*Biological Journal of the Linnean Society* (1987) *31*: 75–88. With 6 figures

# Complex clines: the predictivity of complicated patterns of geographic variation portrayed by multivariate analysis

R. S. THORPE

*Department of Zoology, University of Aberdeen, Aberdeen AB9 2TN*

*Received 28 October 1986, accepted for publication 26 January 1987*

The stability of complex patterns of geographic variation was investigated by assessing the congruence between multivariate ordinations derived from randomly chosen real characters. Two series of populations were analysed representing two situations with complex patterns of geographic variation. The first, a 'Eurasian' series of populations, showed a strongly structured hierarchical pattern, the second, an 'eastern' series of populations, showed a more subtle complex pattern of smooth clines and steps. The characters were selected from a total of 81 (Eurasian) or 61 (eastern) within-population independent characters from six different systems. The congruence between ordinations of the geographical populations was measured by the rotational fit statistic, $R^2$. Three procedures were used to compare ordinations based on from two to up to 80 characters randomly chosen to give: A, completely independent character sets; B, subsets compared to the total set; and C, potentially overlapping sets. All three procedures showed that congruence between the ordinations was asymptotic in relation to character number. This relationship was described by one of two mathematical models (procedure B did not result in a hyperbolic model as found with simple patterns of geographic variation). Generally speaking, once a sufficient number of characters are used, the complex patterns of geographic variation are stable, reliable and predictive and not substantially influenced by character choice. The strongly structured hierarchical pattern required 15 or so characters to achieve reliability whilst the more subtle patterns required 20 or so characters. However, the addition of further characters does improve reliability in both cases. The greater percentage of variance portrayed by three-dimensional ordinations compared to two-dimensional ordinations is achieved at the cost of lower congruence when a sufficient number of characters are used. If case studies of geographic variation were to adopt these procedures (preferably using completely independent character sets; procedure A) the reliability of their results would be indicated).

KEY WORDS: Geographic variation – taxonomic congruence – clines – multivariate analysis – race – error estimation – resampling.

CONTENTS





0024–4066/87/050075 + 14 $03.00/0

## INTRODUCTION

Many studies of geographic variation implicitly aim to determine general or predictive patterns of variation. Since each independent character may have a somewhat different pattern of geographic variation, multivariate ordination analysis is widely used to portray general trends (Garrison, 1976; Rightmire, 1976; Schmidly & Hendricks, 1976; Eger, 1977; Iverson, 1977; Birks, 1978; Havera & Nixon, 1978; Jackson, 1978; Matson & Friesen, 1978; Thorpe & McCarthy, 1978; Williams & Genoways, 1978, 1980; Dowler & Genoways, 1979; Parkinson, 1979; Prentice, 1979; Schmidly & Brown, 1979; Smith, 1979; Thorpe, 1979, 1980, 1984a, b, 1986; Crome *et al.*, 1980; Duncan, 1980; Hartmann, 1980; Diersing, 1981; Diersing & Hoffmeister, 1981; Rogers & Schmidly, 1981; Thorpe, Corti & Capanna, 1985; Thorpe, Watt & Baez, 1985).

Previous studies of the relationship between the reliability of patterns of geographic variation portrayed by multivariate analysis and the number of characters employed have been based on a simple pattern, i.e. a smooth cline or a sharp step (categorical variation). These studies (Thorpe, 1985a, b, c), using data from the grass snake, *N. natrix*, show that these simple patterns of geographic variation may be reliably portrayed by multivariate analysis when only a few randomly selected characters are chosen. In these cases the cause of the variation was phylogenesis influencing a wide range of independent features.

These studies were a necessary starting point. However, in nature one rarely finds such simple patterns in isolation. In practice the patterns of geographic variation may turn out to be far more complicated, i.e. various smooth clines and sharp steps in different geographic directions and character dimensions. Are these complex patterns 'stable' and if so how many characters are needed for their reliable portrayal? If the pattern of geographic variation is complicated should it be portrayed by a two or a three dimensional ordination? This paper attempts to contribute towards a more general answer to these questions by determining the specific answers for the complex geographic variation in two series of populations of the grass snake, *N. natrix*.

The mainland Eurasian populations of *N. natrix* are primarily divided into two distinct parapatric semispecies, *N. n. natrix* in the east and *N. natrix helvetica* in the west (Thorpe, 1979). Within the eastern and western forms there is still substantial geographic variation encompassing several traditional subspecies which I do not recognise. In the west there is a sharp step in the Pyrenees between the Iberian populations and those to the north together with a smooth cline from the low countries through to southern Italy (Thorpe, 1984a). The eastern form has a greater range and has an even more complex pattern of geographic variation. There are several lines of divergence from the south-central populations of Greece and Turkey. In the south-east the populations of Syria and south-eastern Turkey diverge; in the north-east the populations diverge clinally from the Caucasus to Siberia; and in the north the populations of central Europe diverge. There is a sharp step between these northern populations and the south-central populations where they meet in the Balkans and north-western Turkey (Thorpe, 1984b).

An analysis of the geographic variation in the mainland Eurasian populations should indicate a clear east–west division irrespective of the substantial variation within the eastern and western forms if the pattern is to be predictive. This strong hierarchical variation is one of the complex patterns to be analysed in relation to character number. The various clines and steps within the eastern form are more subtle and this provides a less strongly structured complex pattern that can be analysed in relation to character number.

