



**HAL**  
open science

## Overcoming the Spurious Groups Problem in Between-Group PCA

Jean Thioulouse, Sabrina Renaud, Anne-Béatrice Dufour, Stéphane Dray

► **To cite this version:**

Jean Thioulouse, Sabrina Renaud, Anne-Béatrice Dufour, Stéphane Dray. Overcoming the Spurious Groups Problem in Between-Group PCA. *Evolutionary Biology*, 2021, 10.1007/s11692-021-09550-0 . hal-03348141

**HAL Id: hal-03348141**

**<https://hal.science/hal-03348141v1>**

Submitted on 17 Sep 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

[Click here to view linked References](#)

<b>Evolutionary Biology manuscript No.</b> (will be inserted by the editor)
--

---

## Overcoming the Spurious Groups Problem in Between-Group PCA

Jean Thioulouse · Sabrina Renaud ·  
Anne-Béatrice Dufour · Stéphane Dray

Received: date / Accepted: date

**Abstract** Several papers have recently raised the occurrence of some problems with between-group Principal Component Analysis (bgPCA). This method inflates the differences between the groups, and can even display completely artificial differences when none exist, for example when applied to random numbers tables with many variables (columns) and few individuals (rows). Lately, cross-validation has been proposed as a way to circumvent this problem. Here we present some tools and several functions of the **ade4** package for the **R** statistical software to compute a bgPCA, test the presence of statistically significant groups, perform a cross-validation of this analysis and compute associated statistics. We also describe how to use these functions to avoid running into the spurious groups problem. Several examples, including a real data set and random numbers tables, are used to validate this approach in var-

---

✉ J. Thioulouse

Université de Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive  
UMR 5558, F-69622 Villeurbanne, France.  
E-mail: jean.thioulouse@univ-lyon1.fr  
ORCID: 0000-0001-7664-0598

S. Renaud

Université de Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive  
UMR 5558, F-69622 Villeurbanne, France.  
E-mail: sabrina.renaud@univ-lyon1.fr  
ORCID: 0000-0002-8730-3113

A.B. Dufour

Université de Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive  
UMR 5558, F-69622 Villeurbanne, France.  
E-mail: anne-beatrice.dufour@univ-lyon1.fr  
ORCID: 0000-0002-9339-4293

S. Dray

Université de Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive  
UMR 5558, F-69622 Villeurbanne, France.  
E-mail: stephane.drays@univ-lyon1.fr  
ORCID: 0000-0003-0153-1105

ious experimental and numerical conditions. The integrated framework of the duality diagram, as implemented in **ade4**, allows to extend this approach to other multivariate analysis methods beyond principal component analysis, which could prove useful in the case of other types of variables. The **R** code and the real data table used to make the computations and graphs of this paper are available as supplementary material.

**Keywords** Geometric morphometrics · Multivariate analysis · ade4 · Between-group analysis · Spurious groups · Random permutation test · Leave-one-out cross-validation

## 1 Introduction

In many fields, multivariate data are measured on individuals (samples, specimens, populations, etc.) belonging to different groups. Researchers aim to check whether differences exist between the groups and which variables mainly drive them. Canonical Variate Analysis (CVA or DA, Discriminant Analysis) was extensively used, in the context of geometric morphometrics (*e.g.*, Debat et al., 2003; Leinonen et al., 2006; Valenzuela-Lamas et al., 2011), to evidence differentiation among groups of individuals using multivariate data. However, a paper by Mitteroecker and Bookstein (2011) alerted the community of geometric morphometrics on the fact that CVA may lead to spurious results when the number of variables exceeded the number of individuals. The same year, Kovarovic et al. (2011) and Viscosi and Cardini (2011) also noted the paramount importance of carefully choosing the predictor variables and the need to address the overfitting problem by cross-validation when using CVA. In fact, some statisticians even recommended (without demonstration) that the number of individuals should be at least equal to ten times the number of variables (Weinberg and Darlington, 1976). This is probably excessive, but the issue is becoming more and more problematic with the raise of 3D data in morphometrics, that could deliver hundreds and even thousands of shape variables (*e.g.*, Harbers et al., 2020). Evin et al. (2013) noted that performing CVA on principal components “may help not only to maximize the cross-validated accuracy but also to reduce the noise in the data, as well as the positive bias in classification accuracy when groups have different sample sizes”. Hence, CVA remains quite popular, but it is now regularly associated with a procedure of dimensionality reduction (Chiari and Claude, 2012; Cucchi et al., 2017; Dianat et al., 2017; Harbers et al., 2020).

Between-Group Analysis (BGA) was first proposed, in two papers (in French) by Dolédec and Chessel (1987; 1989), as an alternative to CVA when the number of variables exceeds the number of individuals. These papers explained how to disentangle the variations in ecological data tables according to the categories of a qualitative variable. In 2002, Culhane et al. (2002) used the **ADE-4** software (Thioulouse et al., 1997), a former standalone version of the **ade4** package for **R**, to apply BGA to microarray data sets. In the field of geometric morphometrics, the method soon appeared, under the denomination of bgPCA (between-group Principal Component Analysis), as an alternative to circumvent the problems of CVA when the number of variables exceeded the number of individuals and the method is now increasingly used (*e.g.*, Alméjija et al., 2013; Gunz et al., 2012; Ledevin and Koyabu, 2019; Souquet et al., 2019). However, a series of papers by Cardini et al. (2019), Bookstein

(2019) and Cardini and Polly (2020) have recently alerted morphometricians, and biologists at large, about the fact that bgPCA, as CVA, can also exhibit spurious groups, even when applied to randomly generated data without any group structure.

As noted by Cardini et al. (2019), “These are important problems for a method mainly designed for the analysis of variation among groups when there are very large numbers of variables and relatively small samples. In such cases, users are likely to conclude that the groups they are comparing are much more distinct than they really are.” These authors use the mean overlap index,  $O_{ij}$  to measure the dispersion of bgPCA groups, and Cardini and Polly (2020) showed that a permutation test of the global  $R^2$  (i.e., the part of total variance that is due to differences among groups) associated to a cross-validation of bgPCA coordinates can help discard spurious cases. Several packages, particularly **Morpho** for the **R** statistical software (Schlager, 2017) and the **MorphoJ** software (Klingenberg, 2011), provide many functions to apply (cross-validated) CVA and bgPCA to geometric morphometric data sets and to perform associated permutation tests, particularly PERMANOVA, with the *adonis* function of the **vegan** package for **R** (Oksanen et al., 2019).

In this paper, we propose a procedure for identifying, quantifying and overcoming the spurious groups problem in BGA. We prefer to use the abbreviation BGA (for Between-Group Analysis) instead of bgPCA (for between-group Principal Component Analysis) because this procedure is not limited to Principal Component Analysis. It can be applied to any type of analysis available in the **ade4** package and thus deals with different types of data, not only quantitative variables. This procedure involves a permutation test of the global  $R^2$  (*randtest* function), the use of cross-validated coordinates (new *loocv* function) and we propose a new measure of the discrepancy between the original BGA and the cross-validated coordinates by comparing their respective mean overlap indexes. We show that this discrepancy can be seen as a “spuriousness index”. We use real and simulated data to show how these different tools can be used to avoid misinterpretations of BGA outputs. The procedure can be easily implemented using the functions *randtest* and *loocv*, available in the **ade4** package for **R**. The whole package is presented by Thioulouse et al. (2018), with one chapter dedicated to BGA and DA. All the **R** scripts used in this paper to draw figures and compute analyses and simulations are available as supplementary material. They can be used to redo and check all the graphs and computations.

## 2 Material and Methods

### 2.1 Statistical tools

We propose the combination of different tools to avoid misinterpretations of BGA outputs due to the presence of spurious groups. These tools are provided by the **ade4** package for **R** (Thioulouse et al., 2018) that offers the same between- and within- group analyses as other geometric morphometric software (e.g., the **Morpho** package). One advantage of **ade4** implementation is its flexibility so that BGA framework can be applied to the analysis of a table of quantitative variables (bgPCA), qualitative variables (by multiple correspondence analysis) or contingency tables (by correspondence analysis) and has been extended to many other types of more complex methods devoted to the analysis of two or three tables (Franquet et al.,

1995; Dray et al., 2015). See Thioulouse et al. (2018) for more details about these analyses.

