p \rangle|, |\langle v, p \rangle|)$$

is constant, giving as desired $\rho([p], \gamma) \geq \pi/4$.

REFERENCES

AFSARI, B. (2011). Riemannian L^p center of mass: Existence, uniqueness, and convexity. *Proc. Amer. Math. Soc.* **139** 655–673. MR2736346

BAILEY, I. W. and SINNOTT, E. W. (1915). A botanical index of cretaceous and tertiary climates. *Science* **41** 831–834.

BEEM, J. K. and PARKER, P. E. (1991). The space of geodesics. *Geom. Dedicata* **38** 87–99. MR1099923

BHATTACHARYA, R. and PATRANGENARU, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.* **31** 1–29. MR1962498

BHATTACHARYA, R. and PATRANGENARU, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Ann. Statist.* **33** 1225–1259. MR2195634

BOOKSTEIN, F. L. (1978). *The Measurement of Biological Shape and Shape Change*, 2nd ed. *Lecture Notes in Biomathematics* **24**. Springer, New York.

BOOKSTEIN, F. L. (1986). Size and shape spaces for landmark data in two dimensions (with discussion). *Statist. Sci.* **1** 181–222.

BREDON, G. E. (1972). *Introduction to Compact Transformation Groups. Pure and Applied Mathematics* **46**. Academic Press, New York. MR0413144

BRENNER, W. (1902). Klima und Blatt bei der Gattung Quercus. *Flora oder Allgemeine botanische Zeitun* **90** 114–160.

CHOQUET, G. (1954). Theory of capacities. *Ann. Inst. Fourier (Grenoble)* **5** 131–295. MR0080760

DRYDEN, I. L. and MARDIA, K. V. (1998). *Statistical Shape Analysis*. Wiley, Chichester. MR1646114

EDELMAN, A., ARIAS, T. A. and SMITH, S. T. (1998). The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.* **20** 303–353. MR1646856

FRÉCHET, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré Probab. Statist.* **10** 215–310. MR0027464

GOODALL, C. (1991). Procrustes methods in the statistical analysis of shape. *J. Roy. Statist. Soc. Ser. B* **53** 285–339. MR1108330

GOWER, J. C. (1975). Generalized Procrustes analysis. *Psychometrika* **40** 33–51. MR0405725

HENDRIKS, H. and LANDSMAN, Z. (1996). Asymptotic behavior of sample mean location for manifolds. *Statist. Probab. Lett.* **26** 169–178. MR1381468

HENDRIKS, H. and LANDSMAN, Z. (1998). Mean location and sample mean location on manifolds: Asymptotics, tests, confidence regions. *J. Multivariate Anal.* **67** 227–243. MR1659156

HOTZ, T., HUCKEMANN, S., GAFFREY, D., MUNK, A. and SLOBODA, B. (2010). Shape spaces for pre-alingend star-shaped objects in studying the growth of plants. *J. Roy. Statist. Soc. Ser. C* **59** 127–143.

HUCKEMANN, S. (2011a). Dynamic shape analysis and comparison of leaf growth. Unpublished manuscript. Available at arXiv:1002.0616v1.

HUCKEMANN, S. (2011b). On the meaning of mean shape. Unpublished manuscript. Available at arXiv:1002.0795v1.

HUCKEMANN, S. (2011c). Inference on 3D Procrustes means: Tree bole growth, rank-deficient diffusion tensors and perturbation models. *Scand. J. Statist.* To appear. DOI:10.1111/j.1467-9469.2010.00724.x.

HUCKEMANN, S. and HOTZ, T. (2009). Principal component geodesics for planar shape spaces. *J. Multivariate Anal.* **100** 699–714. MR2478192

HUCKEMANN, S., HOTZ, T. and MUNK, A. (2010a). Intrinsic MANOVA for Riemannian manifolds with an application to Kendall's space of planar shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32** 593–603.

HUCKEMANN, S., HOTZ, T. and MUNK, A. (2010b). Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. *Statist. Sinica* **20** 1–100. MR2640651

JUNG, S., FOSKEY, M. and MARRON, J. S. (2010). Comment on "Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions," by S. Huckemann, T. Hotz and A. Munk. *Statist. Sinica* **20** 63–65. MR2640653

JUNG, S., FOSKEY, M. and MARRON, J. S. (2011). Principal arc analysis on direct product manifolds. *Ann. Appl. Statist.* **5** 578–603.

