



## A Classification of Pure Malt Scotch Whiskies

Francois-Joseph Lapointe; Pierre Legendre

*Applied Statistics*, Vol. 43, No. 1 (1994), 237-257.

Stable URL:

<http://links.jstor.org/sici?sici=0035-9254%281994%2943%3A1%3C237%3AACOPMS%3E2.0.CO%3B2-%23>

*Applied Statistics* is currently published by Royal Statistical Society.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/rss.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

For more information on JSTOR contact [jstor-info@umich.edu](mailto:jstor-info@umich.edu).

©2003 JSTOR

## General Interest Section

# A Classification of Pure Malt Scotch Whiskies

By FRANÇOIS-JOSEPH LAPOINTE and PIERRE LEGENDRE†

*Université de Montréal, Canada*

[Received November 1992. Revised September 1993]

### SUMMARY

Single-malt Scotch whiskies are produced by 109 distilleries in Scotland. The layman may wonder what are the major types of single malts that can be recognized, and what are their chief characteristics and best representatives, whether there is a geographical component in that classification and whether the various categories of characteristics lead to the same classification. This paper provides an answer to these questions, applying an array of statistical methods to a database derived from a connoisseur's description of these liquors. The taster's literary descriptions of Scotches were turned into a numerical database (109 Scotches  $\times$  68 binary variables). A first classification was produced by distance computation and hierarchical clustering. Since it was significantly related to the regions of Scotland, a second classification was computed with a spatial contiguity constraint, to divide Scotland into regions where Scotches are homogeneous in their organoleptic characteristics. To explore the congruence of the categories of characteristics, the five databases corresponding to nose, colour, body, palate and finish characteristics were compared by using statistical tests of significance: among the raw data tables, among distance matrices and among classifications derived from these distance matrices. Most types of characteristic lead to congruent results, despite the loss of information that occurs when moving from one level to the next.

*Keywords:* Canonical analysis; Classification; Constrained clustering; Double-permutation test; Mantel test; Redundancy analysis; Scotch whisky; Single-malt whiskies

## 1. Introduction

Whisky (or whiskey, from the Gaelic *uisgebeatha*, 'water of life') is technically defined as an alcoholic liquor obtained by distillation of a fermented starchy compound, usually a grain (Funk and Wagnalls, 1950). Although distillation, including that of fermented grain, had been invented by the Chinese at least as far back as the 7th century AD, five centuries before it was introduced or rediscovered by Europeans (Needham, 1974; Temple, 1986), it is in Scotland that whisky has known its full development in diversity. The Scotch whisky distilleries catalogued by Jackson (1989) amount to 109, for a total of over 300 whiskies sold as single malts, which excludes the innumerable blended whiskies made of assemblages of liquors of different qualities or brands.

It is interesting that Scotland alone has developed such a diversity of whiskies, matured and sold as single malts. Jackson (1989) mentions only four other pure malts in the world, one from Ireland and three from Japan. Single malts are well known by amateurs to differ widely in nose, colour, body, palate and finish. The layman interested in discovering the diversity of these tasting sensations may wonder how to approach the problem: what are the main types of single-malt Scotches, and in what way do they differ? This is the type of question that came to us after acquainting

† *Address for correspondence:* Département de Sciences Biologiques, Université de Montréal, CP 6128, Succursale A, Montréal, Québec, H3C 3J7, Canada.  
E-mail: LAPOINF@ERE.UMONTREAL.CA

ourselves with single-malt whiskies during and after the 3rd Conference of the International Federation of Classification Societies held at Heriot-Watt University in Edinburgh, Scotland, in August 1991.

Jackson (1989) produced a connoisseur's guide to the malt whiskies of Scotland. This guide contains a description of single malts from each of the 109 distilleries of Scotland. We decided to use that information to produce a classification of single-malt whiskies, to answer the following questions.

- (a) What are the major types of single-malt whiskies that can be recognized? What are their chief characteristics and the best representatives?
- (b) What is the geographic component in that classification?
- (c) Do the various categories of characteristics—nose, colour, body, palate and finish—lead to the same classification?

These problems raised several methodological questions. How do we obtain a quantitative assessment of similarity, from the literary descriptions of nose and taste written by connoisseurs? And how do we compare the nose, colour, body, palate and finish data sets with one another? Several published procedures are available for doing so, such as comparing raw data sets (canonical analysis), derived distance matrices (matrix correlations) or dendrograms (consensus measures). Do they all lead to the same conclusions? To our knowledge, all three families of methods have never been compared on the same data sets. So, the present paper may be interesting to statisticians for its methodological component as well as to Scotch fans for its conclusions.

## 2. Characteristics of Pure Malt Scotch Whiskies

Our analyses are based on pure malt Scotch whiskies only, i.e. a liquor obtained from fermented (malted) barley, produced in a pot still in a single distillery located in Scotland and aged in oak casks for at least three years (Jackson, 1989). Pure malt whiskies not made in Scotland (e.g. Bushmills Malt from Ireland) as well as blended Scotch whiskies (e.g. Johnnie Walker) were therefore not considered in this study.