We used several functions of the **ade4** (Thioulouse et al., 2018) and **adegraphics** (Siberchicot et al., 2017) packages for **R**. The main functions used in the present context are *bca*, for *between class analysis* (“class” and “group” have the same meaning here) to compute between-group analyses, the *randtest* function, to perform randomisation tests on the percentage of between-group variance, and the *loocv* function to compute *leave-one-out cross-validation*. The *s.class* function allowed to display groups on the BGA factor maps and on the cross-validated maps. All these functions are generic (S3 class), which means that they can be applied to several types of analyses, and that the suitable computation method will be automatically chosen.

### 2.1.1 Percentage of explained variation and permutation procedure

The *randtest* function performs a permutation test based on a criterion equal to the percentage of total variation explained by between-group differences ( $R^2$ ). This percentage is useful to assess the validity of BGA results: a low value means that group separation should be considered with caution. A permutation procedure is associated where the rows of the data table (individuals) are permuted and thus randomly assigned to the groups and the  $R^2$  is recomputed for each permutation. The simulated p-value of the test is equal to the proportion of permutations that produces an  $R^2$  value higher (or equal) than the value computed for observed data. This test is equivalent to the PERMANOVA test of the **vegan** package (*adonis* and *adonis2* functions) used for example by Cardini and Polly (2020) and it gives the same results. But the *randtest* function is much faster (10 to 200 times faster) for large data tables, because it is not distance-based and does not need to compute distance matrices. Note that this procedure is based on a permutation of original raw data and assumes that sampling units (individuals) are exchangeable and thus independent. When there is a risk that spatial or phylogenetic autocorrelation occurs between individuals, adapted procedures should be preferred such as randomization of residuals as implemented in the **RRPP** package (Collyer and Adams, 2018) or Moran’s Spectral Randomization (MSR, Wagner and Dray, 2015) available in the **adespatial** package. In this paper, we consider only the case of independence between individuals but we provide R code in the supplementary material showing how to run the testing procedure with both **RRPP** and **adespatial** when phylogenetic dependence occurs.

### 2.1.2 Cross-validation

The *loocv* function implements a leave-one-out procedure: each row of the data table (individual) is removed, one at a time. A BGA is computed on each of these new tables, and the missing row is projected as additional element on this BGA outputs. This allows to compute new coordinates for this individual, based on the group means computed using all other individuals. If group means are really different (due to real differences between groups), then these new cross-validated coordinates is very close to the coordinates obtained in the BGA of the complete table. In this case, the cross-validated factor map will look similar to the factor map of the original analysis. Conversely, if group means are not really different, with just spurious group on the BGA factor map coming from geometrical constraints linked to the size of

vector subspaces (Rohlf, 2021), then the new coordinates will be different from the coordinates obtained in the BGA of the complete table, and the cross-validated factor map will not show spurious groups.

The *loocv* function is useful to get a visual confirmation of the result of the permutation test. If the *randtest* permutation test is statistically significant, then the cross-validated factor map should display the same groups as the original BGA factor map. In the case where there is no real difference between groups (BGA spurious groups), but the p-value of the test is anyway (just by chance) lower than the chosen significance threshold, then the cross-validated factor map will not show the spurious groups. This should prevent users from drawing false conclusions based only on the p-value of the test. The *loocv* outputs include several statistics that can be used to measure the discrepancy between BGA and the cross-validation.

### 2.1.3 The mean overlap index

The mean overlap index  $\bar{O}_{ij}$  (Cardini et al., 2019) is equal to the average, for all the pairs of groups, of the proportion of individuals in a group that are closer to the mean of the other group. This index can be computed for the observed and also for the cross-validated data sets. It can be computed for the full data space, for the full BGA space (i.e., all axes) or for only some axes (for example the two axes used to draw a factor map). In this paper, it is always computed in the full BGA space.

When the groups are very different, their means are very far apart and the groups do not overlap, so  $\bar{O}_{ij}$  will be equal to 0. Conversely, when there is no group structure, they will completely overlap, and  $\bar{O}_{ij}$  will be near to 0.5 because in this case any individual is equally likely to be closest to any other group mean. This overlap index can be computed for both the BGA ( $bga\bar{O}_{ij}$ ) and for the cross-validation ( $xval\bar{O}_{ij}$ ).

In the presence of spurious groups, the BGA mean overlap decreases (compared to a BGA with no spurious group effect) and  $bga\bar{O}_{ij}$  will be lower than  $xval\bar{O}_{ij}$ . So the difference ( $xval\bar{O}_{ij} - bga\bar{O}_{ij}$ ) can be used as an index of spuriousness. This difference varies between 0 (for  $xval\bar{O}_{ij} = 0.5$  and  $bga\bar{O}_{ij} = 0.5$ , i.e. no spurious group effect) and 0.5 (for  $xval\bar{O}_{ij} = 0.5$  and  $bga\bar{O}_{ij} = 0$ , i.e. maximum possible effect). It can therefore be expressed as a percentage of the maximum possible spurious group effect:

$$\Delta\bar{O}_{ij} = (xval\bar{O}_{ij} - bga\bar{O}_{ij})/0.5 * 100$$

These statistics are computed by the *loocv* function.

## 2.2 Simulation study

We design a simulation study to evaluate how the combined use of the procedures described above can help to avoid misinterpretations of BGA outputs. We firstly generate tables of random numbers using the *rnorm* function of **R** (R Core Team, 2020) to simulate data where no real groups exist but spurious groups can appear when increasing the number of variables relative to the number of individuals. A real data set on the house mouse was also analyzed and we added random variables to evaluate how the different statistical tools perform when real groups exist but the structure is degraded by random noise. In this second simulation, we considered random variables without correlation structure (*rnorm*) or with a correlation structure

fixed to the one computed on real house mouse data (using the *mvnorm* function of the **MASS** package).

### 2.2.1 Simulation with no group differences

We generated random numbers drawn from a normal distribution with mean equal to 0 and standard deviation equal to 1 using *rmorm* function. These numbers were then arranged in tables with varying numbers of individuals ( $N_i$ ) and variables ( $N_v$ ). Arbitrary (random) groups of rows were defined in these tables, so these groups corresponded to no real structure. The number of groups ( $N_g$ ) varies in the different simulations.

In this simulation, we evaluated the performance of the  $R^2$  permutation test when no differences exist between groups. We also checked how the statistics computed by the *loocv* function ( $\hat{O}_{ij}$  and  $\Delta\hat{O}_{ij}$ ) changes when varying the total number of individuals and variables.

### 2.2.2 Simulation with group differences

We used a real morphometric data set describing populations of the Western European house mouse (*Mus musculus domesticus*). This subspecies has been shown to display geographic differentiation in its dental morphology, between continental populations (Renaud et al., 2017b) but especially regarding insular populations (Renaud et al., 2011, 2015, 2018). The sampling therefore included mice from the following locations: (1) the neighboring localities of Montpellier (N=13) and Frontignan (N=30), in South-Eastern France along the Mediterranean coast; (2) the locality of Gardouch (N=68) near Toulouse, in South Western France; (3) mice from Lombardy, Northern Italy (N=40); (5) populations from various areas in Corsica (Fango, Bonifacio and Bavella, for a total of N=63) in order to document insular differentiation. All mice were trapped in a commensal context, except those of Frontignan that are characterized by their feral way of life (Renaud et al., 2017b). Details regarding the sampling can be found in previous publications (Renaud et al., 2011, 2015, 2018).

The character used here to quantify this geographic differentiation is the shape of the occlusal surface of the first upper molar (UM1). It was described using 64 points sampled at equal curvilinear distance along the outline, the first point being located at the anterior-most part of the tooth. The points along the outline were analyzed as sliding semi-landmarks (Bookstein, 1991, 1997; Cucchi et al., 2013). Using this approach, the outline points are adjusted using a generalized Procrustes superimposition (GPA) standardizing size, position and orientation, while retaining the geometric relationships between specimens (Rohlf and Slice, 1990). During the superimposition, semi-landmarks were allowed sliding along their tangent vectors until their positions minimize the shape difference between specimens, the criterion being bending energy (Bookstein, 1997). Because the first point was only defined on the basis of a maximum of curvature at the anterior-most part of the UM1, some slight offset might occur between specimens, introducing noise into the data set. The first point was therefore considered as a semi-landmark allowed to slide between the last and second points. Each molar tooth was therefore described by a set of 128 aligned coordinates. The GPA was performed using the **R** package **geomorph** (Adams and Otárola-Castillo, 2013).