The random subsampling and comparison procedure developed in Thorpe (1985a, b, c) and employed in this paper to evaluate the relationship between congruence and character number can be related to the more generally used re-sampling procedures that have gained prominence over the last decade with the availability of inexpensive computation. These re-sampling procedures such as the bootstrap (Efron, 1979) jackknife (Quenouille, 1956; Bissell & Ferguson, 1975; Tukey, 1958), cross-validation (Geisser, 1975; Stone, 1974), repeated replication (McCarthy, 1966) and balanced subsampling (Hartigan, 1969) differ to one another in some respects. They are reviewed by Diaconis & Efron (1983), Efron (1982), Efron & Gong (1983) and Miller (1974). These procedures are not widely used in systematic studies although Gibson, Baker & Moeed (1984) give an example of the use of the jackknife in morphometrics and Felsenstein (1985) the bootstrap in phylogenetic reconstruction. These procedures are generally employed to resample the specimens (cases) to establish the confidence of an estimator with an unknown distribution (although see Felsenstein, 1985). However, in Thorpe (1985a, b, c) and the present paper the cases (group means) are kept constant and the characters are randomly subsampled. Moreover, my procedure involves the direct comparison of random subsets to derive a statistic (congruence) rather than estimating a statistic for each set, and its error or confidence across several sets.

## METHOD

### *Populations*

As indicated in the introduction two series of mainland populations were used, i.e. Eurasian and eastern. The Eurasian series is composed of 26 populations representing both eastern and western forms and the geographic variation within them. The populations, figured in Thorpe (1979), are 1, 3–13, 15, 16, 20–26, 30, 32, 33, 35 and 42. The eastern series is composed of 17 populations representing the variation within this semispecies, i.e. populations 9–13, 20–26, 30, 32, 33, 35 and 42 (also figured in Thorpe, 1979). Every analysis was run for both the Eurasian and eastern series using male specimens.

### *Characters*

The basic set of characters was composed of independent characters (a low within-population correlation) from six character systems, i.e. colour pattern, scalation, internal morphology dentition, dermal sense organs and body proportions adjusted against snout-vent length. Eighty-eight could be recorded from the populations considered in this study. When tested by a one-way analysis of variance significant between group (geographic) variation was found for 81 characters in the Eurasian populations and 61 characters in the eastern populations. These characters are listed in Appendix 1.

### *Multivariate analysis and congruence coefficients*

The pattern of affinities between populations is summarized by principal component/coordinate analysis on normalized group means. This gives very similar results to canonical variate analysis with this data set because of the very low within-population correlation between characters.

When a simple pattern of geographic variation is analysed the pattern can be portrayed in a one-dimensional ordination (Thorpe, 1985a, b, c) using the 'first' coordinate, i.e. the one associated with the largest eigen value. When the pattern is more complex more dimensions are generally required. For complex patterns two dimensional ordinations are frequently used (Garrison, 1976; Eger, 1977; Birks, 1978; Havera & Nixon, 1978; Jackson, 1978; Matson & Friesen, 1978; Thorpe & McCarthy, 1978; Parkinson, 1979; Prentice, 1979; Schmidly & Brown, 1979; Smith, 1979; Crome, Carpenter & Frith, 1980; Thorpe, 1980, 1984a, b; Diersing, 1981; Diersing & Hoffmeister, 1981; Thorpe, Watt & Baez, 1985; Thorpe, 1986) but three dimensional projections are also useful (Schmidly & Hendricks, 1976; Iverson, 1977; Gould & Woodruff, 1978; Williams & Genoways, 1978, 1980; Dowler & Genoways, 1979; Thorpe, 1979; Duncan, 1980; Rogers & Schmidly, 1981; Thorpe, Corti & Capanna, 1982). Three dimensions are the most that can be conveniently illustrated, although there have been attempts at illustrating four or more coordinates as for example in Birks (1978). For a given analysis, increasing the number of dimensions (ordination axes) increases the percentage of the total variation that is represented. However, the reliability of the patterns portrayed in these could increase or decrease with a greater number of dimensions. Consequently, every analysis is run with both two and three dimensional ordinations, i.e. the first two or first three principal coordinates.

Simple patterns, portrayed by single principal coordinates, can be compared by a product-moment correlation coefficient (Thorpe, 1985a, b, c) but the two- and three-dimensional ordinations used in this study are best compared using Gower's (1971) rotational fit, $R^2$ as in Davies (1981), Davies & Boratynski (1979) and Quartau & Davies (1985). The ordinations are scaled, rotated, translated and reflected to minimize the squared distance between corresponding populations. Maximum congruence, that is, identical ordinations, results in an $R^2$ of zero, and increasing values of $R^2$ represent decreasing congruence between the geographic variation as portrayed by the ordinations.

### *Procedure for the selection of character sets and their comparison*

Two procedures, referred to as A and B, have previously been used to investigate the relationship between character number and the congruence of simple patterns of geographic variation (Thorpe, 1985a, b, c). Procedure A is based on the random selection and subsequent comparison (comparison of character sets refers to computing the congruence between the principal coordinate ordinations derived from these character sets) of completely independent character sets of equal size whilst procedure B is based on the comparison of the total character set to a randomly selected subset of characters. The mathematical relationship between these procedures is discussed in Thorpe (1985b). In this study a third procedure, C, is also employed with minimal additional computation. In procedure C, equal size character sets are randomly selected from the total set without the constraint of independence and their ordinations compared. Consequently, in this procedure, when the number of characters, $n$, is low the sets will tend to be independent but as $n$ is increased the sets will tend to progressively overlap until the probability of complete overlap equals unity when $n$ equals $N$ (i.e. the total number of characters). Therefore in procedure C the congruence must be absolute ($R^2 = 0$) when $n$ (the sub-sample size) equals $N$ (the total number of characters).

When different researchers study the geographic variation in an organism the character sets they choose need not be completely independent and if there is a limited number of readily recordable characters the probability of overlap in their character sets will increase with the number of characters employed. Procedure C will tend to reflect this.

