NOTES

Are probability estimates from the permutation model of Mantel's test stable?

DONALD A. JACKSON AND KEITH M. SOMERS

Department of Zoology, University of Toronto, Toronto, Ont., Canada M5S 1A1

Received May 4, 1988

JACKSON, D. A., and SOMERS, K. M. 1989. Are probability estimates from the permutation model of Mantel's test stable? Can. J. Zool. **67**: 766–769.

Matrix comparison tests (i.e., Mantel's test or quadratic assignment) are employed with increasing frequency to measure the concordance between genetic, behavioural, morphological, ecological, and geographic distances. Such tests compare an observed measure of matrix association with a null distribution derived from a randomly generated subset of all possible permutations of one of the original matrices. Typically, 500–2000 randomly permutated matrices are used to generate the statistical distribution and thereby estimate the probability of obtaining an observed association between two matrices. We demonstrate that a considerable error (i.e., 5–6%) may be associated with probability estimates based on such low numbers of permutations. To ensure the stability of the probability estimates (i.e., increase the reliability of the test), we recommend the use of a minimum of 10 000 permutations, and 100 000 permutations if the observed probability approaches a critical significance value (e.g., 0.05).

JACKSON, D. A., et SOMERS, K. M. 1989. Are probability estimates from the permutation model of Mantel's test stable? Can. J. Zool. **67** : 766–769.

On a de plus en plus tendance à utiliser un test de comparaison de matrices (test de Mantel) pour évaluer la concordance entre les distances génétiques, éthologiques, morphologiques, écologiques et géographiques. Ce test confronte la mesure du degré d'association observée entre deux matrices à une distribution nulle dérivée d'un sous-ensemble aléatoire de toutes les permutations possibles de l'une des matrices originales. Typiquement, de 500 à 2000 matrices obtenues par permutations aléatoires donnent lieu à la distribution statistique et elles permettent donc d'estimer la probabilité d'obtenir par hasard la valeur qui a effectivement été observée. Nous démontrons ici qu'une erreur considérable (i.e., 5–6%) peut être associée aux estimations de probabilité basées sur des nombres aussi faibles de permutations. Pour assurer la stabilité des estimations de probabilité (donc pour augmenter la finesse du test), nous recommandons d'utiliser un minimum de 10 000 permutations, et de 100 000 permutations si la probabilité observée avoisine la valeur significative critique (e.g., 0,05).

[Traduit par la revue]

The comparison of two matrices of intersite or intersample distances has been a long-standing problem in biology (e.g., Gower 1971; Sokal 1979). As an example, one might wish to evaluate the similarity between a matrix of genetic or morphological distances among several populations and the actual geographic distances among those same populations. Although we can calculate a correlation coefficient between the distances in the two matrices much like the cophenetic correlation used in cluster analysis (e.g., see Sokal and Rohlf 1962; or Sneath and Sokal 1973, pp. 278–283), we are unable to evaluate the statistical significance of such a correlation because the distances within each matrix are not independent (i.e., the distances ab and ac are not independent of the distance bc; Gower 1971). Mantel (1967) resolved this problem by employing a permutation test to establish the statistical significance of the association between two matrices (also see Mantel and Valand 1970). Mantel's test avoids the problems of matrix interdependence by employing randomization procedures (e.g., see Edgington 1987).

In Mantel's approach, one or more statistics measuring the degree of association between two matrices are compared with a null distribution of the same association statistic. Three statistics are generally available to measure intermatrix association (e.g., Dietz 1983). These statistics are Mantel's Z, Spearman's ρ, and Kendall's τ. The last two are rank correlations, whereas Mantel's Z resembles a product–moment correlation in the sense that it is a summation of cross-products

between the elements of the two matrices. A null distribution is generated by repeatedly calculating the association between one of the original matrices and each of a large number of randomized or reordered versions of the second matrix (see Douglas and Endler 1982 for examples). For small matrices, all permutations of the second matrix are used, but with larger matrices a subset of all possible permutations is employed.

Mantel's test of matrix association was originally based on the assumption that the null distribution derived from a large number of randomized matrices was asymptotically normal (e.g., Mantel and Valand 1970). In this approach, the expected value and standard error of the association statistic are determined with formulae assuming a normal distribution (see Mantel 1967; Douglas and Endler 1982; Manley 1985), and a t-statistic is calculated as the deviation between the observed and expected values divided by the standard error. However, this assumption of normality is incorrect (Mielke 1978) and a permutation-based model is now advocated (e.g., Dietz 1983).

