A review on combination-based tests for shape analysis

Chiara Brombin[1]   Luigi Salmaso[2]

Abstract

This paper reviews main features of the combination-based approach for shape analysis. It is worth noting that inference in shape analysis is a crucial point and robust testing is much needed. Our nonparametric permutation-based approach provides a suitable and powerful testing also in presence of many correlated landmarks and a limited sample size. An important feature of combination-based tests is the finite sample consistency property (Pesarin and Salmaso, 2010 and Brombin and Salmaso, 2013) which allows to gain power in the testing procedure by increasing the number of variables rather than the sample size, provided that added variables yield additional information.

Key Words: Correlated data, nonparametric combination, permutation tests

1. Introduction and motivation

Statistical shape analysis is considered a cross-disciplinary field characterized by flexible theory and techniques, potentially adaptable to any appropriate configuration matrix. Specific applications of shape analysis may be found in archaeology, architecture, biology, geography, geology, agriculture, genetics, medical imaging, security applications such as face recognition, entertainment industry (movies, games), computer-aided design and manufacturing, and so on. Inferential methods known in the shape analysis literature make use of configurations of landmarks optimally superimposed using a least-squares procedure, or analyse matrices of interlandmark distances, e.g. by means of Euclidean Distance Matrix Analysis (EDMA). In the two-sample case, a practical method for comparing the mean shapes in the two groups is to use Procrustes tangent space coordinates and if data are concentrated (i.e. close in shape or presenting small variations in shape), calculate the Mahalanobis distance and then Hotelling's T^2 test statistic. Under the assumption of isotropy, another simple approach is to work with statistics based on the squared Procrustes distance and then consider Goodall's F test statistic (Goodall, 1991).

These tests are based on quite stringent assumptions, such as the equality of covariance matrices, the independency of variation within and among landmarks or the multinormality of the model describing landmarks.

As pointed out in Good (2000), the assumption of equal covariance matrices may be unreasonable in certain applications, the multinormal model in the tangent space may be doubted and sometimes there are fewer individuals than landmarks, implying over-dimensioned spaces (course of dimensionality) and loss of power for Hotelling's T^2 test (Blair *et al.*, 1994). Hence an alternative procedure is to consider a permutation approach. Further limitations of traditional inferential procedures have been highlighted in Terriberry *et al.* (2005). Actually, useful shape models contain parameters lying in non-Euclidean spaces. More specifically, in morphometrics, four kinds of parameters should be taken into account: nuisance parameters (e.g.

[1]University Centre of Statistics in the Biomedical Sciences (CUSSB), Vita-Salute San Raffaele University, Via Olgettina 58, 20132 Milan, Italy

[2]Department of Management and Engineering, University of Padova, Stradella San Nicola, 3 36100 Vicenza, Italy

translation and rotation); geometric parameters, such as shape coordinates; statistical parameters, such as mean differences or correlations; and finally another set of geometric parameters, such as partial warp scores or Procrustes residuals (Slice *et al.*, 1996). Some of these parameters may have a large variance, may be highly correlated, or have completely different scales, thus invalidating analyses and final results. Hence traditional statistical tools designed for Euclidean spaces must be used with particular care when they are applicable. On the contrary, permutation tests are appealing because they make no distributional assumptions, requiring only that the data in each group are exchangeable under the null hypothesis. Terriberry *et al.* (2005) presented an application of NPC methodology in shape analysis, but the properties of the method itself are not further investigated. On the strength of these considerations, we suggest an extension of NPC methodology to shape analysis. In fact, under very mild and reasonable conditions, the NPC method is found to be consistent, unbiased and extendable to unconditional inferences. We remark that in the parametric approach, this extension is possible when the data set is randomly selected by well-designed sampling procedures on well-defined population distributions.

We emphasize that permutation tests require homogeneous covariance matrices in order to guarantee exchangeability only in H_0, thus relaxing more stringent assumptions required by parametric tests since they do not require homoscedasticity in the alternative.

Firstly we review inferential methods known in the shape analysis literature, highlighting some drawbacks to the use of Hotelling's T^2 test. Then, focussing on the two-sample case, through an exhaustive comparative simulation study, we evaluate the behaviour of traditional tests along with nonparametric permutation tests using multi-aspect (MA) procedures and domain combinations as well. In this nonparametric framework we also analyze the case of heterogeneous and dependent variation at each landmark. Furthermore we examine the effect of superimposition on the power of NPC tests. Due to their nonparametric nature, we may assert that the suggested tests provide efficient solutions which allow us to deal with data sets including many informative landmarks and few specimens.