*Procedure A: Comparison of independent character sets*

This procedure compares the ordinations derived from completely independent character sets in the following way.

(1) The number of characters, $n$, is a defined set.

(2) $2n$ characters are randomly selected from the total set of significant characters (i.e. 81 for the Eurasian populations and 61 for the eastern populations).

(3) The $2n$ set of characters is divided into two mutually exclusive sets with $n$ characters each.

(4) Principal component/coordinate analyses are run on these two sets of $n$ characters.

(5) The congruence in the pattern of geographic variation is computed as Gower's rotational fit, $R^2$, between these principal coordinate ordinations based on both the first two and first three principal coordinates.

(6) Steps 2–5 are repeated 10 times for a given level of $n$.

(7) The mean congruence (mean $R^2$) is computed for a given level of $n$ across the ten cycles.

(8) Steps 2–7 are repeated with the number of characters, $n$, varied: 2, 3, 4, 5, 6, 8, 10, 12, 15, 20, 20, and 30 for the eastern and Eurasian populations and in addition 35 and 40 for just the Eurasian populations. The maximum number of characters that can be used in this procedure is the integer of half the number of total characters, i.e. 30 for the eastern populations and 40 for the Eurasian populations.

*Procedure B: Comparison of a subset to the total set of characters*

In B the same general procedure as above is followed except that in step 5 the principal coordinates produced from the character subsets are not compared to one another but to the principal coordinates produced from the total set of characters. For the ten cycles in step 6 this gives 20 comparisons to the total analysis. Since procedure A stops at half the number of characters this is extended (ten cycles at a time) for 35, 40, 50 and 60 characters for the eastern populations and extended for 50, 60, 70 and 80 characters for the Eurasian populations.

*Procedure C: Comparison of potentially overlapping character sets*

This follows the same general procedure as above except at step 5 potentially overlapping character sets of the same character number are compared. For the eastern populations this gives 180 comparisons from two to 30 characters and 45 comparisons from 35 to 60 characters. For the Eurasian populations this gives 180 comparisons for two to 40 characters and 45 comparisons from 50 to 80 characters.

The mean rotational fit is plotted against the number of characters. A curve is fitted to this scatter according to a given mathematical model by a least squares fit, weighted by the reciprocal of the variance of $R^2$ for a given number of characters. The full scatter and fitted curve is presented for the two dimensional analyses and just the fitted curve for the three dimensional analyses.

There are essentially three facets to the methodology: first, there are two series of groups, an Eurasian series and an eastern series; second, there are three procedures for comparing ordinations derived from randomly selected character sets of varying size, i.e. procedure A, independent sets, procedure B, subset to total set, and procedure C, potentially overlapping sets; and third, the ordination may be based on either the first two or first three principal coordinates. This result is 12 congruence × character number, plots, to which curves are fitted.

## RESULTS

*Eurasian series: procedure A: independent sets*

The congruence between patterns of geographic variation based on entirely independent character sets has an asymptotic relationship to the number of characters in the set. The curvilinear relationship between congruence ($R^2$) and character number ($n$) conforms to the asymptotic formula

$$R^2 = a + b(e^{-cn}). \quad (1)$$

In the two dimensional analysis (Fig. 1) the congruence is poor ($R^2 \simeq 1.0$)

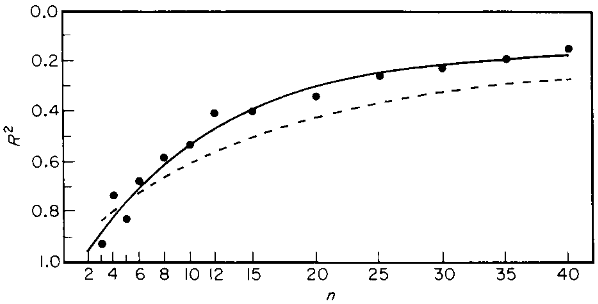


Figure 1. Plots of congruence (Gower's rotational fit, $R^2$) against character number ($n$) for procedure A, Eurasian populations; congruence between ordinations derived from completely independent random character sets. ●, Mean $R^2$ for the two dimensional ordinations; ——, curvilinear fit for the mean $R^2$ of the two-dimensional ordinations; – – – –, curvilinear fit for the mean $R^2$ of the three-dimensional ordinations. See text for interpretation of $R^2$.

Table 1. Estimated values of parameters for the curvilinear fits of congruence ($R^2$) against character number for the two dimensional ordinations

| Procedure | Formula | Fig. | Parameters for 'Eurasian' set | | | Fig. | Parameters for 'Eastern' set | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | a | b | c | | a | b | c |
| A | 1 | 1 | 0.156 | 0.971 | 0.096 | 4 | 0.242 | 1.136 | 3.095 |
| B | 2 | 2 | −0.030 | 2.685 | 0.333 | 5 | −0.108 | 3.348 | 0.301 |
| C | 2 | 3 | −0.098 | 4.084 | 0.300 | 6 | −0.436 | 5.443 | 0.226 |

when the number of characters is low but it improves steadily so that at 15 characters the congruence is fair ($R^2 < 0.4$) and it finally reaches an asymptote of 0.156 (Table 1). The congruence when three dimensions are considered is generally poorer than for two dimensions (when $n = 15$, $R^2 > 0.5$), and has a poorer asymptote ($a = 0.203$).