In the permutation model, an expected distribution of a given association statistic is generated by repeatedly randomizing one of the initial distance matrices and calculating a null distribution of expected values from the randomly generated intermatrix associations. Instead of assuming an asymptotically normal distribution, the observed statistic is contrasted with the expected distribution and evaluated according to the probability of obtaining as extreme a measure of association.
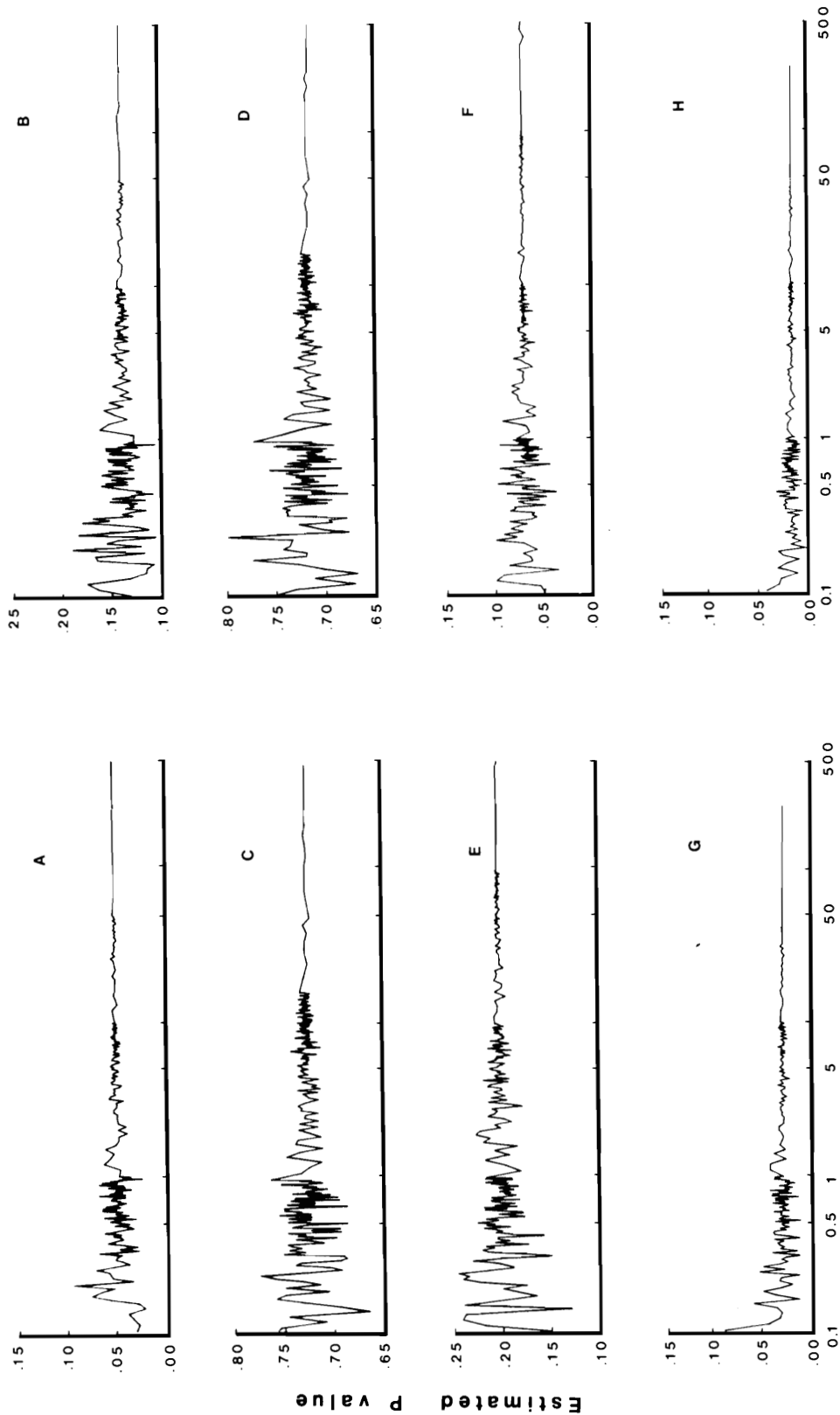
Fig. 1. Estimated probabilities versus the number of permuted matrices for Mantel's statistics (A, C, E, G) and Spearman's rank correlation r_s (B, D, F, H). Data sets are the following: ZOO6 (A, B), BIRD14 (C, D), ANTH19 (E, F), FISH52 (G, H).

Ideally, all possible permutations should be completed with permutation-based tests, but computational requirements (i.e., time) prohibit this approach (e.g., see Dietz 1983). As a result, an arbitrary number of permutations is specified *a priori*, and this number typically ranges from 500 to 2000 (e.g., Dietz 1983; Manley 1986; Harvey et al. 1988).