The data table is available as supplementary material. As it contains the measurements of 128 variables for 214 individuals, we assume that the numerical conditions ensure that there is no risk of spurious groups in the analysis of this data table. We expect that BGA applied on this data set allows to identify true differences between the groups. We mimic the effect of spurious groups by adding to the house mouse data several sets of 128 columns of random numbers. These sets of 128 columns were generated in two ways:

- Numbers drawn from univariate normal distributions with means and standard deviations equal to the means and standard deviations of the 128 variables of the house mouse data set, with no covariance structure.
- Numbers drawn from a multivariate normal distribution with means and standard deviations equal to the means and standard deviations of the variables of the house mouse data set and with the same covariance matrix.

These simulations allowed us to check the influence of the number of variables and of the spurious group effect when slight real biological differences exist between groups in the data set. We evaluated the performance of the  $R^2$  permutation test to identify these differences, showed how factorial maps are affected by introducing random variables and evaluated if the cross-validation provide better representation of individuals in factorial maps.

### 3 Results

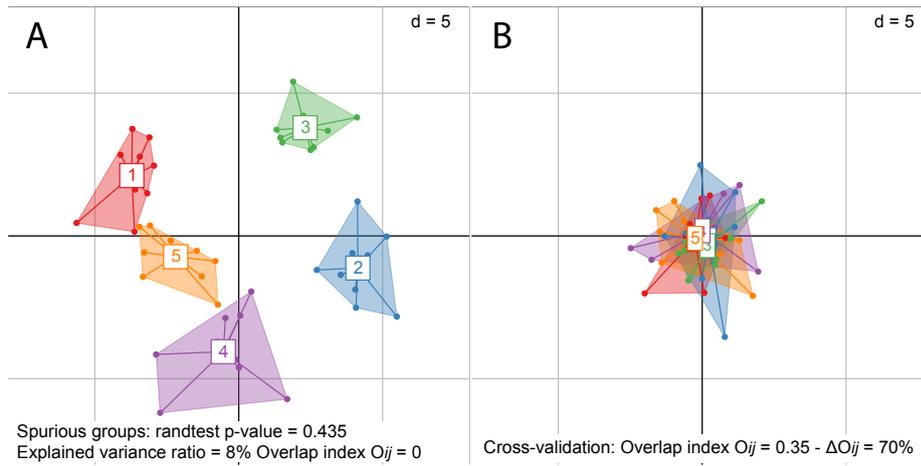
#### 3.1 Simulation with no group differences

Figure 1 shows the factor map of the BGA (Figure 1A) and of the cross-validation (Figure 1B) of a table of 360 random normal variables simulated for 50 individuals belonging to five groups (10 individuals per group). On the left (Figure 1A), the BGA factor map shows spurious groups, and on the right (Figure 1B), the cross-validated factor map computed with the *loocv* function shows that the spurious groups effect has been removed.

Note that the p-value of the *randtest* permutation test of this BGA was equal to 0.435. This means that the test was not tricked by the spurious groups appearing on the BGA factor map. This permutation test is performed in the full data space and it is not influenced by the group separation observed on Figure 1A. This separation of the five groups occurs only in the BGA vector space. So when a BGA leads to a non significant permutation test, all the other results should be discarded and the factor maps should not even be drawn.

However, as for any statistical test, there is always a small probability to get low p-values when doing many tests, even on a table of random numbers. In any case, on 100 tests, it is expected that five will have a p-value less than 0.05, even if the null hypothesis  $H_0$  (no difference between groups) is true. This corresponds to Type I error and is not related to the spurious groups problem, but both can happen simultaneously (see Supplementary Figure 1). In such situations, incautious users could conclude that the five spurious groups visible on Figure 1A correspond to a real structure of the data set. Several ways exist to avoid this problem

The first and most usual way is to use a lower significance threshold for the p-value of the test. This is particularly appropriate here because the spurious groups



**Fig. 1** Factor map of the BGA of a random numbers table (five groups, 10 individuals by group, 360 variables). The five groups are represented by convex hulls surrounding the 10 individuals of each group. A: Factor map showing the spurious groups. The p-value of the *randtest* permutation test is equal to 0.435 and the ratio of between-groups to total variance is equal to 8%. The mean overlap index  $bgaO_{ij}$  is equal to 0 (the five groups are completely apart). B: Cross-validated factor map computed with the *loocv* function, showing that the spurious group effect have been removed. The mean overlap index  $xvalO_{ij}$  is equal to 0.35, reflecting the large overlap of the five groups. The index of spuriousness  $\Delta O_{ij}$  is equal to 70% and the comparison of Figure 1A and 1B confirms the fact that BGA groups on Figure 1A are not supported by cross-validation.

problem appears mostly for large data sets with hundreds or thousands of variables. In this case, even tiny effects can be evidenced and may provide low p-values. Type II error rate will be inflated, but the large size of these data sets should allow to compensate for this problem. Choosing a significance threshold equal to 0.01, 0.005 or even 0.001 instead of 0.05 is therefore a good way to prevent misinterpretations.

Another way to prevent over-optimistic conclusions is to look at the percentage of between-group variance. Here, it is equal to only 8%, and interpretations based on such a low percentage of explained variance should be considered with caution, particularly if the test is not statistically highly significant.

A third way is to use the cross-validation procedure and compare the BGA factor map and the cross-validated factor map. The large difference between Figure 1A and Figure 1B would allow to exclude any erroneous conclusion. The index of spuriousness  $\Delta O_{ij}$  measures the discrepancy between the BGA and the cross-validation. The high value observed here (70%) means that the separation of groups on Figure 1A is probably spurious. The theoretical maximum of 100% would be obtained only for groups completely separated in the BGA space and completely superimposed in the cross-validation space.

### 3.2 Evaluation of the $R^2$ permutation test

In order to test the validity of the *randtest* permutation test in different situations, we varied the number of groups in the simulated data tables ( $N_g = 2, 3, 5, 10$  and

30 groups), while keeping the total number of individuals ( $N_i$ ) constant and equal to 300 (*i.e.*, respectively 150, 100, 60, 20 and 10 individuals per group). The number of variables ( $N_v$ ) was equal to 100, 300, 600, 1200, 1800, 2400 and 4800. Results are summed up in Tables 1, 2 and 3.

**Table 1** Number of *randtest* tests returning a p-value lower than 0.05 on 1000 tests, with 1000 permutations for each test. Columns: number of groups, rows: number of variables. The mean number of significant tests under  $H_0$  is equal to 50 and the 95% interval of its theoretical distribution is [37, 64].

$N_v \setminus N_g$	2	3	5	10	30	Tot
100	41	61	48	43	50	243
300	53	44	62	48	59	266
600	55	54	53	41	50	253
1200	39	59	52	41	52	243
2400	49	59	49	44	41	242
4800	49	57	56	45	52	259
Total	286	334	320	262	304	1506

**Table 2** Number of *randtest* tests returning a p-value lower than 0.01 on 5000 tests, with 1000 permutations for each test. Columns: number of groups, rows: number of variables. The mean number of significant tests under  $H_0$  is equal to 50 and the 99% interval of its theoretical distribution is [33, 69].

$N_v \setminus N_g$	2	3	5	10	30	Tot
100	35	44	54	40	36	209
300	48	49	39	41	52	229
600	42	45	46	52	48	233
1200	34	42	43	46	47	212
2400	50	38	42	65	46	241
4800	56	50	41	40	48	235
Total	265	268	265	284	277	1359

These three tables show that the *randtest* permutation test always performed as expected: the proportion of tests for which the p-value was lower than a given threshold (0.05, 0.01, and 0.001) was always nearly equal to this threshold, in all the tested situations, *i.e.*, for all the numbers of individuals, groups, and variables. This was true even in the worst situations ( $N_i/N_v$  ratio=0.06), which demonstrates that this test has a correct level of Type I error and is thus not sensitive to the spurious groups problem of BGA.

Another problem that can affect permutation tests concerns the minimum number of individuals that are needed to obtain the desired number of permutations (generally 1000). If the number of individuals is too low, then the total number of possible permutations will be less than 1000. In this case, the distribution of the criterion might not be approximated satisfactorily, and the result of the test (p-value) could be biased. The minimal number of individuals needed to have at least 1000 distinct permutations for the *randtest* function of the **ade4** package is given in table 4. These numbers are relatively low and should not be an obstacle in usual studies.

**Table 3** Number of *randtest* tests returning a p-value lower than 0.001 on 10000 tests, with 10000 permutations for each test. Columns: number of groups, rows: number of variables. The mean number of significant tests under  $H_0$  is equal to 10 and the 95% interval of its theoretical distribution is [4, 17].

