Aust. N. Z. J. Stat. **46**(4), 2004, 615–629

ASSESSING CONGRUENCE AMONG DISTANCE MATRICES: SINGLE-MALT SCOTCH WHISKIES REVISITED

PIERRE LEGENDRE[1]* AND FRANÇOIS-JOSEPH LAPOINTE[1]

Université de Montréal

Summary

A test of congruence among distance matrices is described. It tests the hypothesis that several matrices, containing different types of variables about the same objects, are congruent with one another, so they can be used jointly in statistical analysis. Raw data tables are turned into similarity or distance matrices prior to testing; they can then be compared to data that naturally come in the form of distance matrices. The proposed test can be seen as a generalization of the Mantel test of matrix correspondence to any number of distance matrices. This paper shows that the new test has the correct rate of Type I error and good power. Power increases as the number of objects and the number of congruent data matrices increase; power is higher when the total number of matrices in the study is smaller. To illustrate the method, the proposed test is used to test the hypothesis that matrices representing different types of organoleptic variables (colour, nose, body, palate and finish) in single-malt Scotch whiskies are congruent.

Key words: coefficient of concordance; congruence; data combination; distance matrices; power analysis; single-malt Scotch whiskies; statistical test, simulations.

1. Introduction

During multivariate data analysis, scientists often have to compare matrices containing different types of variables from the same objects and decide if they are congruent with one another, before using them jointly in statistical analysis. Such problems are commonly found in the fields of population genetics, phylogenetics, ecology, anthropology, archaeology, psychometry, sociology and food science. There is thus a need for a test to assess whether several data matrices contain congruent information, i.e. information having mutual agreement or conformity. The Mantel test of matrix correspondence (Mantel, 1967; Mantel & Valand, 1970), which is widely used in the above-mentioned fields, offers a way of testing such a hypothesis for two distance matrices: if two rectangular data matrices contain congruent information, the distances derived from them should be significantly correlated.

This paper proposes a new test of the significance of congruence among distance matrices (CADM). It can be seen as a generalization of the Mantel test to more than two matrices. CADM helps decide whether several distance matrices should be analysed jointly or separately. We show that the proposed test has the correct rate of Type I error over a range of numbers of

Received August 2002; revised July 2003; accepted January 2004.

*Author to whom correspondence should be addressed.

[1] Département de sciences biologiques, Université de Montréal, CP 6128, succursale Centre-ville, Montréal (Québec) Canada H3C 3J7. e-mail: Pierre.Legendre@umontreal.ca.

Acknowledgments. This research was supported by NSERC grant OGP0007738 to the first author and NSERC grant OGP0155251 to the second author. The application portion of this paper was written while the first author was the host of the Department of Statistics, University of Waikato (January and March 2001). The authors thank their Waikato colleagues for their interest in the primary materials discussed in the example.

objects (n) and data matrices (p_D) for any significance level α, as well as good power. We also show that *a posteriori* tests on individual incongruent matrices have, individually, correct Type I error rates in the presence of any number of congruent matrices in the study. Single-malt Scotch whiskies serve to illustrate the applicability of the test to real data: we test the hypothesis that matrices representing different types of organoleptic variables (colour, nose, body, palate and finish), estimated over 109 whiskies, are congruent.

The CADM method has distinct properties that differentiate it from previously described procedures: (i) the test is based on the comparison of distance matrices, each one representing a separate dataset; if distance matrices are computed from rectangular tables of variables, the chosen distance functions must be appropriate to each dataset; (ii) the null hypothesis (H_0) for the test is the lack of congruence (defined above) of all distance matrices; (iii) if the null hypothesis is rejected, *a posteriori* tests of incongruence can be performed for each matrix, in turn, to identify incongruent members in a set of distance matrices; complementary Mantel tests can also be used to identify groups of congruent matrices; (iv) using CADM, data that readily come in the form of distance matrices can be compared to rectangular data tables, which can easily be transformed into distance matrices; observational or experimental data that readily come in the form of distance matrices (e.g. DNA hybridization data) do not have to be artificially transformed into rectangular data tables for comparison; (v) distance matrices derived from data tables containing different numbers of variables receive the same weight in the basic form of the congruence analysis; if this is not deemed appropriate, a weighted version of the test is available.

2. The CADM test

Consider a collection of raw object-by-variable data tables and/or distance matrices obtained from field or laboratory observations. The CADM method proceeds as follows:

1. For each raw data table, compute an $(n \times n)$ distance matrix among the n objects using an appropriate distance function; quantitative variables that are expressed in different physical dimensions should be standardized, or ranged in the [0, 1] interval using their minimum and maximum values, prior to distance calculation. Distance functions can vary from dataset to dataset. With no loss of generality, we talk only about distance matrices in this paper, with the understanding that similarity matrices can be used in the same fashion.
2. Unfold the upper (or lower) off-diagonal portion of each distance matrix into a vector and write this vector to a row (i) of a work table. We assume for the moment that all distance matrices are symmetric. However, non-symmetric distance matrices can also be handled by the method, as shown in Section 7.
3. Transform the values in each row of the work table into ranks.
4. Compute W = Kendall's coefficient of concordance among the unfolded and ranked distance matrices; the formulae are given in Section 3. Transform W into Friedman's χ^2 statistic, which is a pivotal statistic appropriate for testing. This provides the reference statistic (χ^2_{ref}) for the test. Actually, W ($= W_{\text{ref}}$) as well as the sum of squared R_j values (SSR_{ref}) is equivalent to χ^2 for permutation testing, where R_j is the sum of the ranks in each column (j) of the table of ranked values (Section 3). Within a given permutation test, the three statistics W, χ^2 and SSR are monotonic to one another; thus

they are equivalent statistics for permutation testing, producing the same permutational probability.