*Eurasian series: procedure B: subset to total set*

The curvilinear relationship between mean congruence and character number is asymptotic (Fig. 2) conforming to the formula

$$R^2 = a + be^{n^{-c}} \tag{2}$$

In the two-dimensional analysis the $R^2$ is greater than 0.6 when the number of characters is low but it improves to $R^2 \simeq 0.2$ when $n = 15$, to $R^2 \simeq 0.1$ when $n = 25$ and to $R^2 = 0.05$ when half the total number and characters are used and it finally reaches an asymptote of $-0.030$ (Table 1). The curve for the three-dimensional analysis rises less steeply so that at $n = 15$, $R^2 \simeq 0.25$, at $n = 25$, $R^2 = 0.15$, and at half the total number of characters $R^2 \simeq 0.08$.

*Eurasian series: procedure C: potentially overlapping sets*

As expected with low numbers of characters the mean congruence for two dimensions is similar to procedure A, i.e. $R^2 \simeq 1.0$ and with maximum characters the congruence is absolute, $R^2 \simeq 0.0$ as with procedure B (Fig. 3). The mean congruence for two dimensions is asymptotic in relation to character number and improves to $R^2 \simeq 0.35$ when $n = 15$, to $R^2 \simeq 0.2$ when $n = 25$,

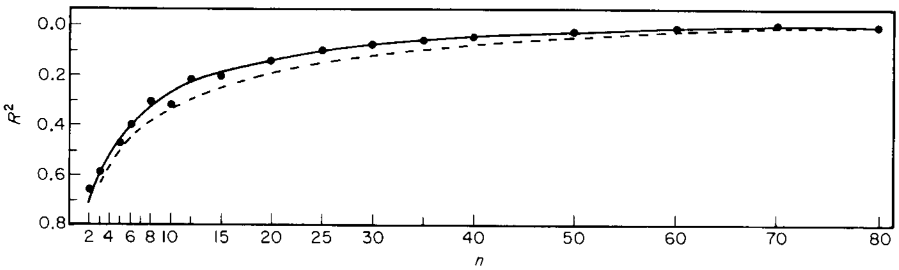


Figure 2. Plots of congruence ($R^2$) against character number ($n$) for procedure B, Eurasian populations; congruence between ordinations derived from the total set and a random subset of characters. See Fig. 1 for explanation of symbols.

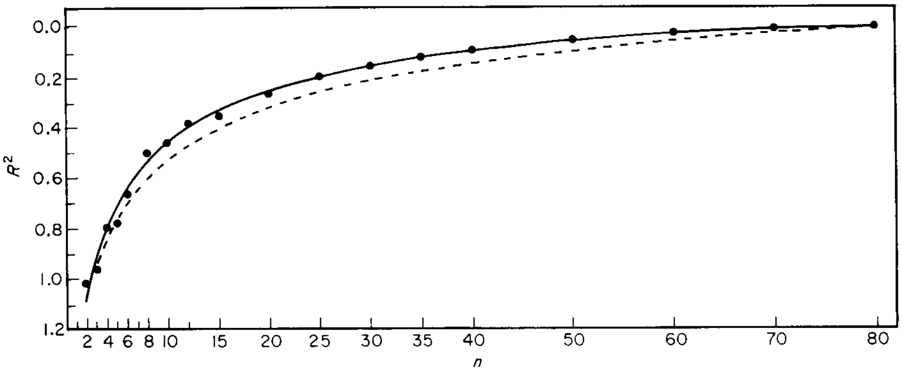


Figure 3. Plots of congruence ($R^2$) against character number ($n$) for procedure C, Eurasian populations; congruence between ordinations derived from potentially overlapping random character sets. See Fig. 1 for explanation of symbols.

and to $R^2 \simeq 0.1$ when half the total number of characters are used and finally reaches an asymptote of $-0.098$ (Table 1). The curve is fitted by Formula 2. The congruence rises less steeply for the three-dimensional analysis such that when $n = 15$, $R^2 > 0.4$, when $n = 25$, $R^2 > 0.25$ and when half the total number of characters are used $R^2 \simeq 0.15$.

*Eastern series: procedure A: independent sets*

Once again the relationship between mean congruence and character number is asymptotic (Fig. 4) and the curve is fitted by Formula 1. The congruence is poor, $R^2 > 1.0$, when the number of characters is low but improves steadily when the numbers of characters is increased to $R^2 \simeq 0.55$ when $n = 15$ and to $R^2 \simeq 0.35$ when $n = 25$, and finally reaches an asymptote of 0.242 (Table 1). The three-dimensional congruence is broadly similar to that for two dimensions

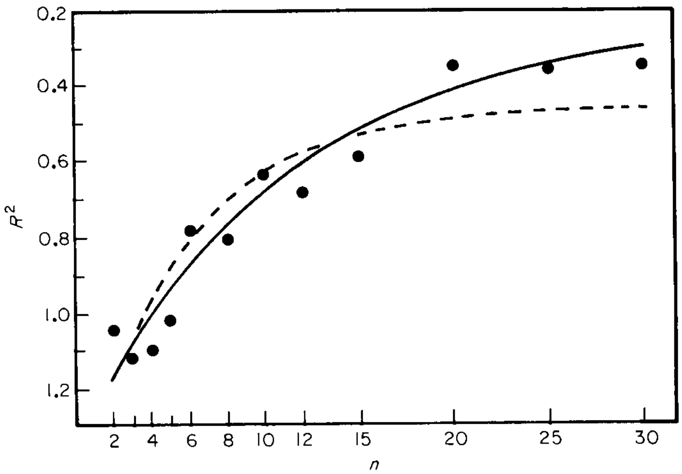


Figure 4. Plots of congruence ($R^2$) against character number ($n$) for procedure A, eastern populations; congruence between ordinations derived from completely independent random character sets. See Fig. 1 for explanation of symbols.