$N_v \backslash N_g$	2	3	5	10	30	Tot
100	10	9	6	9	6	40
300	6	13	6	4	11	40
600	11	10	9	11	9	50
1200	9	15	7	7	10	48
2400	9	8	10	8	5	40
4800	10	8	11	4	9	42
Total	55	63	49	43	50	260

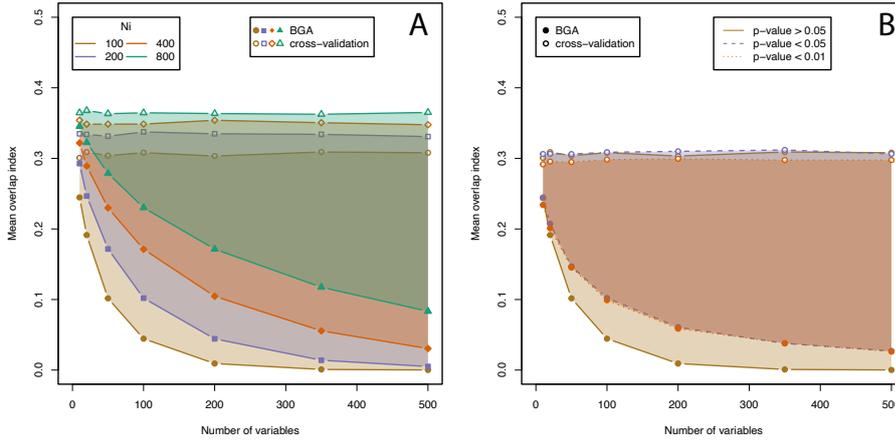
**Table 4** Number of groups, number of individuals per group and total number of individuals needed to have at least 1000 distinct permutations of the *randtest* test of the **ade4** package. This table was obtained by a simulation study.

$N_g$	$N_i/N_g$	$N_i$
2	12	24
3	6	18
4	4	16
5	4	20
6	3	18
7	3	21
8	3	24

### 3.3 Evaluation of the cross-validation procedure

We tested the validity of the *loocv* cross-validation procedure in several situations, by doing the BGA and cross-validation of tables of random normal variables. The number of groups in the simulated data tables was kept constant and equal to five. The total number of individuals ( $N_i$ ) was equal to 100, 200, 400 and 800. The number of variables ( $N_v$ ) was equal to 10, 20, 50, 100, 200, 350, and 500. Results are summed up in Figure 2, that shows the variations of the mean overlap index for the BGA (full symbols) and for the cross-validation (open symbols) when the number of variables in the random numbers table increases, and for several values of the number of individuals. The overlap index of BGA is directly a measure of the spurious group effect: as groups are separating, the overlap index decreases.

Figure 2A shows that the number of variables has no influence on the mean overlap index of cross-validation (curves with open symbols), demonstrating that cross-validated factorial maps are not sensitive to the spurious group effect. The mean overlap index of cross-validation slightly increases (from 0.31 to 0.37) as the number of individuals increases from 100 to 800. This result was expected for random number tables, because the probability of any individual being nearer to another group increases with the number of individuals. The mean overlap index of the BGA (curves with full symbols) decreases as the number of variables increases, because the groups are more and more separated by the spurious groups effect in the BGA space. For tables with only 100 individuals (brown curves and circles), the mean overlap index of BGA decreases very fast when the number of variables increases, but the



**Fig. 2** Values of the mean overlap index for the BGA (full symbols) and for the corresponding cross-validation (empty symbols) for increasing numbers of variables and individuals. The number of groups is always equal to five and each point represents the mean of 100 simulations. The number of variables (x-axis) varies from 10 to 500. A: The number of individuals ( $N_i$ ) is equal to 100 (brown curves and circles), 200 (blue curves and squares), 400 (red curves and diamonds) and 800 (green curves and triangles). B: Dashed and dotted lines also correspond to the overlap index computed for tables of random numbers (100 individuals), but these particular tables were selected by an iterative procedure to give significant (p-value < 0.05, dashed blue lines) and very significant (p-value < 0.01, dotted red lines) permutation tests of the difference between groups.

decrease gets slower as the number of individuals increases. For tables with 500 variables, the mean overlap index of BGA stays very low even with 800 individuals (green curves and triangles). This means that increasing the number of individuals is not enough to overcome the spurious groups problem.

As a consequence of trends observed for the cross-validation mean overlap index ( $xval\bar{O}_{ij}$ ) and the BGA mean overlap index ( $bga\bar{O}_{ij}$ ), the statistic  $\Delta\bar{O}_{ij}$  can be used as a spuriousness index. For 800 individuals and 200 variables it is equal to 39% of the maximum theoretical possible spurious group effect, and for 500 variables it is equal to 57%. Hence, the values of the spuriousness index also depend on the number of individuals and variables. For example, Figure 1 shows that it can reach 70% for  $N_i/N_v = 50/360$ .

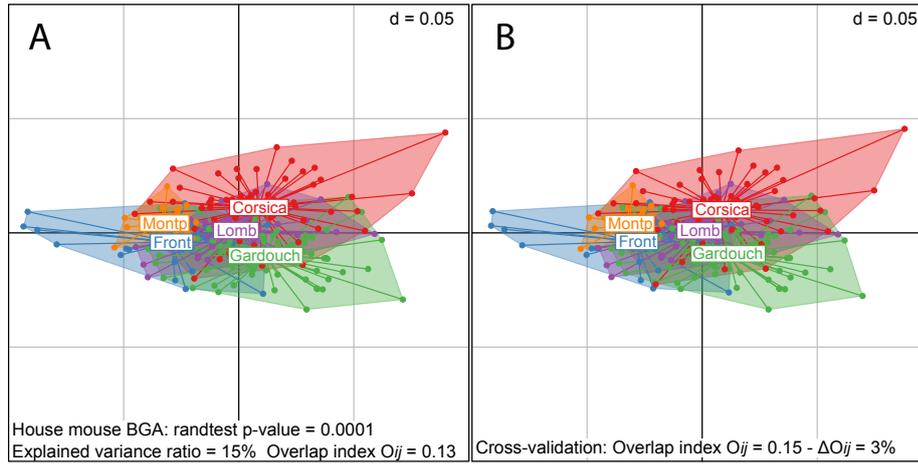
Lastly, it is important to note that all trends reported above are also observed if the  $R^2$  permutation test returns significant results (type I error, dotted and dashed lines on Figure 2B). This shows that spurious groups can be detected by a high value of  $\Delta\bar{O}_{ij}$  even when the permutation test is highly significant.

### 3.4 Simulation with group differences

When the number of variables is high and the groups are only slightly different it may be difficult to disentangle the spurious group effect from the small true differences between the groups. This situation was simulated by gradually adding several sets of columns of random numbers to the house mouse data. These columns of random

numbers were generated in two ways: with and without covariance structure. Our aim was to check the ability of the BGA to recover the groups structure when random noise is added to the data set, and to evaluate the capacity of cross-validation to remove the spurious group effect while keeping the biological signal identifiable.

### 3.4.1 Analysis of the house mouse data



**Fig. 3** Factor map (first two axes) of the BGA on the house mouse data set (5 regions, 214 individuals, 128 variables). The percentages of projected inertia on the two axes are equal to 68% (axis 1, horizontal) and 19% (axis 2, vertical). The scale of the graph is given by the value  $d$  in the upper right corner, it corresponds to the size of the background grid mesh. The five regions are represented by convex hulls surrounding all the individuals of each region. A: factor map of the BGA. The  $p$ -value of the  $R^2$  permutation test is equal to 0.0001, the percentage of explained variance is equal to 15% and the mean overlap index  $bga\bar{O}_{ij}$  is equal to 0.13. B: cross-validated factor map using the *loocv* procedure. The mean overlap index  $xval\bar{O}_{ij}$  is equal to 0.15. The spuriousness index  $\Delta\bar{O}_{ij}$  is equal to 3% and reflects the good agreement between Figure 3A and 3B, confirming the fact that BGA groups are fully supported by cross-validation.