Sokal (1979) recommended Mantel's test to facilitate studies of geographic variation. Since then, the technique (also known as quadratic assignment; see Dow and Cheverud 1985) has been used by an increasing number of biologists to reveal congruence between matrices of morphometric, genetic, behavioral, and geographic distances (e.g., Douglas and Endler 1982; Dillon 1984; Douglas and Matthews 1985; Schnell et al. 1985, 1986; Lynch and Baker 1986). Unfortunately, with no established criteria for the number of randomly permutated matrices, biologists must arbitrarily choose a reasonably large number of permutations and assume that the results of Mantel's test are unaffected by this choice (Dietz 1983). Because no studies have established whether the number of permutated random matrices affects the estimated significance of an observed measure of intermatrix association (i.e., the reliability of the test), we addressed the following questions: (*i*) does the number of randomly permutated matrices used to generate the expected distribution affect the probability associated with Mantel's test? (*ii*) if the probability does stabilize, what number of randomly permutated matrices is required? and (*iii*) does matrix size affect the number of randomly permutated matrices required to produce stable probability estimates?

Four data sets of different dimensions were chosen to evaluate these questions: (*i*) ZOO6 consists of two 6×6 matrices contrasting zoogeographic distances in fish community composition and Mahalanobis' distances based on lake morphology and water chemistry for six discrete regions in Ontario (Jackson and Harvey 1989), (*ii*) BIRD14 represents two 14×14 matrices of genetic and geographic distances between 14 sparrow populations (Parkin and Cole 1984), (*iii*) ANTH19 contrasts two 19×19 matrices of genetic and anthropomorphic distances between 19 Yanomama Indian villages (Spielman 1973), and (*iv*) FISH52 consists of two 52×52 matrices of distances comparing fish community composition and geographic distances between 52 south central Ontario lakes (D. A. Jackson, unpublished data).

Mantel's tests were generated using the permutation algorithm outlined by Dietz (1983). Expected distributions of intermatrix associations were calculated separately for different numbers of randomly permutated matrices ranging from 100 to 500 000. Because the ZOO6 matrix was small, the exact probability of the observed matrix association was calculated, as well as values for the entire range of permutations. Technically, the additional permutations are unnecessary, yet we are not convinced that biologists using Mantel's test recognize such details, since "rules of thumb" of 500–2000 permutations are evident from slightly larger matrices (e.g., Dietz 1983; Manley 1986). In contrast, it was not feasible to calculate all possible permutations for the larger matrices. For example, the number of possible permutations for BIRD14, ANTH19, and FISH52 is 8.72×10^{10}, 1.22×10^{17}, and 8.07×10^{67}, respectively. Therefore, large numbers of permutations were done with all matrices to ensure comparability among results. In addition, a maximum of 250 000 matrices was calculated for the 52×52 data set because of computational time constraints.

To evaluate the stability of the probability estimates for each data set, the probability levels for two commonly used association statistics, Mantel's Z and Spearman's ρ, were plotted against the number of randomly permutated matrices (Figs. 1A–1H). We omitted a third statistic, Kendall's τ, from this analysis because of the excessive computational requirements (Dietz 1983). Preliminary comparisons between results from Z, ρ, and τ indicated similar patterns of variability (D. A. Jackson and K. M. Somers, unpublished data).

All analyses displayed fluctuating probability estimates that gradually stabilized as the number of matrices increased. Probability estimates based on 500–2000 matrices, the range of values commonly employed in current studies, showed considerable fluctuation (Figs. 1A–1H). Probabilities varied 5–6% for ZOO6, 6–7% for BIRD14, 5–6% for ANTH19, and 3–4% for FISH52. Similar patterns are evident for both Mantel's Z and Spearman's ρ, although the statistic with the smaller probability also tends to vary less. In all cases, variation in the probability estimates is greatly reduced for tests based on 50 000 permutated matrices. Stable probability estimates (i.e., less than 0.1% variability) are evident for tests incorporating 100 000 or more permutated matrices.

Edgington (1987) elaborates on the general utility of randomization methods for testing statistical significance and provides formulae for confidence limits on probabilities derived from randomization tests. Unfortunately, such confidence limits require the exact probability associated with all possible permutated matrices. Variability in the probability estimates for the ZOO6 data (i.e., Fig. 1A) resembles confidence limits based on the formulae given by Edgington (1987, pp. 54–56), emphasizing that some variability is expected, and that this variability depends on the number of permutations and the exact probability (i.e., compare Figs. 1A–1H).

If an unsuspecting biologist chose 1000 randomly permutated matrices as a reasonable number of matrices to generate the expected distribution for Mantel's statistic, the ZOO6 data would indicate a marginally significant association between fish community composition and patterns in lake morphology and water chemistry ($P = 0.047$). However, the selection of 1500 or 2000 matrices would provide different probabilities (i.e., $P = 0.060$ and $P = 0.048$, respectively, whereas the exact probability from the 6! permutations would give $P = 0.0514$). Obviously, reliable inferences from Mantel's test on large matrices require a large number of randomly permutated matrices whenever the observed probability approaches a critical value, such as 0.05. Alternatively, the range in variation in the probability estimate is 5–6% for 500 to 2000 matrices and decreases thereafter. Variability in the probability estimate fell to less than 1% for tests based on 10 000 to 50 000 matrices, and probabilities derived from analyses incorporating 100 000 or more matrices only varied by 0.1%.