2. Testing for shape data

The statistical community has shown an increased interest in shape analysis in the last decade and particular efforts have been addressed towards the development of powerful statistical methods based on models for shape variation of entire configurations of points corresponding to the locations of morphological landmarks. Rohlf (2000) reviews the main tests used in the field of shape analysis and compares the statistical power of the various tests that have been proposed to test for equality of shape in two populations. Although his work is limited to the simplest case of homogeneous, independent, spherical variation at each landmark and the sampling experiments emphasize the case of triangular shapes, it allows practitioners to choose the method with the highest statistical power under a set of assumptions appropriate for the data. By means of a simulation study, he found that Goodall's F test had the highest power followed by T^2 test using Kendall tangent space coordinates. Power for T^2 tests using Bookstein shape coordinates was good if the baseline was not the shortest side of the triangle. The Rao and Suryawanshi shape variables had much lower power when triangles were not close to being equilateral. Power surfaces for the EDMA-I T statistic revealed very low power for many

shape comparisons including those between very different shapes. Power surface for the EDMA-II Z statistic depended strongly on the choice of baseline used for size scaling.

All these tests are based on quite stringent assumptions. In particular, the tests based on the T^2 statistic; e.g. T^2 tests using Bookstein, Kendall tangent space coordinates, Rao and Suryawanshi shape variables, such as Rao-d (1996) and Rao-a (1998) require independent samples and homogeneous covariance matrices in both H_0 and H_1, and shape coordinates distributed according to the multivariate normal distribution. We remark that Hotelling's T^2 is derived under the assumption of population multivariate normality and it is not recommended unless the number of subjects is much larger than that of landmarks (Dryden and Mardia, 1998). It is well known in the literature that Hotelling's T^2 test is formulated to detect any departures from the null hypothesis and therefore often lacks power to detect specific forms of departures that may arise in practice, i.e. the T^2 test fails to provide an easily implemented one-sided (directional) hypothesis test (Blair *et al.*, 1994).

Goodall's F test requires a restrictive isotropic model in which configurations are isotropic normal perturbations from mean configurations, and assumes that the distributions of the squared Procrustes distances are approximately chi-squared distributed.

If we consider the methods based on interlandmark distances, EDMA-I T assumes independent samples and the equality of the covariance matrices in the two populations being compared (Lele and Cole, 1996), while EDMA-II Z assumes only independent samples and normally distributed variation at each landmark.

In order to complete the review on main tests used in shape analysis, we recall the pivotal bootstrap methods for k-sample problems, in which each sample consists of a set of real (the directional case) or complex unit vectors (the two-dimensional shape case), proposed in the paper by Amaral et al. (2007). The basic assumption here is that the distribution of the sample mean shape (or direction or axis) is highly concentrated. This is a substantially weaker assumption than is entailed in tangent space inference (Dryden and Mardia, 1998) where observations are presumed to be highly concentrated. For mathematical, statistical and computational details we refer to Amaral et al. (2007). As pointed out in Good (2000), the assumption of equal covariance matrices may be unreasonable, especially under the alternative, the multinormal model in the tangent space may be doubted and sometimes there are few individuals and many more landmarks, implying over-dimensioned spaces and loss of power for Hotelling's T^2 test. Hence when sample sizes are too small, or the number of landmarks is too large, it is essentially inefficient to assume that observations are normally distributed. An alternative procedure is to consider a permutation version of the test (see Good, 2000; Dryden and Mardia, 1993; Bookstein, 1997; Terriberry *et al.*, 2005). Permutation methods are distribution-free and allow for quite efficient solutions which may be tailored for sensitivity to specific treatment alternatives providing one-sided as well as two-sided tests of hypotheses (Blair *et al.*, 1994).

In the wake of these considerations, we propose an extension of the NPC methodology (Pesarin and Salmaso, 2010) to shape data. Generally, permutation tests require homogeneous covariance matrices under H_0 in order to guarantee exchangeability, thus relaxing the stringent assumptions of parametric tests. This is consistent with the notion that the true H_0 implies the equality in multivariate distribution of observed variables.

3. NonParametric Combination (NPC) and shape analysis

Let $\mathbf{X}_1$ be the $n_1 \times (k \times m)$ matrix of aligned data (e.g. specimens or individuals) in the tangent space in the first group. By the term 'aligned' we mean that we are considering the shape coordinates obtained through GPA. We recall that the direct analysis of databases of landmark locations is not convenient because of the presence of nuisance parameters, such as position, orientation and size. Usually, in order to carry out a valuable statistical shape analysis, a GLS or GPA superimposition is performed to eliminate non-shape variation in configurations of landmarks and to align the specimens to a common coordinate system (Rohlf and Slice, 1990). Hence the GPA procedure is performed to estimate a mean shape and to align the specimens to it.

Similarly $\mathbf{X}_2$ is the $n_2 \times (k \times m)$ matrix of aligned specimens in the tangent space, i.e. the second group of subjects. Let $\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}$ be the $n \times (k \times m)$ matrix of aligned specimens in the tangent space, i.e. our data set (that may be written as $\mathbf{X} = \mathbf{X}_1 \uplus \mathbf{X}_2$), where $n = n_1 + n_2$. Hence $\mathbf{X}$ is a matrix of data with specimens in rows and landmark coordinates in columns. The unit-by-unit representation of $\mathbf{X}$ is then $\{X_{hji}, i = 1, \ldots, n, j = 1, 2, h = 1, \ldots, km; n_1, n_2\}$.