Figure 3 shows the BGA of the house mouse data set. The  $R^2$  permutation test is very highly significant ( $p$ -value  $< 0.0001$ ) and the percentage of between-group variance is equal to 15%. The cross-validated factor map (Figure 3B) is very similar to the original one (Figure 3A), which suggests that the groups observed on Figure 3A are not spurious. The values of the overlap index for the two figures are very similar (0.13 and 0.15) and the spuriousness index  $\Delta\bar{O}_{ij}$  is thus very low (3%). This confirms that the distinction between the five regions is well supported by cross-validation. The null hypothesis of no difference between the five regions can be rejected with a very low risk.

The first axis underlines the geographic differentiation between the two neighboring localities of Montpellier and Frontignan (South-Eastern France along the Mediterranean coast) on the left, opposed to the Gardouch locality (South Western France) and the Corsica populations on the right. Mice from Lombardy (Northern Italy) have an intermediate position. The second axis opposes Corsican populations

(upward) to the mice from the Gardouch locality (downward), revealing the effect of insular differentiation.

Given previous results, this structure was to be expected. Corsican mice display first upper molar with an elongated anterior part (Renaud et al. 2011). In contrast, among continental populations, the mice from Montpellier and Frontignan display a particularly short and massive molar shape (Renaud et al. 2017).

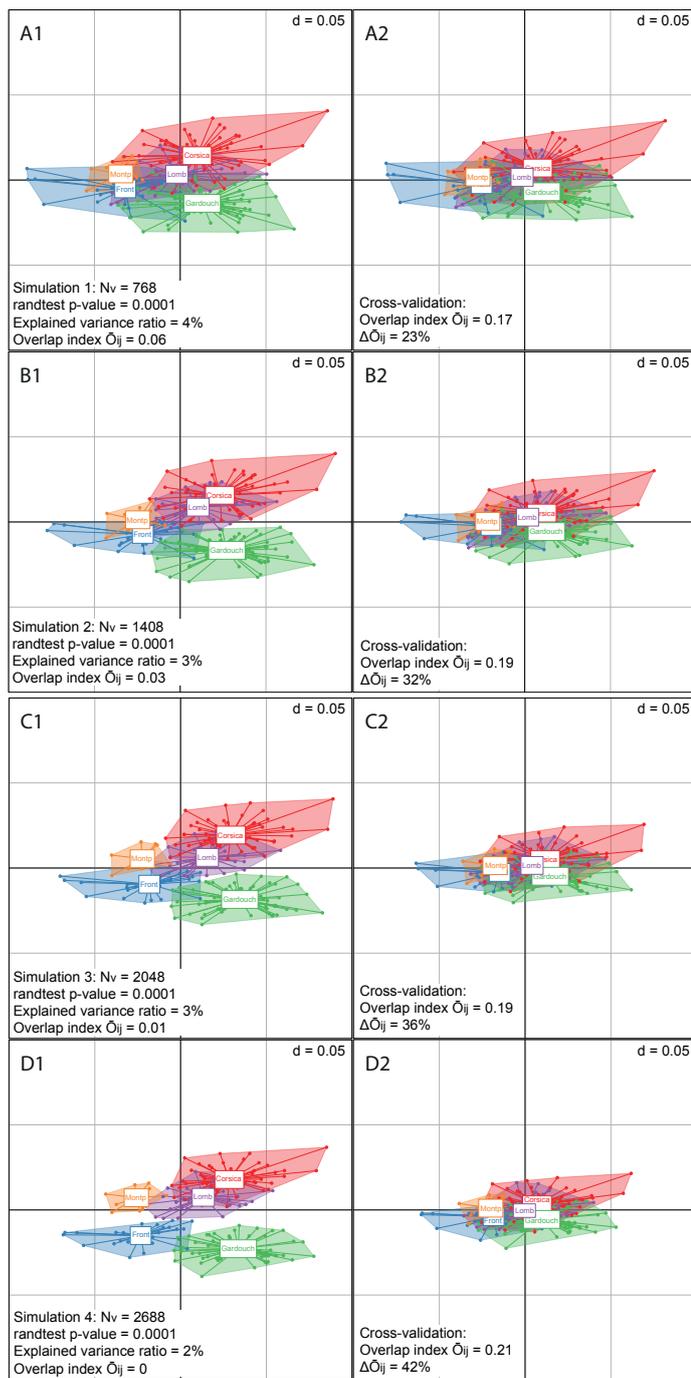
### 3.4.2 Simulations with additional random variables

Figure 4 displays the results of adding columns of random numbers with no covariance structure to the house mouse data table. This figure can be compared for reference to Figure 3 (BGA and cross-validation of the house mouse data table). When columns of random numbers are added gradually (from A to D), the percentage of explained variance drops abruptly (from 15% to 4%, 3% and 2%) because of the strong increase of the total variance due to higher number of random variables. However, it has no effect on the significance of the  $R^2$  permutation test. The p-value stays constant and equal to 0.0001, even for high number of random variables ( $N_i/N_v = 214/2688$ ). Hence, the permutation test stays extremely significant and is still able to detect the differences between the five regions despite the columns of random numbers that were added. The BGA overlap index drops from 0.13 for the house mouse data to 0.06, 0.03, 0.01 and 0 respectively. The cross-validation overlap index increases slowly from 0.15 to 0.17, 0.19 and 0.21. It is also computed in the full BGA space. As a consequence, the spuriousness index ( $\Delta\bar{O}_{ij}$ ) increases from 3% to 23%, 32%, 36% and 42%.

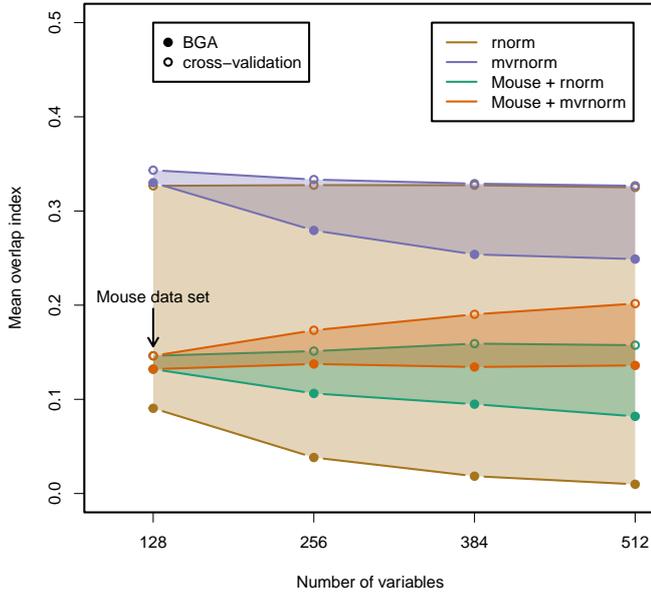
Figure 4D1 (BGA with 2688 variables) shows that the five regions are much more separated than they are on Figure 3. This comes from the spurious groups effect, and Figure 4D2 shows that this effect is adequately fixed by the cross-validation. In fact on Figure 4D2 the five regions are even too close together compared to the original data represented on Figure 3. Cross-validation seems to slightly over-correct the spurious groups effect so that the relative positions of groups are less separated.

Figure 5 shows the values of the mean overlap index of BGA and of cross-validation in several real and simulated situations. For tables of purely random numbers (rnorm, brown curves and circles), the BGA mean overlap index decreases because of the spurious groups effect and the cross-validation adequately fixes it. For tables of purely random numbers but with the same correlation matrix as the mouse data set (mvnorm, purple curves and circles), the mean overlap index also decreases but it stays much higher. When tables of random numbers are successively added to the mouse data table, the spurious groups effect slowly decreases the BGA mean overlap index (green curve and full circles). Cross-validation (green curve and empty circles) fixes this and the mean overlap index is even higher than the one obtained for the mouse data table alone (over-correction). When tables of random numbers with the same correlation matrix as the mouse data table are successively added to the mouse data table (orange curves and full circles), the BGA mean overlap index does not decrease, even for 512 variables. And the cross-validation mean overlap index (orange curves and empty circles) increases even more.

The decrease of the BGA overlap index clearly shows the apparition of the spurious group effect that adds up to the real regions separation (also seen in Figure 4). The cross-validation overlap index increases slightly, and shows the tendency of cross-validation to overcompensate the spurious groups effect. The spuriousness index also



**Fig. 4** BGA (left) and cross-validation (right) factor maps of the house mouse data set with added columns of random numbers with no covariance structure.