5. Permute each distance matrix using a 'matrix permutation' procedure, as in the Mantel test (Mantel, 1967; Mantel & Valand, 1970; Legendre, 2000). The need for a permutation test in CADM is discussed at the end of this section. Compute a χ^{2*} (or W^*, or SSR^*) value under permutation. 'Matrix permutation' is an algorithm in which the rows and corresponding columns of the matrix are rewritten as if the objects had been permuted in the original rectangular data matrix and the distances recomputed; in computer programs, even this rewriting step can be avoided by indirect addressing of the distance matrix elements, using a vector of permuted object numbers. Rewrite the permuted distance matrices to the rows of the work table.

   5a. For the global test of significance, all distance matrices are permuted at random, independently of one another. The null hypothesis for this test is the monotonic independence (or incongruence) of all matrices. The alternative hypothesis is that at least two matrices are congruent, having similar rankings of the distances. The test is one-tailed in the upper tail; two matrices with exactly opposite rankings produce a value of 0 for the Kendall statistic.

   5b. In *a posteriori* comparisons, a single distance matrix is permuted at a time. This is repeated for all matrices in turn. The null hypothesis in *a posteriori* tests is the monotonic independence (or incongruence) of the matrix subjected to the test, with respect to all the other matrices in the study. The alternative hypothesis is that this matrix is congruent with at least one other matrix in the set, having similar rankings of the distances (one-tailed test).

6. Repeat step 5 a large number of times to obtain an estimate of the distribution of the χ^2 (or W, or SSR) statistic under permutation. Add the reference value χ^2_{ref} (or W_{ref}, or $\mathrm{SSR}_{\mathrm{ref}}$) to the distribution (Hope, 1968).
7. Calculate the one-tailed probability (P-value) of the data under the null hypothesis as the proportion of values of χ^{2*} (or W^*, or SSR^*) that are larger than or equal to χ^2_{ref} (or W_{ref}, or $\mathrm{SSR}_{\mathrm{ref}}$). The test indicates that the set contains congruent matrices if χ^2_{ref} (or W_{ref}, or $\mathrm{SSR}_{\mathrm{ref}}$) is larger than or equal to most (say, 95% for $\alpha = 0.05$) of the χ^{2*} (or W^*, or SSR^*) values obtained through permutations. If the overall null hypothesis is rejected, *a posteriori* tests can determine which of the individual matrices are congruent.

CADM uses Friedman's χ^2 statistic, or one of the other statistics that are equivalent to it for permutation testing, but in many other respects it differs from Friedman's two-way analysis of variance by ranks, and from the test of Kendall's coefficient of concordance. (i) In CADM, the null hypothesis concerns distance matrices; (ii) testing is done by matrix permutation; (iii) *a posteriori* tests are available in CADM (the latter could also be implemented with Kendall's coefficient of concordance, but to our knowledge it has not been suggested yet in the statistical literature).

Friedman's χ^2 is the statistic used in the CADM test, but the associated P-value is not obtained from the χ^2 distribution. CADM requires a permutation test for the same reason that the Mantel test does: it is the objects that label the rows and columns of a distance matrix that are the permutable units under the null hypothesis, not the individual distance values (Legendre, 2000). A parametric test of the CADM χ^2 statistic would mimic a permutation test where the individual distances would be permuted, which would be wrong since the distances

are not the permutable units. For a symmetric distance matrix, permuting the distance values at random would generate an incorrect permutation set of size $(\frac{1}{2}n(n-1))!$ for n objects, whereas permuting the objects only (which is done by the 'matrix permutation' procedure described above) results in a correct permutation set of size $n!$ (Legendre, Lapointe & Casgrain, 1994). The permutation set is the set of all distinguishable permutations that can be generated by a permutation procedure. Section 9 shows that the permutation test described above has correct rates of Type I error.

3. The Kendall coefficient of concordance

There is a close relationship between Friedman's two-way analysis of variance without replication by ranks and Kendall's coefficient of concordance. They address hypotheses concerning the same data table and they use the same χ^2 statistic for testing. They only differ in the formulation of their respective null hypotheses. To illustrate the difference, consider p judges (rows of the data table) assessing n athletes (columns) in a competition. In Friedman's test, the null hypothesis is that there is no real difference among the n athletes. If that is the case, they should receive random ranks from the various judges, so that the sums of ranks of the athletes should be approximately equal. Kendall's test focuses on the p judges. If the null hypothesis of Friedman's test is true, it means that the judges have produced rankings that are monotonically independent of one another. This is the null hypothesis of Kendall's test of concordance.

Friedman's H_0: The n objects (columns j of the work table) are drawn from the same statistical population.

Kendall's H_0: The p judges (rows i of the work table) produce independent rankings of the objects.

In CADM analysis, the judges (rows of the work table) are replaced by distance matrices; the athletes are replaced by the pairs of objects between which distances are calculated; the objects among which distances are computed must be the same in all distance matrices in the study.

There are two ways found in textbooks for computing Kendall's W statistic (left- and right-hand forms in (1) and (2)); they lead to the same result. One computes first one of the following statistics from the column-marginal sums of ranks R_j received by objects j (Siegel, 1956 p. 234; Siegel & Castellan, 1988 p. 266):

$$S = \sum_{j=1}^{n} (R_j - \bar{R})^2 \quad \text{or} \quad S' = \sum_{j=1}^{n} R_j^2 = \text{SSR}\,. \tag{1}$$

Kendall's W statistic can be obtained from either one of the following formulae:

$$W = \frac{12S}{p^2(n^3 - n) - pT} \quad \text{or} \quad W = \frac{12S' - 3p^2n(n+1)^2}{p^2(n^3 - n) - pT}\,, \tag{2}$$

where T is a correction factor for tied ranks (Siegel, 1956 p. 234; Siegel & Castellan, 1988 p. 266; Zar, 1999 p. 446),

$$T = \sum_{k=1}^{m} (t_k^3 - t_k)\,,$$

in which t_k is the number of tied ranks in each (k) of m groups of ties. The sum is computed over all groups of ties found in all p judges of the data table.