For simplicity, we may assume that the landmark coordinates in the tangent space behave according to the model $X_{hji} = \mu_h + \delta_{hj} + \sigma_h Z_{hji}$, $i = 1, \ldots, n$, $j = 1, 2$, $h = 1, \ldots, km$, where:

- k is the number of landmarks in m dimensions;

- μ_h represents a population constant for the h-th variable;

- δ_{hj} represents treatment effect (i.e. the noncentrality parameter) in the jth group on the hth variable which, without loss of generality, is assumed to be $\delta_{h1} = 0$, $\delta_{h2} \geq (\text{or} \leq) 0$;

- σ_h is the scale coefficient specific to the h-th variable;

- Z_{hji} are random errors assumed to be exchangeable with respect to treatment levels, independent with respect to units, with null mean vector ($\mathbb{E}(\mathbf{Z}) = 0$), and finite second moment.

With reference to the scale coefficients σ_h, we observe that these parameters may be very useful since they reflect the 'intrinsic' biases in the registration of landmarks. There are in fact landmark points readily available, hence easier to capture than others by the operator or machine. As a consequence, they are less variable in their location. Hence landmark coordinates in the first group differ from those in the second group by a 'quantity' δ, where δ is the km-dimensional vector of effects. $\left\{ X_{hji}^*, i = 1, \ldots, n, j = 1, 2, h = 1, \ldots, km; n_1, n_2 \right\}$ indicates a permutation of the original data.

Therefore the specific hypotheses may be expressed as

$$H_0 : \left\{ \bigcap_{h=1}^{km} \left[X_{h1} \stackrel{d}{=} X_{h2} \right] \right\} \quad \text{against } H_1 : \left\{ \bigcup_{h=1}^{km} \left[X_{h1} \stackrel{d}{>} (<) X_{h2} \right] \right\},$$

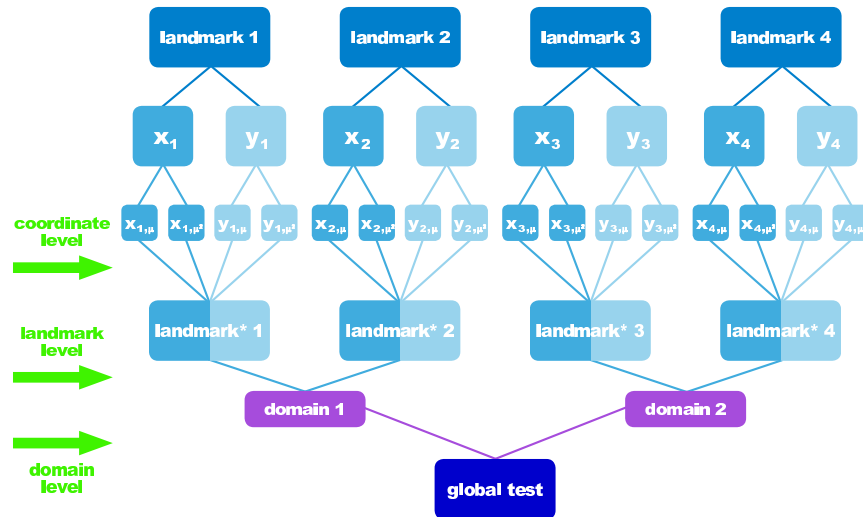where $\stackrel{d}{>}$ stands for distribution (or stochastic) dominance.

Figure 1: Different levels of combination.

Let $T_h^* = \mathrm{Sg}(\delta_h) \sum_{i \leq n_1} X_{h1i}^*$, $h = 1, \ldots, km$, be the km partial tests where $\mathrm{Sg} = +1$ if $\delta_h > 0$ and -1 otherwise. All these tests are exact, marginally unbiased and consistent, and significant for large values, so that the NPC theory properly applies. The hypothesis testing problem is broken down into two stages, considering both the coordinate and the landmark level (and, if present, the domain level). We actually refer to domains as subgroups of landmarks sharing anatomical, biological or locational features. However, we wish to remark that a domain may be seen as a latent variable.

Hence, we formulate partial test statistics for one-sided hypotheses and then we consider the global test T'' obtained after combining at the first stage with respect to m, then with respect to k (of course, this sequence may be reversed). We may also apply the multi-aspect procedure to each coordinate of a single landmark and then consider their combination. For example, if we consider 4 landmarks, first of all we can consider a test for each coordinate (x and y coordinates in the 2D case) of each landmark. Once the aspects of interest have been decided (e.g. first two moments), we can focus on the coordinate level, or on the landmark level after combining coordinates, or on the domain level as well, and finally on the global test (see Figure 1).

One of the main features and advantages of the proposed approach is that by using the multi-aspect procedure and the information about domains, we are able to obtain not only a global p-value, as in traditional tests, but also a p-value for each of the defined aspects or domains. Hence following our procedure it is possible to construct a hierarchical tree, allowing for testing at different levels of the tree (see Figure 1). On the one hand, partial tests may provide marginal information for each specific aspect; on the other, they jointly provide information on the global hypothesis. In this way, if we find a significant departure from H_0, we can investigate the nature of this departure in detail. Furthermore, we can move from the top of the tree to the bottom and, for interpreting results in a hierarchical way, from the bottom to the top. It is worth noting that "intermediate" level p-values need to be adjusted for multiplicity.