We selected 109 different pure malt Scotch whiskies, one from each and every distillery mentioned by Jackson (1989). On each of these whiskies, the following variables were then selected, and coded numerically when necessary.

### 2.1. *Geographic Co-ordinates of Distillery*

Using the map and the geographic information provided by Jackson (1989), each distillery was located on a map of Scotland, from which the longitude and latitude were read on an arbitrary set of Cartesian axes. These variables were treated as  $(X, Y)$  geographic co-ordinates. A geographic distance matrix among distilleries was computed from these co-ordinates to test the hypothesis that distilleries that are close to one another produce whiskies of similar characteristics.

### 2.2. *Region*

The multistate variable 'region' describes the location of the distilleries following the division of Scotland into the three Scotch-producing regions and 13 districts proposed in Jackson (1989). The Highlands region is divided into northern, southern, western and eastern Highlands, Speyside, Midlands and the Islands districts (including

Mull, Skye, Jura and Orkney). The Lowlands are subdivided into the central, west, east, Borders and Campbeltown districts. The island of Islay is coded as a single region in this study; it includes the north shore, south shore and Loch Indaal. We used this variable to establish whether there are any genuine differences in Scotch characteristics between regions and districts. Regions reflect such macrogeographic variables as proximity to the sea, altitude, quality of soil and water, microclimate, etc.

### 2.3. *Distillery Score*

Each distillery has been awarded a score from 1 (poor) to 5 (best) by Jackson (1989), indicating the overall quality of its products. This variable was used only to compute the average distillery rating for each group in Appendixes A and B.

### 2.4. *Scotch Score*

Jackson awarded scores to Scotches from 55 (poor) to 95 (best), on a scale from 0 to 100, describing his judgment of their quality. This variable was used only as an indicator of Scotch quality in Appendixes A and B.

### 2.5. *Age*

When more than one pure malt was produced by any given distillery and described in Jackson (1989), the Scotch that was 10 years of age, or closest to that age, was selected. This variable was not used in the analyses proper.

### 2.6. *Scotch Characteristics*

All descriptions of colour, smell, taste and appearance are in textual form in Jackson (1989). Two examples follow:

'Appetizing aroma of peat *smoke*, almost incense-like, heather honey with a *fruity* softness';  
'Pronounced fragrance, with a hint of *smoky* dryness, some almond nuttiness, *fruit* and plenty of sweet, crushed-barley maltiness'.

Notice that some of the variables are found in both descriptions (smoky, fruity), whereas others are not. Our first challenge was to transform this type of information into something that could be analysed numerically. We proceeded as follows to recode all available information into binary form. First, all adjectives or substantives used by Jackson to describe his whiskies were catalogued and the frequency of each term computed. All characteristics with more than five citations were retained, and the whiskies were coded as having each of these characteristics (1) or not (0). The 68 variables used in our analysis are the following, divided according to Jackson's five feature types:

- (a) *colour*—white wine, yellow, very pale, pale, pale gold, gold, old gold, full gold, bronze, pale amber, amber, full amber, red, fino sherry;
- (b) *nose*—aromatic, peaty, sweet, light, fresh, dry, fruity, grassy, salty, sherry, spicy, rich;
- (c) *body*—soft, medium, full, round, smooth, light, firm, oily;
- (d) *palate*—full, dry, sherry, big, light, smooth, clean, fruity, grassy, smoky, sweet, spicy, oily, salty, aromatic;
- (e) *finish*—full, dry, warm, big, light, smooth, clean, fruity, grassy, smoky, sweet, spicy, oily, salty, aromatic, quick, long, very long, lingering.

So, the resulting data matrix concerned 109 Scotch whiskies and 68 taster's characteristics, plus two geographic co-ordinates, one multistate region variable and two scores to be used for *a posteriori* assessment of our classification.

### 3. Classification of Pure Malt Scotch Whiskies

Two steps are required to derive a classification from the 68 characteristics measured on the 109 Scotches. First, a distance matrix must be computed to express proximities among Scotches. Then, this matrix must be fed into a clustering algorithm to obtain a hierarchical classification. The methods used in this section are described in all general texts on classification methods, such as the classical book of Sneath and Sokal (1973).

Because of the way that we coded the information, we are justified in counting the joint possession of a characteristic (code 1) as an indication of similarity between two whiskies, while the joint absence of a characteristic (code 0) can be considered neither as an indication of similarity nor an indication of dissimilarity. So the similarity between all pairs of Scotches was computed by using Jaccard's (1901) coefficient of similarity ( $S$ ), which is the number of characteristics possessed (code 1) by both whiskies under comparison divided by the number of characteristics present (code 1) in either whisky. This coefficient assumes that the database is internally consistent and that the use of a descriptor refers to the same organoleptic property in all whiskies where it is found. Under this assumption, the coefficient counts as a difference (dissimilarity), the presence (code 1) of a characteristic in a whisky and its absence in another (code 0).