JUPP, P. E. and KENT, J. T. (1987). Fitting smooth paths to spherical data. *J. Roy. Statist. Soc. Ser. C* **36** 34–46. MR0887825

KENDALL, D. G. (1974). Foundations of a theory of random sets. In *Stochastic Geometry* (*a Tribute to the Memory of Rollo Davidson*) 322–376. Wiley, London. MR0423465

KENDALL, D. G. (1984). Shape manifolds, Procrustean metrics, and complex projective spaces. *Bull. London Math. Soc.* **16** 81–121. MR0737237

KENDALL, D. G. (1989). A survey of the statistical theory of shape. *Statist. Sci.* **4** 87–99. MR1007558

KENDALL, W. S. (1990). Probability, convexity, and harmonic maps with small image. I. Uniqueness and fine existence. *Proc. London Math. Soc.* (3) **61** 371–406. MR1063050

KENDALL, D. G., BARDEN, D., CARNE, T. K. and LE, H. (1999). *Shape and Shape Theory*. Wiley, Chichester. MR1891212

KENOBI, K., DRYDEN, I. L. and LE, H. (2010). Shape curves and geodesic modelling. *Biometrika* **97** 567–584.

KENT, J. T. and MARDIA, K. V. (1997). Consistency of Procrustes estimators. *J. Roy. Statist. Soc. Ser. B* **59** 281–290. MR1436569

KENT, J. T., MARDIA, K. V., MORRIS, R. J. and AYKROYD, R. G. (2001). Functional models of growth for landmark data. In *Proceedings in Functional and Spatial Data Analysis* (K. V. Mardia and R. G. Aykroyd, eds.) 109–115. Leeds Univ. Press, Leeds.

KUME, A., DRYDEN, I. L. and LE, H. (2007). Shape-space smoothing splines for planar landmark data. *Biometrika* **94** 513–528. MR2410005

LE, H. (2001). Locating Fréchet means with application to shape spaces. *Adv. in Appl. Probab.* **33** 324–338. MR1842295

LE, H. and KUME, A. (2000). Detection of shape changes in biological features. *J. Microsc.* **200** 140–147.

LEHMANN, E. L. (1997). *Testing Statistical Hypotheses*, 2nd ed. Springer, New York. MR1481711

MATHERON, G. (1975). *Random Sets and Integral Geometry*. Wiley, New York.

MORRIS, R., KENT, J. T., MARDIA, K. V. and AYKROYD, R. G. (2000). A parallel growth model for shape. In *Proceedings in Medical Imaging Understanding and Analysis* (S. Arridge and A. Todd-Pokropek, eds.) 171–174. BMVA, Bristol.

ROYER, D. L., MEYERSON, L. A., ROBERTSON, K. M. and ADAMS, J. M. (2009). Phenotypic plasticity of leaf shape along a temperature gradient in Acer rubrum. *PLoS ONE* **4** e7653.

TERRAS, A. (1988). *Harmonic Analysis and Symmetric Spaces and Applications. II*. Springer, Berlin. MR0955271

THOMPSON, D. W. (1917). *On Growth and Form*. Cambridge Univ. Press, Cambridge.

WOLFE, J. (1978). A paleobotanical interpretation of tertiary climates in the northern hemisphere. *American Scientist* **66** 694–703.

WOLFE, J. (1993). A method of obtaining climatic parameters from leaf assemblages. U.S. Geological Survey Bulletin 2040, USGPO.

ZIEZOLD, H. (1977). On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. In *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the Eighth European Meeting of Statisticians* (*Tech. Univ. Prague, Prague*, 1974), *Vol. A* 591–602. Reidel, Dordrecht. MR0501230

ZIEZOLD, H. (1994). Mean figures and mean shapes applied to biological figure and shape distributions in the plane. *Biom. J.* **36** 491–510. MR1292314

INSTITUT FÜR MATHEMATISCHE STOCHASTIK
GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN
GOLDSCHMIDTSTR. 7
D-37077 GÖTTINGEN
GERMANY
E-MAIL: huckeman@math.uni-goettingen.de