4. Finite sample consistency of combination-based tests

In order to introduce Finite Sample Consistency property from a practical point of view, let us consider the two independent sample case and assume that the response variables behave according to the following model:

$$X_{hji} = \mu_h + \delta_{hj} + Z_{hji},$$

$i = 1, \ldots, n_j$, $j = 1, 2$, $h = 1, \ldots, k$, where n_j is the sample size, μ_h represents a population constant for the h-th variable; δ_{hj} represents the fixed treatment effect (i.e. the noncentrality parameter) in the j-th group on the h-th variable and Z_{hji} are k-dimensional random errors exchangeable with respect to treatment levels with null mean vector ($\mathbb{E}(\mathbf{Z}) = 0$) and finite second moment.

Let $\bar{X}_{hj}$, $j = 1, 2$, be the sample mean for the h-th variable, S_j the biased sample covariance matrix (with divisors n_1 and n_2) and S the common covariance matrix, given by $S = (n_1 S_1 + n_2 S_2)/(n_1 + n_2 - 2)$.

We define T''^* the nonparametric permutation counterpart of Hotelling T^2 given by

$$T''^* = \sum_{h=1}^{k} \left(\frac{\bar{X}_{h1}^* - \bar{X}_{h2}^*}{s_h^*} \right)^2$$

where the symbol $*$ indicates a permutation of the original data, $\bar{X}_{hj}^*$, $j = 1, 2$ are multivariate permutation sample means and s_h^* are the diagonal elements of S^*. We remark that the underlying dependence structure is nonparametrically and implicitly 'captured' by the permutation procedure (see e.g. Pesarin and Salmaso, 2010). We also emphasize that in a shape analysis framework X_{hji} will indicate the 2D or 3D landmark coordinates.

When carrying out nonparametric permutation tests we use raw coordinates and not the shape coordinates. Hence we do not use the coordinates obtained after filtering out location, scale and rotational effects from the original data.

We have compared the traditional parametric Hotelling's T^2 test (T^2) with the nonparametric T^2-type counterpart (T''^*) showing that the power for the suggested test increases when increasing the number of the processed variables (see Table 1) with the same noncentrality parameter δ, even when the number of covariates (k) is larger than the permutation sample space (see results in Table 2 and Table 3).

We remark that Hotelling's T^2 test considered in Table 1 is computed using raw coordinates and not shape variables. Moreover, when $n_1 = n_2 = 10$ and $k = 19$, the test statistic is constantly equal to 0. Hotelling's T^2 statistic can be related to the F-distribution by the well-known relation

$$T^2 = \frac{n_1 n_2 (n_1 + n_2 - k - 1)}{(n_1 + n_2)(n_1 + n_2 - 2)k} D^2 \sim F_{k, n_1 + n_2 - 1 - k},$$

where D^2 is the Mahalanobis squared distance.

B is the number of permutations (Monte Carlo sampling) used for estimating the permutation distribution, and CMC is the number of Monte Carlo iterations of the simulation procedure. Note that for $n_1 = n_2 = 3$ we explored the whole permutation sample space.

These interesting findings allow us to assess the usefulness of the nonparametric permutation solution for high-dimensional data in small sample size case.

Table 1: Simulations under H_1 ($n_1 = n_2 = 10$, $\mu = 0$, $\delta = 0.40$, $B = CMC = 1000$)

		α=0.01	α=0.05	α=0.10	α=0.20	α=0.30	α=0.50
$k = 15$	T^2	0.027	0.118	0.233	0.419	0.566	0.789
	$T^{''*}$	0.231	0.484	0.623	0.771	0.856	0.941
$k = 16$	T^2	0.026	0.098	0.192	0.361	0.504	0.741
	$T^{''*}$	0.228	0.496	0.633	0.792	0.866	0.946
$k = 17$	T^2	0.019	0.081	0.158	0.325	0.455	0.703
	$T^{''*}$	0.258	0.534	0.681	0.811	0.875	0.950
$k = 18$	T^2	0.013	0.067	0.132	0.269	0.414	0.642
	$T^{''*}$	0.253	0.543	0.667	0.816	0.874	0.956
$k = 19$	$T^{''*}$	0.244	0.544	0.700	0.837	0.905	0.977
$k = 20$	$T^{''*}$	0.318	0.552	0.683	0.825	0.904	0.965
$k = 21$	$T^{''*}$	0.307	0.570	0.693	0.832	0.901	0.962
$k = 22$	$T^{''*}$	0.340	0.618	0.744	0.845	0.906	0.964
$k = 23$	$T^{''*}$	0.344	0.629	0.750	0.857	0.918	0.974
$k = 24$	$T^{''*}$	0.338	0.622	0.741	0.862	0.919	0.973
$k = 25$	$T^{''*}$	0.365	0.656	0.774	0.880	0.930	0.970