While recent studies have focused on choosing between measures of matrix association (e.g., Mantel's Z, Spearman's ρ, Kendall's τ; Hubert 1978; Dietz 1983), we found that the number of matrices used to estimate the distribution of expected values affects probability estimates irrespective of the association measure. Comparisons of results from four data sets from 6 to 52 observations indicated that probability estimates from Mantel's Z and Spearman's ρ statistics stabilized when 100 000 or more matrices were employed. Matrix dimensions showed no obvious impact on the probability esti-

mates, although results from the largest matrices stabilized at smaller numbers of randomly permutated matrices.

Current studies employing Mantel's test use 500 to 2000 random matrices to estimate probabilities. Because probability estimates based on a small number of permutations are unstable, we recommend that 10 000 to 50 000 permutated matrices are the minimum for reliable probability estimates. If the observed probability approaches a critical value (e.g., 0.05), 100 000 randomly permutated matrices may be required to accurately evaluate intermatrix association.

Acknowledgements

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

DILLON, R. T. 1984. Geographic distance, environmental difference and divergence between isolated populations. Syst. Zool. **33**: 69–82.

DOUGLAS, M. E., and ENDLER, J. A. 1982. Quantitative matrix comparisons in ecological and evolutionary investigations. J. Theor. Biol. **99**: 777–795.

DOUGLAS, M. E., and MATTHEWS, W. J. 1985. Does morphology predict ecology? Hypothesis testing within a freshwater fish community. Am. Zool. **25**: 698. (Abstr.)

DOW, M. M., and CHEVERUD, J. M. 1985. Comparison of distance matrices in studies of population structure and genetic microdifferentiation: quadratic assignment. Am. J. Phys. Anthropol. **68**: 367–373.

EDGINGTON, E. S. 1987. Randomization tests. 2nd ed. Marcel Dekker, Inc., New York.

GOWER, J. C. 1971. Statistical methods of comparing different multivariate analyses of the same data. In Mathematics in the archaeo-logical and historical sciences. Edited by J. R. Hodson, D. G. Kendall, and P. Tautu. Edinburgh University Press, Edinburgh. pp. 138–149.

HARVEY, L. E., DAVIS, F. W., and GALE, N. 1988. The analysis of class dispersion patterns using matrix comparisons. Ecology, **69**: 537–542.

HUBERT, L. J. 1978. Generalized proximity function comparisons. Br. J. Math. Stat. Psychol. **31**: 179–192.

JACKSON, D. A., and HARVEY, H. H. 1989. Biogeographic associations in fish assemblages: local versus regional processes. Ecology, **70**. In press.

LYNCH, A., and BAKER, A. J. 1986. Congruence of morphometric and cultural evolution in Atlantic island chaffinch populations. Can. J. Zool. **64**: 1576–1580.

MANLEY, B. F. J. 1985. The statistics of natural selection on animal populations. Chapman and Hall (Meuthen), New York.

——— 1986. Randomization and regression methods for testing for association with geographical, environmental and biological distances between populations. Res. Popul. Ecol. (Kyoto), **28**: 201–218.

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

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

MIELKE, P. W. 1978. Clarification and appropriate inferences for Mantel and Valand's nonparametric multivariate analysis technique. Biometrics, **34**: 277–282.

PARKIN, D. T., and COLE, S. R. 1984. Genetic variation in the house sparrow Passer domesticus, in the East Midlands of England. Biol. J. Linn. Soc. **23**: 287–301.

SCHNELL, G. D., WATT, D. J., and DOUGLAS, M. E. 1985. Statistical comparison of proximity matrices: applications in animal behavior. Anim. Behav. **33**: 239–253.

SCHNELL, G. D., DOUGLAS, M. E., and HOUGH, D. J. 1986. Geographic patterns of variation in offshore spotted dolphins (Stenella attenuata) of the eastern tropical Pacific Ocean. Mar. Mammal Sci. **2**: 186–213.

SNEATH, P. H. A., and SOKAL, R. R. 1973. Numerical taxonomy, the principles and practice of numerical classification. W. H. Freeman and Co., San Francisco.

SOKAL, R. R. 1979. Testing statistical significance of geographic variation patterns. Syst. Zool. **28**: 227–231.

SOKAL, R. R., and ROHLF, F. J. 1962. The comparison of dendrograms by objective methods. Taxon, **11**: 33–40.

SPIELMAN, R. S. 1973. Differences among Yanomama Indian villages: do the patterns of allele frequencies, anthropometrics and map locations correspond? Am. J. Phys. Anthropol. **39**: 461–480.