Kendall's W statistic is simply the variance of the column sums of ranks R_j divided by the maximum possible value that this variance can take; this occurs when all judges (or all distance matrices in the CADM test) are in total agreement. Hence $0 \leq W \leq 1$. To obtain the formulae for W given above, one must know that the sum of all ranks in the data table is $\frac{1}{2}pn(n+1)$ and that the sum of squares of all ranks is $\frac{1}{6}p^2n(n+1)(2n+1)$. Friedman's χ^2 statistic is obtained from W using the following formula:

$$\chi^2 = p(n-1)W\,.$$

4. Weighted form of CADM

A weighted form of CADM is obtained by including row weights w_j in the calculation of the column marginal sums of ranks R_j. W must be calculated using the formulae in the left-hand column of Section 3; the formulae in the right-hand column assume equal weights. The sum of the weights should be equal to the number of matrices p. This option is available in the computer program mentioned at the end of the Conclusion. Default values for the weights are 1.

In most cases, users prefer to give equal weights to all distance matrices that are compared in a CADM test, especially when the variables used to obtain the distance matrices are of different natures. This procedure requires fewer statements; it is an application of the principle of parsimony (Ockham's razor) which states that we should make as few assumptions as possible when formulating hypotheses. Still, it may be justifiable in some cases to weight the distance matrices differently. For example, in biological applications, some practitioners prefer to give higher weight to the table of morphological characters than to the table of molecular data, acknowledging the fact that several genes are involved in the coding of each morphological character (de Queiroz, Donoghue & Kim, 1995; Huelsenbeck, Bull & Cunningham, 1996).

The application domain may also suggest weighting the distance matrices proportionally to the number of variables in each data table from which a distance matrix is computed. A better weighting scheme, which takes into account the covariance structure of the variables in each table, is to compute the number of non-zero eigenvalues of a principal component analysis of each data table. The numbers of non-zero eigenvalues, which give the ranks of the tables' covariance matrices, can be used to calculate weights for weighted CADM analysis.

5. *A posteriori* tests of congruence

As mentioned in Section 2, in *a posteriori* comparisons, one distance matrix is permuted at a time, and this is repeated for all matrices in turn. To preserve a correct or approximately correct experimentwise error rate, the probabilities should be adjusted for multiple testing. Wright (1992) recommends the Holm (1979) procedure for non-independent tests. This procedure is less conservative than an ordinary Bonferroni adjustment.

A posteriori tests are useful for identifying the matrices that are not congruent with the other matrices in the study, as can be seen in the example, but they do not tell us which, if any, groups of matrices are congruent among those for which the null hypothesis of independence (or incongruence) is rejected.

6. Complementary information: Mantel tests based upon ranks

There is a close relationship between Spearman's correlation coefficient r_S and Kendall's W statistic: W can be directly calculated from the mean ($\bar{r}_S$) of the pairwise Spearman correlations r_S using the following relationship (Zar, 1999 p. 448):

$$W = \frac{(p-1)\bar{r}_S + 1}{p},$$

where p is the number of variables (or judges) among which Spearman's correlation coefficients are computed. For two variables (or judges) only, W is simply a linear transformation of r_S: $W = \frac{1}{2}(r_S + 1)$. So, in that case, a permutation test of W for two distance matrices is the exact equivalent of a permutation test of r_S for the same matrices.

We transpose this reasoning from simple variables to distance matrices: Mantel tests based upon ranks, using the Spearman statistic (as suggested by Mantel, 1967, and Dietz, 1983), are thus the exact equivalent of CADM tests conducted on two distance matrices. They can be used to determine the groups of congruent distance matrices in studies involving several datasets. Since the alternative hypothesis of the CADM test is one-tailed (see step 5 of the test procedure described in Section 2), the Mantel test of the Spearman statistic should also be one-tailed (H_1: positive correlation between the ranks of the distances in the two matrices).

7. Non-symmetric distance matrices

In biology, laboratory techniques such as DNA-hybridization or comparative serology can produce non-symmetric similarity or distance matrices (e.g. Casgrain *et al.*, 1996; Lapointe, Kirsch & Hutcheon, 1999). Non-symmetric matrices are also known in other fields such as sociometry (e.g. Coleman, 1964 pp. 444–455). If at least one of the matrices is non-symmetric, the full distance matrices, except diagonal entries, may be written to row vectors of the work table, in step 2 of the procedure (Section 2).

Any non-symmetric distance matrix $\boldsymbol{D} = [d_{ij}]$ can be made symmetric by averaging the corresponding entries in the upper and lower triangular portions and writing the result in the upper and lower triangular portions, so that $\boldsymbol{D}' = [\frac{1}{2}(d_{ij} + d_{ji})]$; this is the transformation that should be applied in most cases. Another approach is to compute one of the skew-symmetric matrices $\boldsymbol{D}'' = [\frac{1}{2}(d_{ij} - d_{ji})]$ or $\boldsymbol{D}''' = [\frac{1}{2}(d_{ji} - d_{ij})]$, for $i \leq j$, and use these matrices for CADM analysis. In some cases, the relevant information is found in the skew-symmetric matrices; see for instance Casgrain *et al.* (1996).

The CADM method allows non-symmetric distance matrices to be compared to symmetric or other non-symmetric matrices using all distances in the upper and lower triangular portions of the matrix, as described in step 2 of the CADM test procedure. In our computer program, we made sure that in such a case the statistics computed under permutation only involve permuting the distances written in the same triangular portion of the matrices. Before using this option instead of the averaging option described above, one should make sure that the two triangular portions of the matrix represent clearly different information and, if several matrices are non-symmetric, that all upper triangular portions correspond to one another, and similarly for the lower triangular portions.

8. Simulation procedure

We carried out simulations to check the Type I error and power of the test of congruence among distance matrices. Type I error concerns rejecting the null hypothesis when the data

conform to this null hypothesis. To be valid, a test of significance should have a rate of rejection of the null hypothesis that is not larger than the nominal (α) significance level of the test (Edgington, 1995 p. 37) when the null hypothesis (H_0) is true. On the other hand, a test of significance should be able to reject the null hypothesis when H_0 is false; the frequency of rejection of H_0 in these circumstances is referred to as the power of the test.