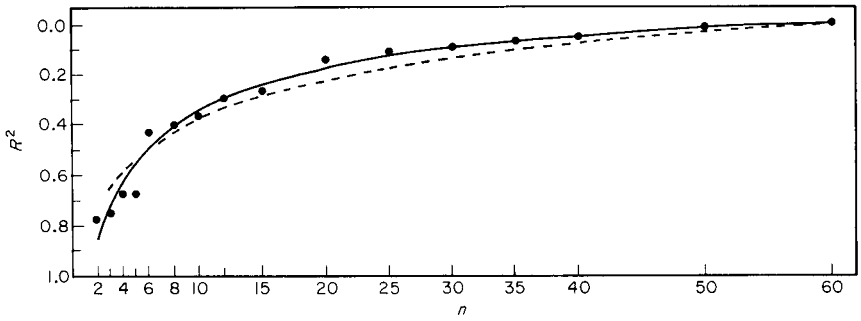


Figure 5. Plots of congruence ($R^2$) against character number ($n$) for procedure B, eastern populations; congruence between ordinations derived from the total set and a random subset of characters. See Fig. 1 for explanation of symbols.

until $n = 15$, but as the number of characters increases the congruence is poorer than that for two dimensions, i.e. the asymptote is 0.461.

*Eastern series: procedure B: subset to total set*

The curvilinear relationship between mean congruence and character number is asymptotic (Fig. 5) and is fitted by Formula 2. In the two-dimensional analysis the congruence rises from $R^2 \simeq 0.8$ when the number of characters is low to $R^2 \simeq 0.25$ when $n = 15$, to $R^2 \simeq 0.12$ when $n = 25$ and $R^2 = 0.10$ when half the characters are used and finally reaches an asymptote of $-0.108$ (Table 1). The congruence rises less steeply in the three-dimensional analysis so that $R^2 \simeq 0.15$ when half the characters are used.

*Eastern series: procedure C: potentially overlapping sets*

The relationship between mean congruence and character number is asymptotic (Fig. 6) and is fitted by Formula 2. In the two-dimensional analysis the congruence is poor ($R > 1.0$) when the number of characters is low and improves to $R^2 \simeq 0.20$ when half the total number of characters are used and

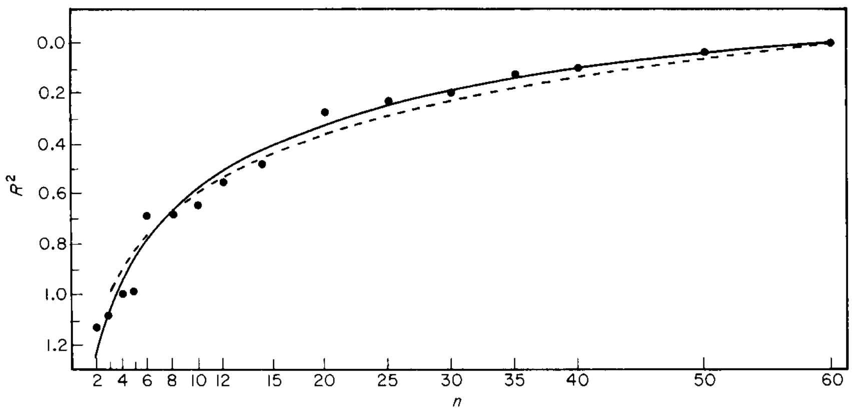


Figure 6. Plots of congruence ($R^2$) against character number ($n$) for procedure C, eastern populations; congruence between ordinations derived from potentially overlapping random character sets. See Fig. 1 for explanation of symbols.

finally reaches an asymptote of −0.436 (Table 1). The congruence for the three-dimensional analysis is generally poorer, reaching $R^2 \simeq 0.25$ when half the total number of characters are used.

## DISCUSSION

### *Two-dimensional ordinations*

The complex patterns of geographic variation within both the strongly structured Eurasian series and more subtly structured eastern series appear to be broadly stable and predictive insofar as the congruence between patterns is asymptotic in relation to character number. That is, there is consistently good mean congruence between patterns of geographic variation portrayed when a sufficient number of characters are used. The comparison of ordinations based on completely independent character sets (procedure A) provides the most rigorous evidence for this. The asymptotic congruence values from procedure A were better for the strongly structured patterns in the Eurasian series than the more subtle pattern of the eastern series, but in both cases the ultimate congruence levels indicated that there was a very small residual element of the pattern that was not stable or predictive but was dependent on the choice of character.

These conclusions are supported by the comparison of subset to total set (B) and of potentially overlapping sets (C). In both cases the congruence must be absolute when the number of characters is maximum. If the congruence increases in a straight line, or only a shallow curve such that the congruence is still improving substantially as the number of characters approaches maximum, then one cannot assume that the pattern of geographic variation would remain the same if characters were added beyond the maximum available in the study. If, on the other hand, the congruence improves very steeply in a sharp hyperbola to give good congruence whilst the number of characters is still low then it is reasonable to assume that the pattern portrayed by the maximum number of characters is stable and is unlikely to substantially alter with the addition of further characters.

In no case for procedures B and C was the curve a sharp hyperbola as was found with the simple patterns of geographic variation (Thorpe, 1985a, b). However, in the strongly structured Eurasian series the curve is quite steep so that at half the total characters the congruence is not very far from absolute. In the eastern series, with its more subtle pattern of variation, the curve is less steep. These results confirm those of procedure A, that is, the patterns are broadly stable (more so in the Eurasian than eastern species) but with a residual element that is dependent on choice of character.

How many characters are needed to portray these complex patterns of geographic variation? One can see from the comparison of the ordinations from independent character sets that beyond 15 characters for the strongly structured Eurasian series and twenty for the subtly structured eastern series the improvement in congruence slows down with the addition of further characters. This is confirmed by the comparison of ordinations in procedures B and C. Consequently one can conclude that complex patterns of geographic variation need between 15 and 20 characters before they are portrayed with reasonable confidence and that the number of characters required depends on whether the patterns are subtle or strong and hierarchical. However, the addition of further characters is not a waste of time as they do result in improved congruence.