Table 2: Simulations under H_1 ($n_1 = n_2 = 3$, $\mu = 0$, $\delta = 0.40$, $B = MC = 1000$)

		α=0.10	α=0.20	α=0.30	α=0.50
$k = 3$	$T^{''*}$	0.059	0.194	0.337	0.554
$k = 18$	$T^{''*}$	0.097	0.278	0.408	0.618
$k = 20$	$T^{''*}$	0.090	0.264	0.390	0.611
$k = 25$	$T^{''*}$	0.117	0.274	0.422	0.643
$k = 30$	$T^{''*}$	0.100	0.270	0.404	0.647
$k = 35$	$T^{''*}$	0.103	0.280	0.436	0.687
$k = 40$	$T^{''*}$	0.089	0.277	0.442	0.667

Table 3: Simulations under H_1 ($n_1 = n_2 = 3$, $\mu = 0$, $\delta = 1$, $B = MC = 1000$)

		α=0.10	α=0.20	α=0.30	α=0.50
$k = 3$	$T^{''*}$	0.187	0.454	0.629	0.800
$k = 20$	$T^{''*}$	0.324	0.779	0.902	0.977
$k = 50$	$T^{''*}$	0.434	0.907	0.985	0.999

This simple simulation study introduces the Finite Sample Consistency property of permutation tests in an intuitive way, i.e. it can be proved that, for a given and fixed number of subjects, when the number of variables increases and the noncentrality parameter of the underlying population distribution increases with respect to each added variable, then power of multivariate combination-based tests increases (for details we refer to Pesarin and Salmaso, 2010 and Brombin and Salmaso, 2013).

5. Comparative simulation study

Let us assume that our samples are made of configurations of $k = 8$ landmarks in $m = 2$ dimensions characterized by slightly different means. Suppose we are dealing with male and female skull configurations of a particular animal created ad hoc, representing the two independent samples (see Table 4). Since sample means differ slightly from each other in 6 out of 16 coordinates (highlighted in bold in Table 4), we have generated data using such a configuration in order to evaluate the power of the competing tests.

Table 4: Hypothetical mean configurations

Domain	#	Lnd. name	Male		Female	
			x	y	x	y
1	1	nasion	65.00	223.00	65.00	**222.85**
1	2	basion	54.00	-40.00	**53.75**	-40.00
2	3	staphylion	0.00	0.00	0.00	0.00
2	4	prosthion	0.00	35.00	0.00	**34.50**
2	5	nariale	19.00	121.00	**18.90**	121.00
3	6	bregma	70.00	203.00	70.00	203.00
3	7	lambda	110.00	112.00	**109.95**	112.00
3	8	opisthion	104.00	17.00	104.00	**16.88**

In all simulations we have $B = 1000$ CMC iterations and the number of Monte Carlo runs $MC = 1000$. Throughout the simulation study, we have selected both Liptak and Fisher combining functions. Focussing on the two-sample case, we have carried out the simulation study in the same conditions of homogeneous, independent, spherical variation at each landmark, as described in Rohlf (2000). In the complete simulation study (Brombin, 2009) we have compared, in terms of statistical power, traditional approaches for the statistical analysis of shape, such as

- Hotelling's T^2 test using approximate tangent space coordinates and Bookstein shape coordinates,

- Goodall's F test,

- EDMA-I (Lele and Richtsmeier, 1991) and EDMA-II tests (Lele and Cole, 1995; 1996),

- T^2 test using Rao and Suryawanshi shape variables, Rao-d (1996) and Rao-a (1998) only for large sample sizes, e.g. $n_1 = n_2 = 50$.

As regards permutation tests, in particular we have considered (i) permutation Hotelling's T^2, (ii) permutation global tests with and without domains using Liptak

and Fisher combining functions, (iii) permutation multi-aspect (MA) tests with and without domains, considering location and scale aspects and using Liptak and Fisher combining functions.

Let n_j, $j = 1, 2$, be the sample sizes of the two samples and the landmark coordinates be multivariate normally distributed. Three domains have been considered, i.e. baseline (nasion and basion), face (staphylion, prosthion and nariale), and braincase (bregma, lambda and opisthion). We use G to denote the global test obtained after combining all partial tests, and G_d to denote the global test that takes into account the information about domains (hence obtained after combining partial tests on chosen domains). MA, if present, indicates that the a multi-aspect procedure has been applied. T''^* indicates the nonparametric permutation counterpart of Hotelling's T^2.

We wish to point out that in this simulation study we only show global p-values. Hence, in this case, we do not need to cope with the multiplicity problem. For the sake of space, we report only the results of NPC tests with small sample sizes ($n_1 = n_2 = 10$, $\sigma^2 = 0.25$). In Table 5, the better performing tests are highlighted in bold. In particular, we show that the global nonparametric test using Fisher's combining function, in its standard, domain and MA versions, has better power in almost all the situations.

Our simulation study emphasizes the good behaviour in terms of power and the flexibility of the nonparametric permutation solution in shape analysis, since it allows us to carry out a shape analysis even in the presence of small sample sizes and a large number of shape variables. We wish to highlight that, under the null hypothesis, the proportion of rejection of the tests is very close to the given significance level (see Table 5).