Since the null hypothesis of the CADM test is the incongruence of all distance matrices, simulations for Type I error only involve independently-generated distance matrices (independent matrices: IM); p_{IM} is the number of independent matrices in a simulation. To produce each IM matrix, (a) an $(n \times p)$ data matrix is created and filled with pseudo-random N(0, 1) deviates, with n objects in rows and p variables in columns; (b) a distance matrix is created by computing the Euclidean distance among the n objects. The distances within each matrix are transformed into ranks in step 3 of the CADM test.

In simulations to estimate power, the alternative hypothesis is true, meaning that at least some matrices are congruent. These simulations involve various combinations of IM matrices, and some partly similar matrices (PM); p_{PM} is the number of partly similar matrices in a simulation. A set of PM matrices is created as follows: (a) a table containing random normal deviates for n objects and $(p-1)$ variables is generated using a pseudo-random normal N(0, 1) generator. (b) For each member of the set of PM matrices, the following is done: (b.1) copy these $(p-1)$ variables into the $(p-1)$ columns of a new data table; (b.2) fill the pth column using N(0, 1) random deviates; this column differs from one member of the set of PM matrices to another; (b.3) create the PM matrix by computing the Euclidean distance among the n objects. The distances within each matrix are transformed into ranks in step 3 of the CADM test.

For Type I error, simulations are run for different numbers (p_{IM}) of IM. For power, simulations are run for different combinations of IM and PM matrices and numbers of objects (n). The number of variables (p) here is kept to 5 in all simulations. 1000 replicate simulations are run for each result; 999 random permutations are used for the test in each simulation. The rate of rejection of the null hypothesis is given along with its 95% confidence interval.

9. Simulation results

The null hypothesis (H_0) for the CADM test is the incongruence (monotonic independence) of all distance matrices. Hence, in the simulations to measure the rate of Type I error, all matrices are created to be independent of one another. The simulation results show that the test produces correct estimated rates of Type I errors for all significance levels (α) and combinations of number of objects $n = \{5, 10, 20, 50, 100\}$ and number of data matrices $p_{\mathrm{IM}} = \{2, 3, 4, 5, 10, 20\}$. The results for five independently-generated matrices are shown in Figure 1. In all cases, the nominal significance level (α) is included in the 95% confidence interval of the rejection rate; the estimated rejection rate is the proportion of replicate simulations for which the null hypothesis is rejected. These simulations were repeated for data matrices in which the variables were highly correlated within each matrix (but not among matrices). The results were very similar to those reported above. Despite the presence of correlations among variables within a data table, the Euclidean distances among the objects in a file remain unrelated to the distances among the objects in the other files.

The power study is based upon simulations in which the null hypothesis is false by construct; in other words, at least some of the data matrices (p_{PM}) are similar. Simulations for power are run for a total of $p_{\mathrm{IM}} + p_{\mathrm{PM}} = 5$ or 10 matrices. Power of the CADM test is

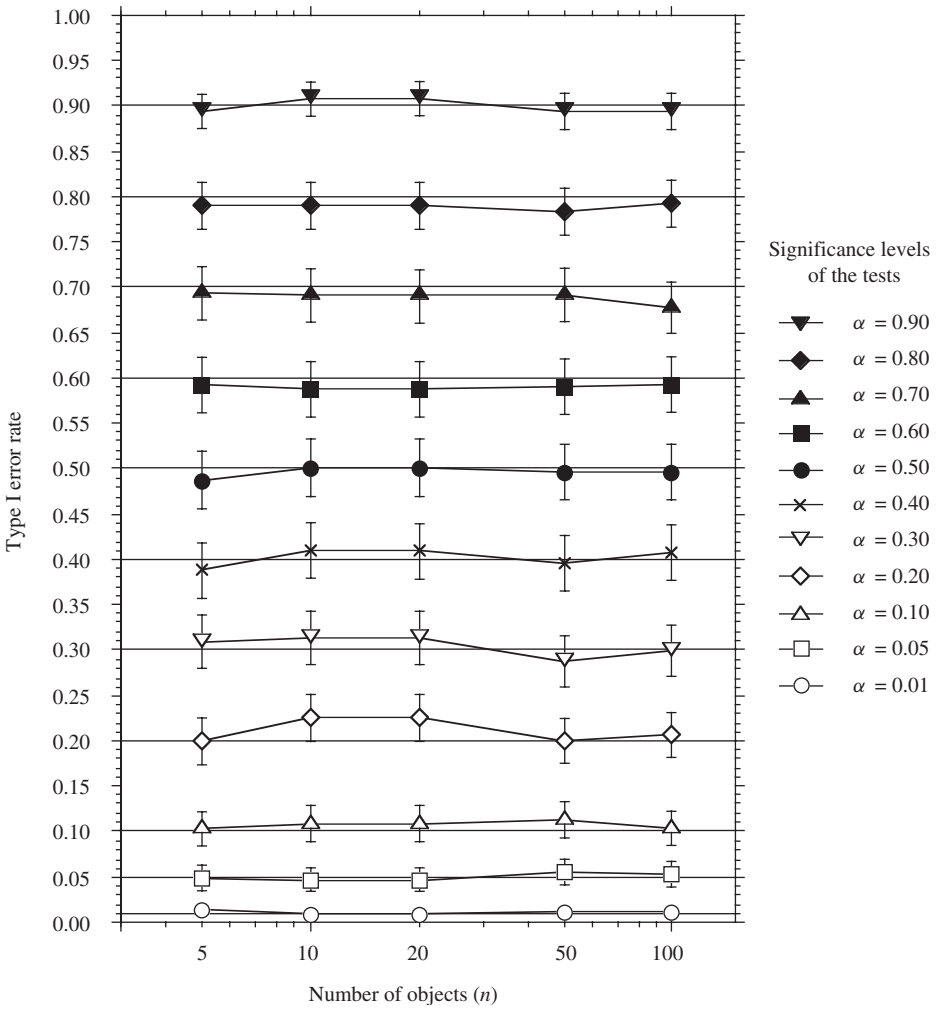

Figure 1. Type I error rates (symbols) and 95% confidence intervals (bars) in CADM simulations involving $p_{\mathrm{IM}} = 5$ independently-generated data tables for various number of objects $n = \{5, 10, 20, 50, 100\}$. The symbols are positioned at the rejection rates of the null hypothesis at the stated significance level.