**Fig. 5** Mean overlap index of BGA (full circles) and of cross-validation (empty circles) for several real and simulated situations. The result for the house mouse data table is indicated by an arrow (BGA and cross-validation). All other points represent the mean of 100 simulations for BGA and cross-validation, as in Figure 2.

increases, but it stays lower than the one obtained for purely random data. If sets of 128 columns of random numbers with the same covariance structure as the original house mouse data table are added, the same patterns are observed but the values of overlap indexes are higher than for independent random variables (Figure 5). This comes from the fact that the random structures added by random number with no group structure but with a fixed covariance matrix is higher than the variations added by random number drawn from a univariate normal distribution.

Table 5 sums up the values of the spuriousness index ( $\Delta\bar{O}_{ij}$ ) in the four situations described in Figure 5.

#### 4 Discussion

Our results confirm the conclusions of the series of papers about spurious groups in between-group PCA, Bookstein (2019), Cardini et al. (2019) and Cardini and Polly (2020) and the influence of the number of variables. We also confirm the usefulness of cross-validation to support BGA results and the need to use tests in the data space to check the significance of the differences between groups. We propose to use the *randtest* permutation test of the **ade4** package for this, and the mean overlap index  $\bar{O}_{ij}$  computed by the *loocv* function to check out the result of the cross-validation

**Table 5** Values of the spuriousness index ( $\Delta\bar{O}_{ij}$ ) in the situations described in Figure 5. All the tables have 214 individuals (rows) and a number of variables (columns) equal to 128, 256, 384 or 512. The situations are: rnorm = tables of purely random numbers, mvnorm = tables of purely random numbers but with the same correlation matrix as the mouse data table. Mouse + rnorm = House mouse data table plus tables of random numbers. Mouse + mvnorm = House mouse data table plus tables of random numbers with the same correlation matrix as the mouse data table. The 2% increase observed in the last three situations for identical numbers of variables is linked to the cross-validation over-correction.

Situation\Variables	128	256	384	512
rnorm	47%	58%	62%	63%
mvnorm	3%	7%	11%	13%
Mouse + rnorm	3%	9%	13%	15%
Mouse + mvnorm	3%	11%	15%	16%

procedure. The case of unbalanced groups has not been tackled here, but similar simulation strategies could be used for estimating this potential problem.

BGA is not the only method sensitive to the emergence of spurious groups on factorial maps when the number of variables is higher than the number of individuals. We have checked with the *plsDA* function of the **Discriminer** package for **R** (Sanchez, 2013) that spurious groups also appear in PLS-DA (Partial Least Square Discriminant Analysis, see for example Barker and Rayens, 2003). In the *plsDA* function, a cross-validation procedure is available, and it is applied by default, which prevents misinterpretation when using PLS-DA.

The spurious groups effect is a geometrical artifact that can appear in any supervised method, even when the number of individuals is higher than the number of variables. The recent paper by Rohlf (2021) explains very clearly this geometrical effect in high dimension. Cross-validation and permutation tests can easily detect this artifact in random numbers tables, but the problem still remains in real data tables, where true groups can be mixed with spurious groups. In this case, the spurious groups effect enlarges the differences between the real groups and it may be hard to disentangle these two sources, even with the help of the cross-validation procedure. This means that a careful selection of variables and having the highest possible number of individuals is still indispensable. In special cases like data sets from 3D methods, this is very difficult and even more caution should be taken in the interpretation.

Concerning BGA, we have shown that when the *randtest* permutation test is statistically significant, the mean overlap index could be used to measure the discrepancy between BGA coordinates and cross-validation results. This is essential to differentiate true groups from spurious groups, particularly on real data sets, in situations where the number of individuals could be low and the between-group differences weak. We also used the mean overlap index to evidence the influence of the correlation structure between variables, by comparing the case of tables of random numbers, with and without correlation structure, and the case of real data.

The covariance structure between "geomorphometric" variables is widespread. It is partly due to the Procrustes approach itself, which scales, rotates and translates data according to least squares methods, and therefore generates correlations among aligned coordinates. The raise of semi-landmark data, being sampled along curves or surfaces, further introduces covariation due to neighbouring effects. Nevertheless, the covariance structure also partly describes biologically meaningful features, expressing

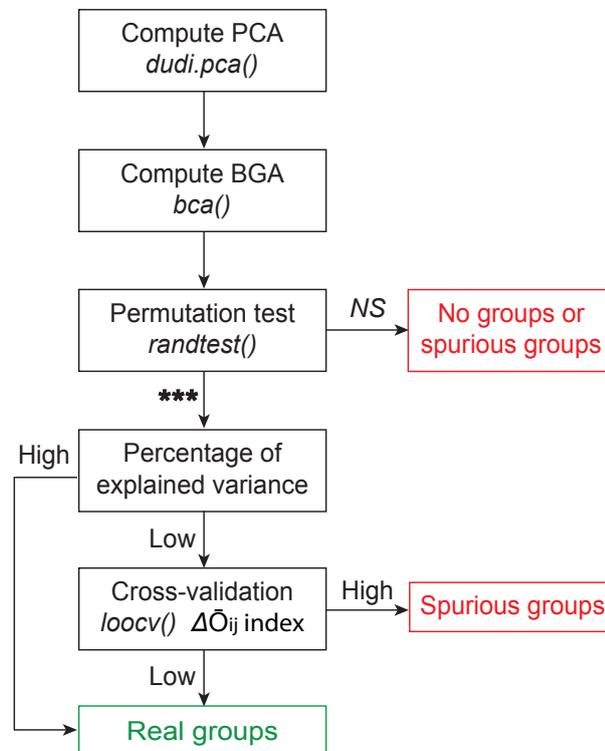
constraints related to the geometry of the object, its function and development, and the underlying standing genetic variation. As such, the covariance structure itself can be the focus of evolutionary studies (Ackermann and Cheverud, 2000; Jamniczky and Hallgrímsson, 2009; Renaud et al., 2017a). Because the main direction of within-group variance corresponds to frequent variants in a population, between-population evolution can be channeled along this phenotypic direction, constituting a line of least resistance to evolution (Renaud and Auffray, 2013; Schluter, 1996).

This covariance structure will differently impact the PCA (and BGA) and the CVA. The PCA, working on the total variance, will be influenced by the structure of the within-group variance. As a consequence, the patterns of between-group differentiation obtained in the space of PCA and BGA axes will tend to be very close (Renaud et al., 2015) and to promote a picture of evolution favored along the line of least resistance constituted by the direction of main within-group variance. In contrast, the CVA, by standardizing within-group variance, will put the focus on evolutionary divergence occurring in other directions. This can evidence changes in more discrete, but still phylogenetically relevant traits (Renaud et al. 2015), although the relationships between groups can be distorted. Since both methods (BGA and CVA) can result in spurious groups, the choice between the two should depend on the role attributed to within-group variance in the interpretation. The present study further demonstrates that such covariance structure in the geomorphometric data will tend to reduce the risk of spurious groups even with low number of individuals, which is a reassuring message for the interpretation of biological data.

We checked that the *randtest* permutation test has good performances in terms of Type I error and power, even in extreme situations (low number of individuals, high number of variables, low p-value threshold). We also showed that the minimum number of individuals needed to compute the test was low enough to be useful even on small data sets. An important consequence is that the  $N_i/N_v$  ratio should not be considered as an immutable constraint. BGA can be used when the number of individuals is low, even with a high number of variables, as long as the spurious group effect is detected and controlled. In the case of non-independent individuals, particularly in the presence of phylogenetic or spatial autocorrelation, additional work will be needed to compare the advantages of the RRPP and MSR procedures. The R code provided in supplementary material shows that the RRPP procedure performs better and is much faster, but MSR is more flexible and can be used in various situations (*e.g.*, controlling simultaneously for spatial and phylogenetic autocorrelations in fourth-corner analysis in Braga et al. (2018) or for spatial autocorrelations in Mantel test in Crabot et al. (2019)).

The *loocv* function can be used for detecting and controlling the spurious group effect: it computes the cross-validation and provides cross-validated factor maps that can be compared to the BGA factor maps. It also provides several statistics to estimate the degree of confidence that can be granted to the BGA. More particularly, the mean overlap index can be used as a measure of the discrepancy between BGA and cross-validation, under the form of the spuriousness index. The integration of these functions in the **ade4** package allows to expand their use to other types of data, particularly to qualitative variables (with Multiple Correspondence Analysis) and to counts tables (with Correspondence Analysis), but also to mixed qualitative/quantitative data sets (Hill & Smith Analysis), and to other methods including those designed for the analysis of two or three tables.