Table 5: Attained α-level and power behaviour: $n_1 = n_2 = 10$, B=CMC=1000, $\sigma^2 = 0.25$

	Attained α-level					
	α=0.01	α=0.05	α=0.10	α=0.20	α=0.30	α=0.50
T''^*	0.011	0.056	0.112	0.211	0.300	0.506
G (Liptak)	0.017	0.054	0.107	0.215	0.307	0.512
G_d (Liptak)	0.014	0.055	0.115	0.213	0.304	0.512
G (Fisher)	0.014	0.058	0.120	0.208	0.307	0.506
G_d (Fisher)	0.013	0.057	0.121	0.205	0.313	0.490
G (Liptak, MA)	0.016	0.051	0.109	0.217	0.311	0.516
G_d (Liptak, MA)	0.013	0.046	0.116	0.218	0.310	0.514
G (Fisher, MA)	0.010	0.058	0.118	0.214	0.311	0.504
G_d (Fisher, MA)	0.012	0.061	0.115	0.210	0.317	0.484
	Power					
	α=0.01	α=0.05	α=0.10	α=0.20	α=0.30	α=0.50
T''^*	**0.087**	**0.229**	**0.320**	**0.496**	**0.611**	**0.775**
G (Liptak)	0.054	0.172	0.279	0.434	0.549	0.741
G_d (Liptak)	0.048	0.165	0.265	0.419	0.526	0.724
G (Fisher)	**0.079**	**0.214**	**0.326**	**0.480**	**0.615**	**0.769**
G_d (Fisher)	0.067	0.203	0.302	0.452	0.581	0.753
G (Liptak, MA)	0.056	0.175	0.275	0.440	0.549	0.748
G_d (Liptak, MA)	0.052	0.170	0.265	0.419	0.527	0.723
G (Fisher, MA)	**0.075**	**0.218**	**0.312**	**0.477**	**0.604**	**0.767**
G_d (Fisher, MA)	0.071	0.203	0.290	0.447	0.568	0.755

6. Combination-based tests with correlated landmarks

In this framework we analyse the case of heterogeneous and dependent variation at each landmark, and evaluate power and the attained α-level. The superimposition step has been included in the routine since GPA superimposition may modify dependency structures among landmarks. In order to obtain a non singular covariance matrix, we have performed an eigenvalue decomposition (ED) of the original variance covariance matrix and transformed the original eigenvalues λ. We have considered transformations such as $\lambda^{1/3}$ and $\lambda^{1/10}$, rescaled by their trace (see the effect of transforming eigenvalues on the scatterplot in Figure 6).

Then we have recalculated the covariance matrix Σ^*, using the relation $\Sigma^* = V\Lambda^* V'$, where Λ^* is a diagonal matrix with the transformed eigenvalues, V is an orthogonal matrix, containing the corresponding eigenvectors and V' means V transposed.

Under the alternative, data have been generated using different means and the same covariance matrix Σ^*. In Table 6 we display hypothetical mean configurations, representing 3D male and female *Macaca fascicularis* monkey skulls (for details, see Frost *et al.*, 2003).
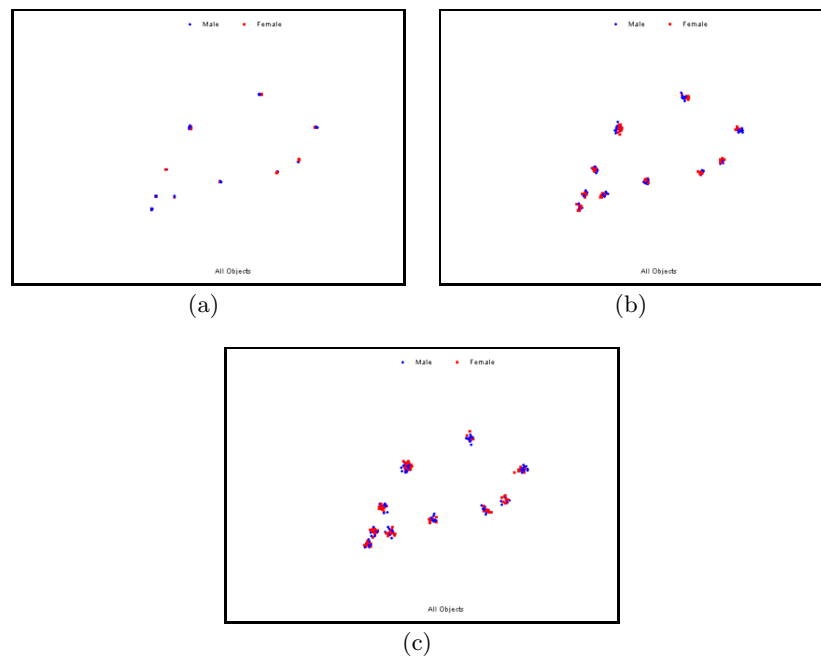

(a) (b)

(c)

Figure 2: Original eigenvalues λ (a), $\lambda^{1/3}$ (b) and $\lambda^{1/10}$ (c)

Let n_j, $j = 1, 2$, denote the sample size in the two groups. In particular we considered the settings: $n_1 = n_2 = 5$, $n_1 = n_2 = 10$, $n_1 = 5$, $n_2 = 10$. In the simulation study we have evaluated power and attained α-level when the number of 3D landmarks k was, in turn, equal to 3,6,9,11. Three domains have been considered: the first includes landmarks 1, 2 and 11; the second includes landmarks from 3 to 7; the third includes landmarks from 8 to 10.