### *Two- or three-dimensional ordinations*

With complex patterns of geographic variation one can generally either represent the population affinities in a two- or three-dimensional ordination where the dimensions are the principal coordinates associated with largest eigen values. In an ordination analysis such as principal coordinate analysis the more coordinates that are used the greater the proportion of the total variation that is expressed.

However, it is apparent from all the analyses (Figs 1–6) that when a sufficient number of characters are employed (i.e. *c.* 15 20) the three-dimensional ordinations are less congruent than the corresponding two-dimensional ordinations. Consequently, the desirable increase in information in three-dimensional ordinations has to be balanced against the undesirable decrease in their predictivity or stability. This loss of predictivity (poorer congruence) is probably due to more subtle elements of the geographic variation being revealed by the ordination axes which represent a lower proportion of the total variation. This lower congruence between subtle patterns is also reflected by the lower congruence between ordinations of the subtle geographic patterns of the eastern series compared to the strong geographic patterns of the Eurasian series. In the supraspecific studies of Davies and his co-workers (Davies, 1981; Davies & Boratynski, 1979) there is no general trend for congruence to deteriorate with an increase of ordination axes for either random or specially selected character sets.

It should be borne in mind that with strongly structured hierarchical data, such as the Eurasian series, it is generally not possible to get a good ordination of the more subtle elements of the pattern by plotting the coordinates assocated with progressively smaller eigen values. For example, the Eurasian series is primarily divided into strongly differentiated eastern and western forms. If one requires an optimal ordination within the western or within the eastern forms this is not available by looking at the minor coordinates of an ordination of the entire Eurasian series as they will be a compromise of the within-eastern and within-western variation. Instead, once one has established the existence of two discrete eastern and western forms one must analyse the variation within them separately. This 'resolution' effect is discussed in Thorpe (1976).

In brief, the complex patterns of geographic variation in *N. natrix* are largely stable, 'overall' patterns that can be predicted by *c.* 15–20 characters. Some of the subtle details are not predictable, i.e. not part of a stable, overall pattern. Whilst the pattern of geographic variation in *N. natrix* is predominantly phylogenetically determined (Thorpe, 1984c) the more subtle details may be influenced by random sampling of characters or individuals or natural selection influencing specific characters. Consequently, one cannot expect a wide range of characters to be influenced by these factors and consequently one cannot expect the fine details of a pattern to be predictable. The subtle eastern pattern is less predictable than the strongly structured Eurasian pattern and the three-dimensional ordinations tend to be less predictable than the two-dimensional ordinations because they encompass more subtle elements of the pattern.

Are phylogenetically determined patterns of geographic variation more stable than those caused by natural selection for specific current factors? As a generalization (Thorpe, in press) phylogenetic causation should result in a wide range of independent characters sharing a common, and therefore stable and predictive, pattern of geographic variation whilst characters adapted to different current selection factors should result in lower congruence. The phylogenetically caused geographic variation in the grass snake results in congruent, stable patterns and conforms to the above expectation but studies across a range of species are needed to reinforce this. Moreover, if these procedures, based on independent characters, were applied to case studies of geographic variation one could assess how reliable the results are, something one is unable to do at present. This is particularly important in studies of geographic variation where predictivity is stated or implied, e.g. studies where taxonomic conclusions are drawn.

The re-sampling procedures are pertinent to the general problem of taxonomic congruence. This and previous studies (Thorpe, 1985a, b, c) show, in clear abrogation of Mickevich & Johnson (1976), that multivariate analysis can give very reliable predictive results. On the other hand, Felsenstein's (1985) bootstrap study of phylogenetic reconstruction revealed considerable uncertainty. Intraspecific phylogenetic reconstruction can be extremely useful in some circumstances (e.g. Thorpe, 1984c) but in any event, because the pattern of geographic variation may or may not be phylogenetically caused, one needs a portrayal of the pattern independent of cause and multivariate analysis can supply this.

## ACKNOWLEDGEMENTS

I wish to thank Professor R. G. Davies (Imperial College) for a CDC version of his program computing rotational fit and Aenea Reid (Aberdeen University Computer Centre) for her advice on implementing NAG algorithms on Aberdeen University's computer.

## REFERENCES

BIRKS, H. J. B., 1978. Geographic variation in *Picea abies* (L.) korsten pollen in Europe. *Grana, 17:* 149–160.

BISSELL, A. F. & FERGUSON, R. A., 1975. The jackknife-toy, tool or two-edged weapon? *Statistician, 24:* 79–100.

CROME, F. H. J., CARPENTER, S. M. & FRITH, H. J., 1980. Geographic variation and taxonomy of the spinifex pigeon *Geophaps plumifera*. *Australian Journal of Zoology, 28:* 135–150.

DAVIES, R. G., 1981. Information theory and character selection in the numerical taxonomy of some male Diaspididae (Hemiptera: Coccoidea). *Systematic Entomology, 6:* 149–178.

DAVIES, R. G. & BORATYNSKI, K. L., 1979. Character selection in relation to the numerical taxonomy of some male Diaspididae (Homoptera: Coccoidea). *Biological Journal of the Linnean Society, 12:* 95–165.

DIACONIS, P. & EFRON, B., 1983. Computer-intensive methods in statistics. *Scientific American, 248(5):* 96–108.

DIERSING, V. E., 1981. Systematic status of *Sylvilagus brasiliensis* and *S. insonus* from North America. *Journal of Mammalogy, 62:* 539–556.