assessed by permuting at random the distances in all distance matrices, independently of one another, as described in step 5 of the procedure (Section 2). The estimated power is the proportion of the 1000 replicate simulations for which the null hypothesis is rejected (Figure 2). As expected, power increases as the number of objects increases and as the number (p_{PM}) of congruent distance matrices increases. For a given number (p_{PM}) of congruent distance matrices, power is higher when the number of non-congruent matrices in the study (p_{IM}) is smaller. Power is high, for example, when there are $p_{\mathrm{PM}} = 3$ partly similar matrices out of a total of five matrices and $n = 10$ or more, or when $p_{\mathrm{PM}} = 4$ partly similar matrices out of 10 matrices and $n = 10$ or more, or $p_{\mathrm{PM}} = 3$ partly similar matrices out of 10 matrices and $n = 20$.

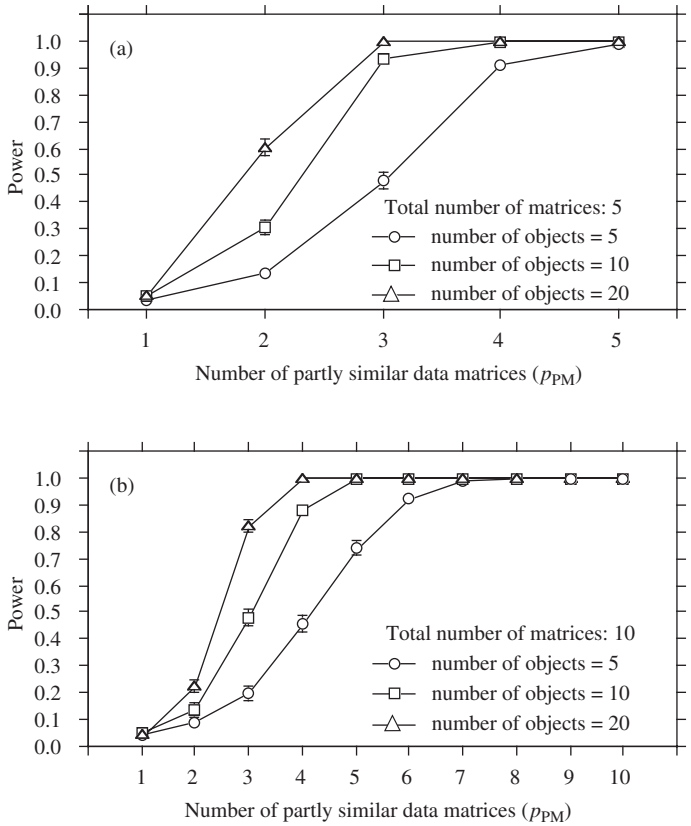

Figure 2. Mean and 95% confidence interval of power of the CADM test (at $\alpha = 0.05$) for various numbers (p_{PM}) of partly similar matrices, for (a) 5 and (b) 10 data matrices. The three curves correspond to different numbers of objects in the matrices. There were 1000 replicate simulations for each combination of parameters. For the leftmost point of each curve, the null hypothesis of independence of the distance matrices is true. The 95% confidence intervals are often so small that the error bars are hidden by the symbols.

Simulations have also been performed to assess the *a posteriori* comparisons obtained by permuting the distance matrices one at a time. The results are presented in Figure 3. When the null hypothesis is true by construct ($p_{PM} = 1$ in the three graphs), the overall test (black circle), in which all matrices are permuted, as well as the tests involving permutation of a single matrix at a time (open symbols), all have correct rejection rates, with values near 0.05 for tests performed using $\alpha = 5\%$ as the significance level. When the null hypothesis is false by construct ($p_{PM} > 1$ in the graphs), permuting one of the 'similar' or 'congruent' matrices produces a test with power greater than α (e.g. open circle and square when $p_{PM} = 2$), but when one of the 'independent' or 'incongruent' matrices is permuted (the other three symbols when $p_{PM} = 2$), the rejection rate is at or near the α significance level. The *a posteriori* tests on individual incongruent matrices have, individually, correct Type I error rates in the presence of any number of congruent matrices in the study. For instance, for $p_{PM} = 3$, permuting any one of the first three matrices (symbols in the graphs: open circle for Rate 1, square for Rate 2, and upward-pointing triangle for Rate 3) rejects the null hypothesis at nearly the same rate as in the global test, but in permutations involving matrices 4 or 5 (symbols in the graph:

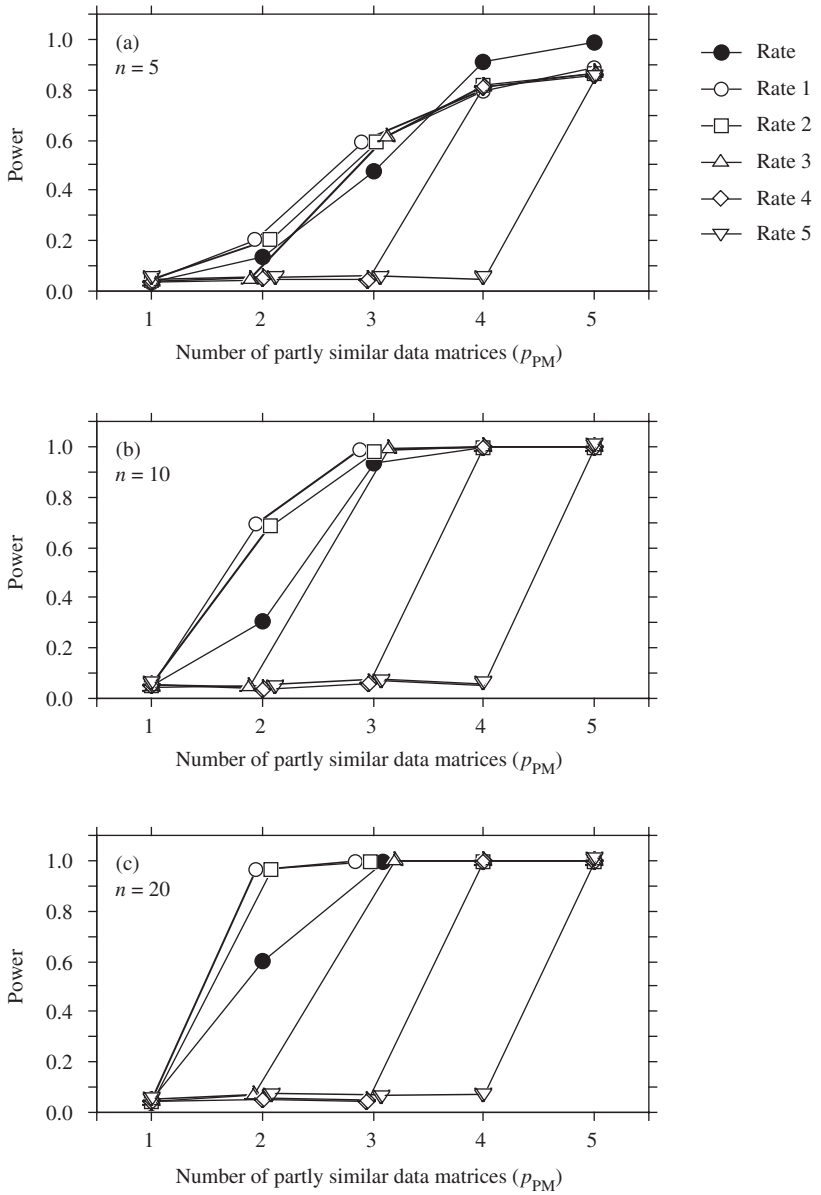

Figure 3. *A posteriori* comparisons in CADM tests, in simulations involving five data tables; n is the number of objects in each table; p_{PM} is the number of matrices that were partly similar in the simulations. When $p_{PM} = 2$, for example, it is the first two matrices that were created to be similar; likewise for $p_{PM} = 3$, 4 or 5. All distance matrices are incongruent when $p_{PM} = 1$. Rate (black circle) is the rejection rate of the global null hypothesis (at $\alpha = 0.05$) after 1000 simulations, computed by permutation of all matrices. Rate 1 (open circle) is the rejection rate of the null hypothesis in tests involving permutations of matrix 1 only; similarly for matrices 2 to 5 (Rate 2 to Rate 5, other open symbols). Some symbols have been moved sideways to improve clarity of the graphs.

diamond for Rate 4 and downward-pointing triangle for Rate 5), the rejection rate is near $\alpha = 0.05$. Power of the *a posteriori* tests increases with n when H_0 is false (compare panels a, b and c of Figure 3). These results show that when the general null hypothesis is rejected, it should be possible in general to identify which of the distance matrices are partly similar and thus congruent, and which are the incongruent ones. This property of the CADM test is further illustrated in the real-case application below.

Additional simulations, not reported here in detail, have shown that the outcome of the CADM test, in terms of Type I error and power, is not affected by the number of variables found in the data tables from which distance matrices are computed. This result was expected: the outcome of a CADM test could not depend on the number of variables in the original data tables, since the method, which only uses distance matrices, does not 'know' how many variables the original datasets contained. In some biological applications, the original data come directly from the lab in the form of a distance matrix (see last paragraph of the Introduction); there are no variables associated with such distance matrices.

10. Interpretation of the CADM test results

CADM results are interpreted as follows:

1. For the overall test of significance, the null hypothesis is the incongruence of all distance matrices. If this hypothesis is not rejected, the matrices should not be used together in statistical analysis. However, if the probability is smaller than or equal to the nominal significance level α (say, 0.05), the null hypothesis is rejected with a probability of Type I error equal to α. The interpretation is that the distance matrices are not all incongruent; there is at least partial congruence among them.
2. The *a posteriori* tests computed to identify the incongruent matrices which cannot be combined in statistical analysis are interpreted as follows. The null hypothesis is that a given matrix is incongruent with respect to all the other matrices in the study.
   - 2a. If the probability is smaller than or equal to the nominal significance level α (say, 0.05), the null hypothesis should be rejected for this matrix with a probability of Type I error equal to α. One concludes that this matrix is congruent with at least one of the other matrices.
   - 2b. If the probability is larger than the nominal significance level α (say, 0.05), the null hypothesis cannot be rejected. One concludes that this matrix differs from all the other distance matrices. The weight of the evidence against the null hypothesis is given by the probability: the higher it is, the more evidence there is that the corresponding distance matrix differs from all the other matrices in the study.
3. In studies involving several datasets, pairwise Mantel tests based upon ranks (Spearman statistics) are useful to determine the groups of congruent distance matrices. When there are many matrices in the study, one can perform clustering or ordination on the table of Mantel statistics; see Section 11 for an example. For smaller sets of matrices, one can simply examine the results of the tests of significance to determine the congruent groups of matrices.

11. Application of CADM: single-malt Scotch whiskies revisited

A few years ago, we published a classification of 109 single-malt Scotch whiskies from 108 distilleries of Scotland (Lapointe & Legendre, 1994), based upon 68 organoleptic variables

derived from tasters' descriptions published in *Michael Jackson's Malt Whisky Companion* (Jackson, 1989). The binary variables (presence–absence, or 1–0) in the five datasets were:

1. Colour (14 variables): white wine, yellow, very pale, pale, pale gold, gold, old gold, full gold, bronze, pale amber, amber, full amber, red, fino sherry.
2. Nose (12 variables): aromatic, peaty, sweet, light, fresh, dry, fruity, grassy, salty, sherry, spicy, rich.
3. Body (8 variables): soft, medium, full, round, smooth, light, firm, oily.
4. Palate (15 variables): full, dry, sherry, big, light, smooth, clean, fruity, grassy, smoky, sweet, spicy, oily, salty, aromatic.
5. Finish (19 variables): full, dry, warm, big, light, smooth, clean, fruity, grassy, smoky, sweet, spicy, oily, salty, aromatic, quick, long, very long, lingering.