## 5 Conclusion



**Fig. 6** Summed up scheme of the use of BGA proposed in the **ade4** package.

The first aim of this paper was to propose to the morphometricians community a set of tools to overcome the problem of spurious groups in BGA. These tools are integrated in the **ade4** package, and their use is summed up in Figure 6. The first step in this approach should always be to use a permutation test of the BGA, even before looking at graphical outputs. The *randtest* procedure of **ade4** is easy to compute and very fast, even on large data tables, but it assumes that the individuals are independent. If this is not the case, an adapted procedure should be preferred, such as randomization of residuals. The resulting p-value should be compared to a sufficiently low threshold, at least 0.01 for large data sets. If it is higher than this threshold, then this approach should be stopped and BGA output graphs should not be considered. In this case, one should conclude that this data table does not allow to exclude the hypothesis of no difference between groups.

If the p-value of the permutation test is lower than the chosen threshold, then one should look at the percentage of between-group variance and, if it is low, start a cross-validation procedure to make sure that the groups evidenced by the BGA are real groups. This second step is needed for two reasons. First, the test may be statistically significant because of a mix of spurious groups effect and of real group differences. In this case, distances between group means are inflated and the dispersion around means is shrunk, leading to a false impression of large between-group differences. To ensure an adequate interpretation of BGA graphs and particularly of group means relative locations, one should try to evaluate the importance of both effects. The second reason is that on 1000 tests, even if the null hypothesis of no difference between the groups is true, the permutation test (like any other statistical tests) will give approximately 10 p-values lower than 0.01. In this situation, the cross-validation step will prevent drawing erroneous conclusions from the BGA outputs.

If the percentage of between-group variance is low, and if the cross-validation map does not support the BGA groups (spuriousness index higher than *e.g.* 50%), then the existence of these groups should be refuted. We hope that these tools will prove useful to morphometricians and more widely to all the researchers who use BGA.

**Acknowledgements** We are very grateful to Julien Claude and another anonymous reviewer of our manuscript, who helped us to substantially improve it.

### Conflict of interest

The authors declare that they have no conflict of interest.

### References

- Ackermann R, Cheverud J (2000) Phenotypic covariance structure in tamarins (Genus *Saguinus*): a comparison of variation patterns using matrix correlation and common principal component analysis. *American Journal of Physical Anthropology* 111:489–501
- Adams C, Otarola-Castillo E (2013) geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution* 4:393–399
- Almécija S, Tallman M, Alba DM, Pina M, Moyà-Solà S, Jungers WL (2013) The femur of *Orrorin tugenensis* exhibits morphometric affinities with both miocene apes and later hominins. *Nature Communications* 4:2888, DOI 10.1038/ncomms3888
- Barker M, Rayens W (2003) Morphometric Least Squares for Discrimination. *Journal of Chemometrics* 17:166–173
- Bookstein FL (1991) *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, Cambridge (UK), New York
- Bookstein FL (1997) Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Medical Image Analysis* 1:225–243
- Bookstein FL (2019) Pathologies of Between-Groups Principal Components Analysis in Geometric Morphometrics. *Evolutionary Biology* 46:271–302

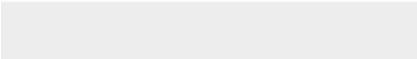
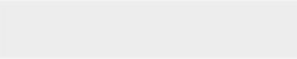
- Braga J, ter Braak CJF, Thuiller W, Dray S (2018) Integrating spatial and phylogenetic information in the fourth-corner analysis to test trait-environment relationships. *Ecology* 99(12):2667–2674, DOI 10.1002/ecy.2530
- Cardini A, Polly PD (2020) Cross-validated Between Group PCA Scatterplots: A Solution to Spurious Group Separation ? *Evolutionary Biology* 47:85–95
- Cardini A, O’Higgins P, Rohlf FJ (2019) Seeing Distinct Groups Where There are None: Spurious Patterns from Between-Group PCA. *Evolutionary Biology* 46:303–316
- Chiari Y, Claude J (2012) Morphometric identification of individuals when there are more shape variables than reference specimens: a case study in Galápagos tortoises. *Comptes Rendus Biologies* 335:62–68, DOI 10.1016/j.crvi.2011.10.007
- Collyer ML, Adams DC (2018) RRPP: An R package for fitting linear models to high-dimensional data using residual randomization. *Methods in Ecology and Evolution* 9(7):1772–1779, DOI 10.1111/2041-210X.13029
- Crabot J, Clappe S, Dray S, Detry T (2019) Testing the Mantel statistic with a spatially-constrained permutation procedure. *Methods in Ecology and Evolution* 10:532–540, DOI 10.1111/2041-210X.13141
- Cucchi T, Kovács Z, Berthon R, Orth A, Bonhomme F, Evin A, Siahsarvie R, Darvish J, Bakhshaliyev V, Marro C (2013) Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Biological Journal of the Linnean Society* 108:917–928
- Cucchi T, Mohaseb A, Peigné S, Debue K, Orlando L, Mashkour M (2017) Detecting taxonomic and phylogenetic signals in equid cheek teeth: towards new palaeontological and archaeological proxies. *Royal Society Open Science* 4:160997, DOI dx.doi.org/10.1098/rsos.160997
- Culhane AC, Perrière G, Considine EC, Cotter TG, Higgins DG (2002) Between-group analysis of microarray data. *Bioinformatics* 18:1600–1608
- Debat V, Bégin M, Legout H, David JR (2003) Allometric and nonallometric components of *Drosophila* wing shape respond differently to developmental temperature. *Evolution* 57(12):2773–2784
- Dianat M, Darvish J, Cornette R, Aliabadian M, Nicolas V (2017) Evolutionary history of the persian jird, *Meriones persicus*, based on genetics, species distribution modelling and morphometric data. *Journal of Zoological Systematics and Evolutionary Research* 55(1):29–45, DOI 10.1111/jzs.12145
- Dolédec S, Chessel D (1987) Rythmes saisonniers et composantes stationnelles en milieu aquatique. I- Description d’un plan d’observations complet par projection de variables. *Acta Oecologica Oecologia Generalis* 8:403–426
- Dolédec S, Chessel D (1989) Rythmes saisonniers et composantes stationnelles en milieu aquatique. II- Prise en compte et élimination d’effets dans un tableau faunistique. *Acta Oecologica Oecologia Generalis* 10:207–232
- Dray S, Pavoine S, Aguirre de Carcer D (2015) Considering external information to improve the phylogenetic comparison of microbial communities: a new approach based on constrained Double Principal Coordinates Analysis (cDPCoA). *Molecular Ecology Resources* 15:242–249, DOI 10.1111/1755-0998.12300
- Evin A, Cucchi T, Cardini A, Vidarsdottir US, Larson G, Dobney K (2013) The long and winding road: identifying pig domestication through molar size and shape. *Journal of Archaeological Science* 40:735–743
- Franquet E, Dolédec S, Chessel D (1995) Using multivariate analyses for separating spatial and temporal effects within species-environment relationships. *Hydrobiolo-*