DIERSING, V. E. & HOFFMEISTER, D. F., 1981. Distribution and systematics of the masked shrew (*Sorex cinereus*) in Ilinois. *Natural History Miscellanea. The Chicago Academy of Sciences, 213:* 1–11.

DOWLER, R. C. & GENOWAYS, H. H., 1979. Variation in *Pappogeomys castronops* (Geomyidae) on the Llano Estacado of Texas and New Mexico. *Southwestern Naturalist, 24:* 577–602.

DUNCAN, T., 1980. A taxonomic study of the *Ranunculus hispidus* Michaux complex in the Western Hemisphere. *University of California Publications in Botany, 77:* 1–125.

EFRON, B., 1979. Bootstrap methods: another look at the jackknife. *Annals of Statistics, 7:* 1–26.

EFRON, B., 1982. *The Jackknife, the Bootstrap and Other Resampling Plans.* SIAM monograph (Society for Industrial and Applied Mathematics) No. 38. Philadelphia.

EFRON, B. & GONG, G., 1983. A leisurely look at the bootstrap, the jackknife and cross-validation. *American Statistician, 37:* 36–48.
EGER, J., 1977. Systematics of the genus *Eumops* (Chiroptera: Molossidae). *Life Sciences Contributions Royal Ontario Museum, 110:* 1–69.
FELSENSTEIN, J., 1985.Confidence limits for phylogenics: an approach using the bootstrap. *Evolution, 39:* 783–791.
GARRISON, R. W., 1976. Multivariate analysis of geographic variation in *Libellula luctuosa* Burmeister. *Pan-Pacific Entomologist, 52:* 181–203.
GEISSER, S., 1975. The predictive sample Reuse method with applications. *American Statistical Association Journal, 70:* 320–328.
GIBSON, A. R., BAKER, A. J. & MOEED, A., 1984. Morphometric variation in introduced populations of the common myna (*Airidotheres tristis*): an application of the jackknife to principal component analysis. *Systematic Zoology, 33:* 408–421.
GOULD, S. J. & WOODRUFF, D. S., 1978. Natural history of *Cerion* VIII: Little Bahama Bank—a revision based on genetics, morphometrics, and geographic distribution. *Bulletin of the Museum of Comparative Zoology, 148:* 371–415.
GOWER, J. C., 1971. Statistical methods of comparing different multivariate analyses of the same data. In F. R. Hodson, D. G. Kendall & P. Tautu (Eds), *Mathematics in the Archaeological and Historical Sciences:* 138–149. Edinburgh: Edinburgh University Press.
HARTIGAN, J. A., 1969. Using subsample values as typical values. *American Statistical Association Journal, 64:* 1303–1317.
HARTMANN, S. E., 1980. Geographic variation analysis of *Dipodomys ordii* using non-metric cranial traits. *Journal of Mammology, 61:* 436–448.
HAVERA, S. P. & NIXON, C. M., 1978. Geographic variation of Illinois grey squirrels. *American Midland Naturalist, 100:* 396–407.
IVERSON, J. B., 1977. Geographic variation in the musk turtle, *Sternotherus minor*. *Copeia, 1977:* 502–517.
JACKSON, J. J., 1978. Differentiation in the genera *Enyalius* and *Strobilurus* (Iguanidae): Implications for Pleistocene climatic changes in Brazil. *Arquivos de Zoologia São Paulo, 30:* 1–79.
MATSON, J. O. & FRIESEN, R. D., 1978. The subspecific status of *Onychomys torridus clarus* Hollister 1913 (Rodentia: Cricetidae). *Bulletin of the Southern California Academy of Sciences, 77:* 116–123.
McCARTHY, P. J., 1966. Replication (an approach to the analysis of data from complex surveys). *National Centre for Health Statistics, 2(14).*
MICKEVICH, M. F. & JOHNSON, M. S., 1976. Congruence between morphological and allozyme data in evolutionary inference and character evolution. *Systematic Zoology, 25:* 260–270.
MILLER, R. G., 1974. The jackknife a review. *Biometrika, 61:* 1–15.
PARKINSON, A., 1979. Morphological variation and hybridization in *Myotis yumanensis sociabilis* and *Myotis lucifugus carissima*. *Journal of Mammology, 60:* 489–504.
PRENTICE, H. C., 1979. Numerical analysis of infraspecific variation in European *Silene alba* and *S. dioica* (Caryophyllaceae). *Botanical Journal of the Linnean Society, 78:* 181–212.
QUARTAU, J. A. & DAVIES, R. G., 1985. Character selection by information content in the numerical taxonomy of some male Batrachomorphus (Homoptera: Cicadellidae). *Zeitschrift für Zoologische Systematik und Evolutionsforschung 23:* 100–115.
QUENOUILLE, M. H., 1956. Notes on bias in estimation. *Biometrika, 43:* 353–360.
RIGHTMIRE, G. P., 1976. Multidimensional scaling and the analysis of human biological diversity in subsaharan Africa. *American Journal of Physical Anthropology, 44:* 445–452.
ROGERS, D. S. & SCHMIDLY, D. J., 1981. Geographic variation in the white throat woodrat (*Neotoma albigula*) from New Mexico, Texas and northern Mexico. *Southwestern Naturalist, 26:* 167–181.
SCHMIDLY, D. J. & BROWN, W. A., 1979. Systematics of short-tailed shrews (genus *Blarina*) in Texas. *Southwestern Naturalist, 24:* 39–48.
SCHMIDLY, D. J. & HENDRICKS, 1976. Systematics of the southern races or Ord's kangaroo rat, *Dipodomys ordii*. *Bulletin of the Southern California Academy of Sciences, 75:* 225–237.
SMITH, M. F., 1979. Geographic variation in genic and morphological characters in *Peromyscus californicus*. *Journal of Mammology, 60:* 705–722.
STONE, M., 1974. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society, Series B, 36:* 111–147.
THORPE, R. S., 1976. Biometric analysis of geographic variation and racial affinities. *Biological Reviews, 51:* 407–452.
THORPE, R. S., 1979. Multivariate analysis of the population systematics of the ringed snake *N. natrix* (L.). *Proceedings of the Royal Society of Edinburgh, 78B:* 1–62.
THORPE, R. S., 1980. Microevolution and taxonomy of European reptiles with particular reference to the grass snake *Natrix natrix* and the wall lizards *Podarcis sicula* and *P. melisellensis*. *Biological Journal of the Linnean Society, 14:* 215–233.
THORPE, R. S., 1984a. Geographic variation in the western grass snake (*N. natrix helvetica*) in relation to hypothesised phylogeny and conventional subspecies. *Journal of Zoology, London, 203:* 345–355.
THORPE, R. S., 1984b. Multivariate patterns of geographic variation between the island and mainland populations of the eastern grass snake (*N. n. natrix*). *Journal of Zoology, London, 204:* 551–561.