We now revisit the whisky data to determine if the five datasets are congruent with one another. We would like to determine if the five datasets contain redundant information in the sense that a qualified taster could, for instance, look at the colour of a whisky and infer many of the characteristics of its nose, body, palate and finish. The analyses published in the above-mentioned classification paper indicated that the 'finish' data contained information somewhat unrelated to the other four datasets. CADM offers a way of conducting a global test of significance of the hypothesis of incongruence of all distance matrices, prior to testing the significance of individual matrices using *a posteriori* CADM tests (H_0: incongruence of the matrix subjected to the test with respect to all the other matrices in the study) or Mantel tests (H_0: monotonic independence of the distances in two matrices).

The Jaccard similarity coefficient, which is appropriate for presence–absence data when double absences should not be emphasized (Legendre & Legendre, 1998 Chapter 7), was computed among whiskies for each dataset separately, producing five similarity matrices. Since we use distances, the similarities were transformed into distances using the transformation $D_{ij} = \sqrt{1 - S_{ij}}$; any other monotonic transformation such as $D_{ij} = 1 - S_{ij}$ would have produced exactly the same CADM results since the distances are transformed into ranks prior to the calculation of the test statistics. Performing the test on the Jaccard similarity matrices would also have produced identical results.

The CADM and Mantel results are shown in Table 1. The overall hypothesis of incongruence of all five distance matrices is rejected, meaning that at least some of the organoleptic groups of variables are congruent with one another. Without or with adjustment for multiple testing, the *a posteriori* tests identify the 'finish' matrix (No. 5) as the incongruent one in the lot. This does not guarantee, however, that the other four distance matrices are all pairwise congruent. The complementary Mantel tests based upon ranks provide the necessary information: matrices 1–4 are congruent to various degrees, except for colour (No. 1) and palate (No. 4) which are incongruent; the permutation probabilities measure the strength of the evidence against the null hypothesis. The 'colour' and 'body' matrices are the most congruent ($P = 0.001$), followed by the 'nose–palate' pair ($P = 0.004$), and finally the pairs 'colour–nose', 'nose–body' and 'body–palate' ($P < 0.05$). These results are illustrated by drawing links representing the strengths of the relationships between matrices on a principal coordinate ordination (Gower, 1966) of the Mantel statistics (Figure 4).

This study shows that organoleptic impressions obtained by smell alone are not sufficient for appreciating the full richness of organoleptic impressions that can be enjoyed during whisky tasting. The finish, or aftertaste, represents a portion of the variation which is incongruent with the other types of data: colour, nose, body and palate. So, to fully appreciate single-

TABLE 1

Results of the (a) overall and (b) a posteriori *CADM tests, using distance matrices derived from the five whisky organoleptic datasets. (c) The complementary Mantel tests are also shown. P = permutational probability, P_H = permutational probability after Holm adjustment, r = Mantel statistic using ranks. All probabilities are based upon 9999 permutations.*

(a) Overall CADM test	H_0: The five distance matrices are incongruent			
Kendall's W:	0.22658			
Friedman's chi-squared:	6667.00524	$P = 0.000$	Reject H_0	
(b) *A posteriori* CADM tests	H_0: This distance matrix is incongruent with the other four			
Matrix 1 (14 colour variables)	$P = 0.002$	$P_H = 0.005$	Reject H_0	
Matrix 2 (12 nose variables)	$P = 0.001$	$P_H = 0.005$	Reject H_0	
Matrix 3 (8 body variables)	$P = 0.001$	$P_H = 0.004$	Reject H_0	
Matrix 4 (15 palate variables)	$P = 0.010$	$P_H = 0.021$	Reject H_0	
Matrix 5 (19 finish variables)	$P = 0.476$	$P_H = 0.476$	Do not reject H_0	

(c) One-tailed Mantel tests H_0: $r = 0$; H_1: $r > 0$
based upon ranks (Spearman correlation)

		Colour	Nose	Body	Palate	Finish
Colour	r	1.0000	0.0344	0.0724	0.0294	–0.0037
	P		0.042	0.001	0.108	0.564
Nose	r		1.0000	0.0479	0.0843	0.0084
	P			0.035	0.004	0.349
Body	r			1.0000	0.0559	0.0153
	P				0.049	0.268
Palate	r				1.0000	–0.0190
	P					0.730
Finish	r					1.0000
	P					

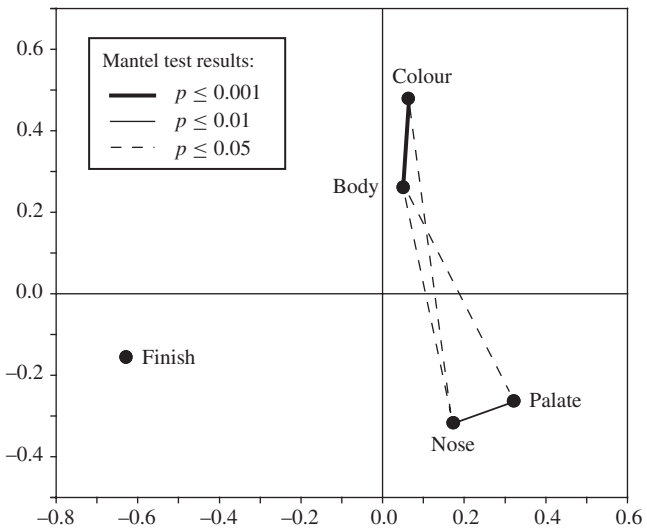

Figure 4. Principal coordinate ordination of the five organoleptic data matrices, based upon the Mantel statistics. Axis I (abscissa) accounts for 28.7% of the variance, and axis II (ordinate) for 26.3%. Results of the Mantel tests that are significant are represented by lines; no adjustment is done for multiple testing since the probabilities are used in the graph in an arbitrary way.

malt Scotch whiskies, they have to be swallowed for the taster to have access to the finish component. This is in agreement with the conclusions of our previous paper (Lapointe & Legendre, 1994).