We use T''^* to denote Hotelling's T^2 permutation counterpart, with G the combination of all partial tests, and G_d the combination using domains. For the sake of space, we present simulation results only for the case in which the number of 3D landmarks k is equal to 6, $n_1 = n_2 = 10$ and transformed λ is $\lambda^{1/3}$(see Table 7).

Table 6: Configurations

Landmark		Male			Female		
#	Lnd. name	x_M	y_M	z_M	x_F	y_F	z_F
1	Inion	17.7752	18.9981	6.9585	17.5252	18.9981	6.9585
2	Bregma	15.9101	16.3499	9.2159	15.9101	16.4499	9.2159
3	Glabella	13.6833	12.7086	7.6433	13.6833	12.7086	7.6433
4	Nasion	13.6799	12.6892	7.5628	13.8299	12.6892	7.5628
5	Rhinion	12.9273	11.2649	5.1792	12.9273	11.2149	5.1792
6	Nasospinale	12.6114	10.5523	3.6257	12.6114	10.5523	3.6257
7	Prosthion	12.4725	10.233	2.8531	12.4725	10.2330	2.8531
8	Opisthion	17.1882	17.8852	5.0014	17.1882	17.8852	5.1514
9	Basion	16.5070	16.7665	4.4799	16.5070	16.7165	4.4799
10	Staphylion	14.6975	13.8755	4.1783	14.6075	13.8755	4.1783
11	Incisivion	13.2442	11.4665	3.5466	13.2442	11.4665	3.5166

In all the simulations under H_0, when using a global test with Fisher's combining function, MA procedure and domain information, the Type I error rate was too large, thus invalidating inferential conclusions. For example, in Table 7, focussing on $\alpha = 0.05$, $G_{MA,F}$ has a corresponding attained $\alpha = 0.112$ and $G_{d,MA,F}$ has a corresponding attained $\alpha = 0.122$. Readers are reminded that at the beginning we decided to evaluate location (mean) and distributional (Anderson–Darling statistic) aspects. Hence GPA superimposition may modify dependency structures, thus altering the final distribution. Fisher's combining function is more sensitive with MA procedures than Tippett's combining function. However, if we change the aspects, e.g. if we consider mean μ and median Me or mean and second moments μ^2, the proportion of rejection of the tests $G_{MA,F}$ and $G_{d,MA,F}$ now achieves the given significance α-level. These results highlight that GPA affects the initial distribution of the data, hence particular care is needed when a MA procedure is performed.

7.

Table 7: Attained α-level and power behaviour: $n_1 = n_2 = 10$, $\lambda^{1/3}$, $k = 6$, $m = 3$, $B = CMC = 1000$

	Attained α-level					
Test	0.01	0.05	0.10	0.20	0.30	0.50
T''^*	0.012	0.056	0.102	0.204	0.295	0.493
G_T	0.000	0.052	0.088	0.195	0.290	0.489
$G_{d,T}$	0.004	0.038	0.103	0.195	0.288	0.475
G_F	0.011	0.055	0.099	0.197	0.292	0.504
$G_{d,F}$	0.011	0.050	0.098	0.196	0.293	0.496
$G_{MA,T}$	0.008	0.031	0.065	0.136	0.206	0.350
$G_{d,MA,T}$	0.008	0.027	0.073	0.143	0.216	0.345
$G_{MA,F}$	0.041	0.112	0.164	0.266	0.340	0.508
$G_{d,MA,F}$	0.043	0.122	0.185	0.305	0.402	0.550
$G_{(\mu,\mathbb{M}e),F}$	0.012	0.055	0.099	0.190	0.288	0.466
$G_{d,(\mu,\mathbb{M}e),F}$	0.011	0.054	0.098	0.191	0.300	0.479
$G_{(\mu,\mu^2),F}$	0.010	0.048	0.109	0.195	0.286	0.509
$G_{d,(\mu,\mu^2),F}$	0.009	0.043	0.096	0.192	0.287	0.485
	Power					
Test	0.01	0.05	0.10	0.20	0.30	0.50
T''^*	0.337	0.621	0.757	0.888	0.930	0.981
G_T	0.000	0.538	0.709	0.855	0.923	0.978
$G_{d,T}$	0.123	0.478	0.683	0.828	0.902	0.971
G_F	0.299	0.564	0.709	0.858	0.916	0.973
$G_{d,F}$	0.286	0.547	0.679	0.819	0.898	0.962
$G_{MA,T}$	0.234	0.456	0.636	0.754	0.851	0.945
$G_{d,MA,T}$	0.255	0.401	0.600	0.755	0.824	0.923
$G_{(\mu,\mathbb{M}e),F}$	0.284	0.571	0.711	0.853	0.918	0.969
$G_{d,(\mu,\mathbb{M}e),F}$	0.164	0.431	0.596	0.797	0.880	0.952
$G_{(\mu,\mu^2),F}$	0.224	0.467	0.612	0.740	0.833	0.922
$G_{d,(\mu,\mu^2),F}$	0.137	0.369	0.495	0.663	0.782	0.903