THORPE, R. S., 1984c. Primary and secondary transition zones in speciation and population differentiation: a phylogenetic analysis of range expansion. *Evolution, 38:* 233–243.
THORPE, R. S., 1985a. Character number and the multivariate analysis of simple patterns of geographic variation: categorical or "stepped clinical" variation. *Systematic Zoology, 34:* 127–139.
THORPE, R. S., 1985b. Clines: character number and the multivariate analysis of simple patterns of geographic variation. *Biological Journal of the Linnean Society, 26:* 201–214.
THORPE, R. S., 1985c. The effect of insignificant characters on the multivariate analysis of simple patterns of geographic variation. *Biological Journal of the Linnean Society, 26:* 215–223.
THORPE, R. S., 1986. Evolution and character congruence in some western Indian Ocean *Phelsuma*: numerical analysis of biochemistry, shape and scalation. *Journal of Zoology, London, 208:* 429–441.
THORPE, R. S., in press. Geographic variation: A synthesis of cause, data, pattern and congruence in relation to subspecies, multivariate analysis and phylogenesis. *Bollettino di Zoologia, 53.*
THORPE, R. S. & McCARTHY, C. J., 1978. A preliminary study, using multivariate analysis, of a species complex of African house snakes (*Boaedon fuliginosus*). *Journal of Zoology, London, 184:* 484–506.
THORPE, R. S., CORTI, M. & CAPANNA, E., 1982. Morphometric divergence of Robertsonian populations/species of *Mus*: a multivariate analysis of size and shape. *Experientia, 38:* 920–923.
THORPE, R. S., WATT, K. & BAEZ, M., 1985. Some interrelationships of the Canary Island lizards of the genus *Gallotia. Bonner Zoologische Beiträge, 36:* 577–584.
TUKEY, J. W., 1958. Bias and confidence in not-quite large samples. *Annals of Mathematical Statistics, 29:* 614.
WILLIAMS, S. L. & GENOWAYS, H. H., 1978. Review of the desert pocket gopher, *Geomys arenarius* (Mammalia: Rodentia). *Annals of the Carnegie Museum, 47:* 541–570.
WILLIAMS, S. L. & GENOWAYS, H. H., 1980. Morphological variation in the southeastern pocket gopher, *Geomys pinetis* (Mammalia: Rodentia). *Annals of the Carnegie Museum, 49:* 405–453.

## APPENDIX

*Abbreviations:* No., number; VS, ventral scale; SS, subcaudal scale; Pos., position; Post., posterior; Ant., anterior; DSR, dorsal scale rows; Red., reduction; Rt., right; Lt., left; RATL, regressed against trunk length.

*Eastern series*

The 61 characters used for the eastern series of populations were: No. VS; No. SS; Pos. of Red. from 23–21, 21–19, 19–17, 4–2 DSR; lower DSR lost at 21–19, 19–17 and 8–6 Red.; Size of neck scales RATL; extent of contact between temporal and lower postocular scales; No. sub-labial, Post. temporal and gular scales; extent of eye, supra-labial, sub-labial streaks and longitudinal stripes, No., size, upper and lower Pos. of lateral blotches, size and upper Pos. of dorsal blotches; No. ventro-lateral blotches; extent of temporal occipital marking; separation of lunar marking and its encroachment onto the posterior upper labial scales; size, separation, length, streaking, separation from parietal and lower curvature of nuchal marking; VS position of the thyroid, left lung, systemic junction, pancreas, Post. tip of right lung, Ant. and Post. tip of liver, Ant. and Post. tip of Rt. and Lt. testes and kidneys; No. of renal arteries; difference in No. renal arteries between Rt. and Lt. kidneys; SS Pos. of hemipenes muscle insertion; No. of maxillary, palative, pterygoid and dentary teeth; head depth and length RATL, body width RATL; No. of temporal and postocular pits.

*Eurasian series*

The 81 characters used for the Eurasian series of populations were as above but excludes the difference in No. of renal arteries between Rt. and Lt. kidneys and includes lower DSR lost at 14–12, 12–10 Red.; Pos. of Red. from 12–10, 10–8, 8–6, 6–4 DSR; extent of contact between sub-labial and Ant. chin shields and between temporal and upper Post. ocular scales; extent of white edges to DSR, No. of dispersed lateral blotches; lower Pos. of dorsal blotch; size of ventro-lateral blotches; extent of occipital line and parietal occipital marking; occurrence of lunar marking; lower curvature of nuchal marking; VS Pos. of heart; VS length of cystic duct; size of cloacal gland and hemipenes; head width RATL.