12. Conclusion

The CADM test opens a new avenue for assessing congruence among several datasets that originate in the form of raw data or distance matrices. Since CADM conducted on two distance matrices is equivalent to a Mantel test based upon ranks, a CADM test for several distance matrices can be seen as a generalization of the Mantel test of matrix correspondence to the multiple-matrix case. In studies involving several datasets, the advantage of CADM over pairwise tests of significance is that the overall ('experimentwise') significance level of the test is correct, whereas pairwise Mantel tests would require a Holm (1979) correction for multiple testing, and even then would only have an approximately correct experimentwise significance level.

The CADM method, with its *a posteriori* tests and Mantel complements, provides a tool for determining which matrices can be used together in an overall multivariate data analysis and which ones should be analysed separately because they contain incongruent information. In the case of raw data tables, rejecting the null hypothesis in the global and all *a posteriori* CADM tests indicates that all variables can be included in a single data table for subsequent analysis. In the case of distance matrices depicting evolutionary data, rejecting the null hypothesis in the global and all *a posteriori* tests indicates that all distance matrices can be used together in a combined analysis. In the field of evolution, another test of the same general hypothesis has been proposed by Huelsenbeck & Bull (1996).

The whisky example was an application with a different perspective: CADM was used to directly test a hypothesis of congruence among five data tables describing the same objects. Another application concerns the classical problem of deciding if datasets can be combined or should be analysed separately. For instance, for an ecological survey repeated at different dates, in which a group of species (e.g. fish) have been counted at several sites, one can use CADM to decide if (and which of) the data tables, collected at the same sites during successive surveys, are congruent and can be combined into a single data table, prior to multivariate analysis. If this is deemed a requirement for fusion of the data, CADM can be complemented by a MANOVA to support the hypothesis that there are no significant differences among the multivariate centroids of the data tables. For community composition data, species counts need to be transformed prior to MANOVA (Legendre & Gallagher, 2001).

A FORTRAN program (CADM: source code, compiled versions for Macintosh and DOS, and program documentation) is available to carry out the calculations described in this paper; see http://www.bio.umontreal.ca/legendre/.

References

CASGRAIN, P., LEGENDRE, P., SIXOU, J-L. & MOUTON, C. (1996). A graph-theory method to establish serological relationships within a bacterial taxon, with example from *Porphyromonas gingivalis*. *J. Microbiol. Methods* **26**, 225–236.

COLEMAN, J.S. (1964). *Introduction to Mathematical Sociology*. New York: The Free Press of Glencoe, Collier-Macmillan.

DE QUEIROZ, A., DONOGHUE, M.J. & KIM, J. (1995). Separate versus combined analysis of phylogenetic evidence. *Ann. Rev. Ecol. Syst.* **26**, 657–681.

DIETZ, E.J. (1983). Permutation tests for association between two distance matrices. *Syst. Zool.* **32**, 21–26.

EDGINGTON, E.S. (1995). *Randomization Tests*, 3rd edn. New York: Marcel Dekker.

GOWER, J.C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**, 325–338.

HOLM, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Statist.* **6**, 65–70.

HOPE, A.C.A. (1968). A simplified Monte Carlo test procedure. *J. Roy. Statist. Soc. Ser. B* **50**, 35–45.

HUELSENBECK, J.P. & BULL, J.J. (1996). A likelihood ratio test for detection of conflicting phylogenetic signal. *Syst. Biol.* **45**, 92–98.

HUELSENBECK, J.P., BULL, J.J. & CUNNINGHAM, C.W. (1996). Combining data in phylogenetic analysis. *Trends Ecol. Evol.* **11**, 152–158.

JACKSON, M. (1989). *Michael Jackson's Malt Whisky Companion: A Connoisseur's Guide to the Malt Whiskies of Scotland.* London: Dorling Kindersley.

LAPOINTE, F-J. & LEGENDRE, P. (1994). A classification of pure malt Scotch whiskies. *Appl. Statist.* **43**, 237–257.

LAPOINTE, F-J., KIRSCH, J.A.W. & HUTCHEON, J.M. (1999). Total evidence, consensus, and bat phylogeny: a distance-based approach. *Mol. Phylogenet. Evol.* **11**, 55–66.

LEGENDRE, P. (2000). Comparison of permutation methods for the partial correlation and partial Mantel tests. *J. Statist. Comput. Simulation* **67**, 37–73.

LEGENDRE, P. & GALLAGHER, E. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia* **129**, 271–280.

LEGENDRE, P. & LEGENDRE, L. (1998). *Numerical Ecology*, 2nd English edn. Amsterdam: Elsevier Science BV.

LEGENDRE, P., LAPOINTE, F-J. & CASGRAIN, P. (1994). Modeling brain evolution from behavior: a permutational regression approach. *Evolution* **48**, 1487–1499.

MANTEL, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220.

MANTEL, N. & VALAND, R.S. (1970). A technique of nonparametric multivariate analysis. *Biometrics* **26**, 547–558.

SIEGEL, S. (1956). *Nonparametric Statistics for the Behavioral Sciences.* McGraw–Hill Series in Psychology. New York: McGraw–Hill.

SIEGEL, S. & CASTELLAN, N.J. JR (1988). *Nonparametric Statistics for the Behavioral Sciences*, 2nd edn. New York: McGraw–Hill.

WRIGHT, S.P. (1992). Adjusted P-values for simultaneous inference. *Biometrics* **48**, 1005–1013.

ZAR, J.H. (1999). *Biostatistical Analysis*, 4th edn. New Jersey: Prentice Hall.