Conclusions

Due to their nonparametric nature, the suggested tests provide efficient solutions in multivariate small-sample-size problems. We have seen how the power of permutation tests based on combination-based tests (Pesarin and Salmaso, 2010) increases when the number of processed variables increases, provided that the noncentrality parameter increases, even when the number of variables is larger than the permutation sample space. These findings are very important, since they show that it is possible to obtain powerful tests in a nonparametric framework by increasing the number of informative variables while leaving the number of cases fixed. Through the application to shape analysis, we have highlighted the power and the flexibility of the nonparametric permutation solution. For a comprehensive overview on permutation tests in shape analysis we refer the reader to our forthcoming book Brombin and Salmaso (2013).

REFERENCES

Amaral, G., Dryden, I. and Wood, A. (2007), "Pivotal bootstrap methods for k-sample problems in directional statistics and shape analysis," *Journal of the American Statistical Association*, 102, 695–707.

Blair, R. C., Higgins, J. J., Karniski, W. and Kromrey, J. D. (1994), "A study of multivariate permutation tests which may replace Hotelling's t^2 test in prescribed circumstances," *Multivariate Behavioral Research*, 29, 141–163.

Bookstein, F. L. (1997), "Shape and the information in medical images: A decade of the morphometric synthesis," *Computer Vision and Image Understanding*, 66, 97–118.

Brombin, C. (2009), *A Nonparametric Permutation Approach to Statistical Shape Analysis*, Ph.D. thesis, University of Padova, Padova, Italy.

Brombin, C. and Salmaso, L. (2013), *Permutation Tests in Shape Analysis*, SpringerBriefs series, Springer, NY.

Dryden, I. L. and Mardia, K. V. (1993), "Multivariate Shape Analysis," *Sankhyā Series A*, 55, 460–480.

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

Frost, S. R., Marcus, L. F., Bookstein, F. L., Reddy, D. P. and Delson, E. (2003), "Cranial allometry, phylogeography, and systematics of large-bodied papionins (primates: Cercopithecinae) inferred from geometric morphometric analysis of landmark data," *The Anatomical Record (Part A)*, 275A, 1048-1072.

Good, P. (2000), *Permutation tests: a practical guide to resampling methods for testing hypotheses*, Springer, New York.

Goodall, C. R. (1991), "Procrustes methods in the statistical analysis of shape," *Journal of the Royal Statistical Society, Series B*, 53, 285–339.

Lele, S. and Cole, T. M. (1995), "Euclidean distance matrix analysis: a statistical review," *Current issues in statistical shape analysis*, 3, 49–53.

Lele, S. and Cole, T. M. (1996), "A new test for shape differences when variance-covariance matrices are unequal," *The Journal of Human Evolution*, 31, 193–212.

Lele, S. and Richtsmeier, J. T. (1991), "Euclidean distance matrix analysis: a coordinate free approach for comparing biological shapes using landmark data," *American Journal of Physical Anthropology*, 86, 415-427.

Pesarin F. and Salmaso L. (2010), *Permutation Tests for Complex Problems: Theory, Applications and Software*, Wiley, Chichester.

Rao, C. and Suryawanshi, S. (1996), "Statistical analysis of shape of objects based on landmark data," *Proceedings of the National Academy of Sciences of the United States of America*, 93, 12132–12136.

Rao, C. and Suryawanshi, S. (1998), "Statistical analysis of shape through triangulation of landmarks: a study of sexual dimorphism in hominids," *Proceedings of the National Academy of Sciences of the United States of America*, 95, 4121–4125.

Rohlf, F. J. (1999), "Shape statistics: Procrustes superimpositions and tangent spaces," *Journal of Classification*, 16, 197–225.

Rohlf, F. J. (2000), "Statistical power comparisons among alternative morphometric methods,"

American Journal of Physical Anthropology, 111, 463–478.

Rohlf, F. J. and Slice, D. E. (1990), "Extensions of the procrustes method for the optimal super-imposition of landmarks," *Systematic Zoology*, 39, 40–59.

Slice, D. E., Bookstein, F. L., Marcus, L. F. and Rohlf, F. J. (1996), "A glossary for geometric morphometrics," *Advances in Morphometrics*, 284, 531–551.

Terriberry, T. B., Joshi, S. and Garig, G. (2005), "Hypothesis testing with nonlinear shape models", in *Information processing in medical imaging*, vol. 3565, Springer Berlin-Heidelberg, pp. 15-26.