- gia 300/301:425–431
- Gunz P, Ramsier M, Kuhrig M, Hublin JJ, Spoor F (2012) The mammalian bony labyrinth reconsidered, introducing a comprehensive geometric morphometric approach. *Journal of Anatomy* 220:529–543
- Harbers H, Neaux D, Ortiz K, Blanc B, Schafberg R, Haruda A, Casabianca F, Studer J, Renaud S, Cornette R, Locatelli Y, Vigne JD, Herrel A, Cucchi T (2020) The mark of captivity: plastic responses in the ankle bone of a wild ungulate (*Sus scrofa*). *Royal Society Open Science* 7:192039, DOI 10.1098/rsos.192039
- Jamniczky H, Hallgrímsson B (2009) A comparison of covariance structure in wild and laboratory muroid crania. *Evolution* 63:1540–1556
- Klingenberg CP (2011) MorphoJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources* 11:353–357
- Kovarovic K, Aiello LC, Cardini A, Lockwood CA (2011) Discriminant function analyses in archaeology: are classification rates too good to be true? *Journal of Archaeological Science* 38:3006–3018
- Ledevin R, Koyabu D (2019) Patterns and constraints of craniofacial variation in colobine monkeys: Disentangling the effects of phylogeny, allometry and diet. *Evolutionary Biology* 46:14–34, DOI 10.1007/s11692-019-09469-7
- Leinonen T, Cano J, Mäkinen H, Merilä J (2006) Contrasting patterns of body shape and neutral genetic divergence in marine and lake populations of threespine sticklebacks. *Journal of Evolutionary Biology* 19:1803–1812
- Mitteroecker P, Bookstein F (2011) Linear discrimination, ordination, and the visualization of selection gradients in modern morphometrics. *Evolutionary Biology* 38(1):100–114
- Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O’Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2019) *vegan: Community Ecology Package*. URL <https://CRAN.R-project.org/package=vegan>, r package version 2.5-6
- R Core Team (2020) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, URL <https://www.R-project.org/>
- Renaud S, Auffray J (2013) The direction of main phenotypic variance as a channel to morphological evolution: case studies in murine rodents. *Hystrix, The Italian Journal of Mammalogy* 24:85–93
- Renaud S, Pantalacci S, Auffray J (2011) Differential evolvability along lines of least resistance of upper and lower molars in island house mice. *PLoS ONE* 6, DOI 10.1371/journal.pone.0018951
- Renaud S, Dufour A, Hardouin E, Ledevin R, Auffray J (2015) Once upon multivariate analyses: when they tell several stories about biological evolution. *PLoS ONE* 10, DOI 10.1371/journal.pone.0132801
- Renaud S, Alibert P, Auffray J (2017a) Impact of hybridization on shape, variation and covariation of the mouse molar. *Evolutionary Biology* 44:69–81
- Renaud S, Hardouin E, Quéré J, Chevret P (2017b) Morphometric variations at an ecological scale: Seasonal and local variations in feral and commensal house mice. *Mammalian Biology* 87:1–12
- Renaud S, Ledevin R, Souquet L, Gomes Rodrigues H, Ginot S, Agret S, Claude J, Herrel A, Hautier L (2018) Evolving teeth within a stable masticatory apparatus in Orkney mice. *Evolutionary Biology* 45:405–424

- Rohlf F, Slice D (1990) Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zoology* 39:40–59
- Rohlf FJ (2021) Why Clusters and Other Patterns Can Seem to be Found in Analyses of HighDimensional Data. *Evolutionary Biology* 48:1–16
- Sanchez G (2013) DiscrMiner: Tools of the Trade for Discriminant Analysis. URL <https://CRAN.R-project.org/package=DiscrMiner>, r package version 0.1-29
- Schlager S (2017) Morpho and rvcg – shape analysis in R. In: Zheng G, Li S, Székely G (eds) *Statistical Shape and Deformation Analysis*, Academic Press, pp 217–256
- Schluter D (1996) Adaptive radiation along genetic lines of least resistance. *Evolution* 50:1766–1774
- Siberchicot A, Julien-Laferrière A, Dufour AB, Thioulouse J, Dray S (2017) adegraphics: An s4 lattice-based package for the representation of multivariate data. *The R Journal* 9:198–212, DOI 10.32614 /R.J-2017-042
- Souquet L, Chevret P, Ganem G, Auffray JC, Ledevin R, Agret S, Hautier L, Renaud S (2019) Back to the wild: does feralization affect the mandible of non-commensal house mice (*Mus musculus domesticus*)? *Biological Journal of the Linnean Society* 126:471–486
- Thioulouse J, Chessel D, Dolédec S, Olivier J (1997) ADE-4: a multivariate analysis and graphical display software. *Statistics and Computing* 7(1):75–83
- Thioulouse J, Dray S, Dufour A, Siberchicot A, Jombart T, Pavoine S (2018) *Multivariate Analysis of Ecological Data with ade4*. Springer, NY, New York, USA
- Valenzuela-Lamas S, Baylac M, Cucchi T, Vigne JD (2011) House mouse dispersal in iron age spain: a geometric morphometrics appraisal. *Biological Journal of the Linnean Society* 102:483–497
- Viscosi V, Cardini A (2011) Leaf morphology, taxonomy and geometric morphometrics: A simplified protocol for beginners. *PLOS One* 6, DOI 10.1371/journal.pone.0025630
- Wagner HH, Dray S (2015) Generating spatially constrained null models for irregularly spaced data using Moran spectral randomization methods. *Methods in Ecology and Evolution* 6(10):1169–1178, DOI 10.1111/2041-210X.12407
- Weinberg SL, Darlington RB (1976) Canonical analysis when number of variables is large relative to sample size. *Journal of Educational Statistics* 1(4):313–332

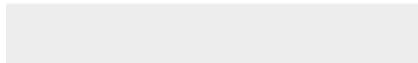


Click here to access/download  
**Supplementary Material**  
xbca.pdf



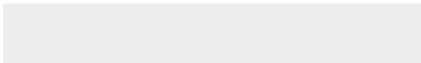


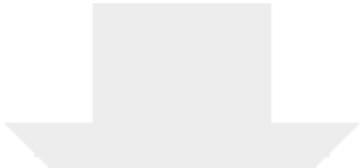
Click here to access/download  
**Supplementary Material**  
SupplementaryFigure1.pdf



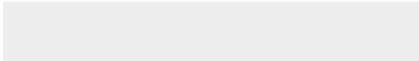
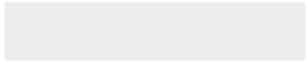


Click here to access/download  
**Supplementary Material**  
SupplementaryFigure1.R



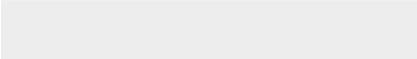
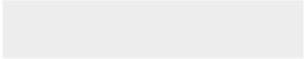


Click here to access/download  
**Supplementary Material**  
facCC.txt





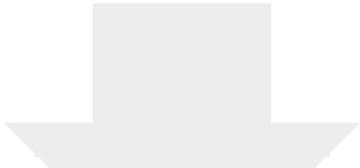
Click here to access/download  
**Supplementary Material**  
tabCC.txt



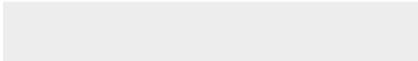


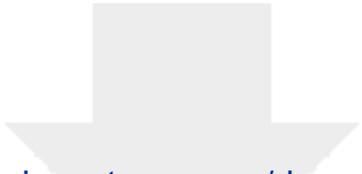
Click here to access/download  
**Supplementary Material**  
Figure5-OverlapSimul.R





Click here to access/download  
**Supplementary Material**  
Figure4-SimulNoCov.R



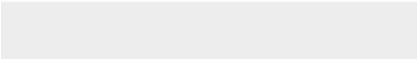
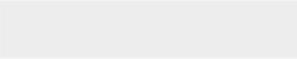


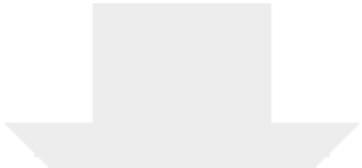
Click here to access/download  
**Supplementary Material**  
Figure2-OverlapIndex.R



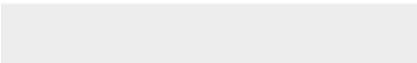


Click here to access/download  
**Supplementary Material**  
Figure1-SpuriousGroups.R



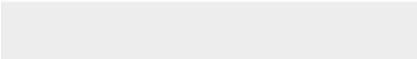
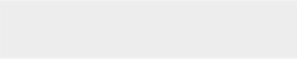


Click here to access/download  
**Supplementary Material**  
Figure3-HouseMouse.R





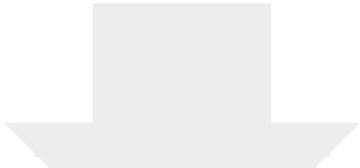
Click here to access/download  
**Supplementary Material**  
Table123-randtest.bcaPar.R



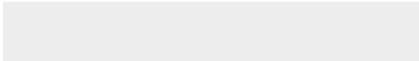
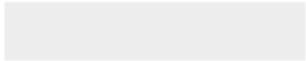


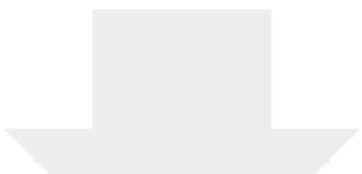
Click here to access/download  
**Supplementary Material**  
Table123-randtest.bca.R





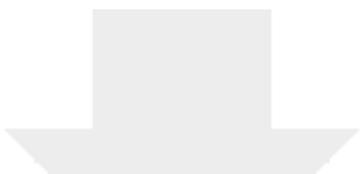
Click here to access/download  
**Supplementary Material**  
RRPPvsMSR.R





Click here to access/download  
**Supplementary Material**  
dOijFig5.txt





Click here to access/download  
**Supplementary Material**  
dOijFig2.txt