When computing the similarities, each of the five feature types was given the same importance. To achieve this, the characteristics were weighted by the inverse of the number of characteristics in their type. The sum of the weighted characteristic similarities was divided by the sum of the weights to rescale the coefficients between 0 (no characteristic in common) and 1 (complete similarity in organoleptic description profiles). The one-complement of this similarity measure is a distance ( $D = 1 - S$ ) bounded between 0 (maximum similarity) and 1 (maximum distance).

A hierarchical classification presented in the form of a dendrogram was derived from the distance matrix by using Ward's (1963) minimum variance clustering method. At each clustering step, this method merges the whiskies or clusters of whiskies that minimize the sum of squared distances to the cluster centroids, so that it tends to form virtually circular compact clusters in a space where individual whiskies would be represented as points. (Other hierarchical clustering methods such as unweighted arithmetic average clustering, also called UPGMA (Sneath and Sokal, 1973), or proportional link linkage with 50% connectedness, which are also known not to distort the metric properties of this reference space, were applied to the similarity matrix and produced almost exactly the same classification. So, our results are largely insensitive to the choice of the clustering method.)

A cophenetic matrix among objects (Sokal and Rohlf, 1962) was also computed from this dendrogram; its use will be discussed later. In that matrix, the distance between two objects is equal to the value of the fusion level where these two objects become members of the same group in the dendrogram. The cophenetic matrix generated by minimum variance hierarchical clustering is ultrametric, i.e. it has the property that  $d_{ij} \leq \max(d_{ik}, d_{jk})$  for all triplets of objects ( $i, j, k$ ) in the matrix (Hartigan, 1967; Johnson, 1967). This means in practice that the classification can

be represented by a dendrogram without reversals (which can result from the centroid clustering methods for instance).

The classification of Scotches obtained by Ward's method is presented in Fig. 1. It depicts large clusters characterizing loose associations of Scotches and smaller clusters that define more closely related products. Depending on the fusion level of clustering, we could for instance define two, three, six or 12 subgroups of Scotches (Fig. 1, top scale). The choice of a partition level then depends on the explanatory power that these clusters have. The first dichotomy divides the 109 distilleries into two subsets of 69 and 40 Scotches respectively, separating full-gold-coloured, dry-bodied and smoky whiskies from amber, aromatic, light-bodied, smooth palate and fruity finish liquors. The larger cluster is again subdivided into two groups of 38 and 31 Scotches, mainly separating amber, full-bodied, fruity, oily and spicy malts from gold-coloured, smooth-bodied, light and grassy whiskies. As we go down the distance scale, smaller and tighter clusters become recognizable. At fusion level 0.74, six groups are distinguished, whereas 12 groups can be recognized at distance level 0.63. The composition of each of these 12 groups is detailed in Appendix A, where the characteristics found in a majority of Scotches in each cluster are presented. Appendix A may be useful to amateurs who want to maximize the organoleptic diversity of their liquor cabinet, or who are looking for particular characteristics.

#### 4. Geographic Interpretation of Scotch Classification

The classification presented in Fig. 1 was derived from the taster's descriptions but may none-the-less bear a significant geographic component characterizing whiskies from different regions of Scotland, since whiskies are affected not only by distillery traditions but also by such regional-dependent factors as water, soil, microclimate, temperature and even air quality (Jackson, 1989).

The effect of geography on the Scotch classification was first assessed by comparing a regional model, represented by a binary matrix, with the cophenetic matrix corresponding to the Scotch dendrogram. In this regional model, a pair of Scotches was coded by distance '0' if both are produced in the same region (referring to the three regions: Highlands, Lowlands and Islay) and by '1' otherwise. Geographic coordinates of the 109 distilleries were also used to build a geographic distance matrix among distilleries that could be compared with the Scotch classification, to test the hypothesis that distilleries that are close to one another produce whiskies with similar organoleptic properties. Finally, both the regional and the geographic models were compared with binary matrices representing partitions of the classification in two, three, six or 12 groups, to evaluate the effect of geography at different levels of the dendrogram; in these model matrices, two whiskies that belong to the same group receive a distance of 0 or 1 if they are found in different groups. Two distinct statistical methods were chosen to test these relations.

The Mantel (1967) test of matrix comparison was first used to evaluate whether the geographic distances are correlated with the cophenetic distances of the dendrogram or with the classification of whiskies in two, three, six or 12 groups; the same test was used to compare the geographic region model with the cophenetic and group model matrices. A measure of resemblance (we used the matrix correlation coefficient, or standardized Mantel statistic) is computed between the values in the two upper (or lower) triangular parts of the square symmetric distance matrices under comparison,

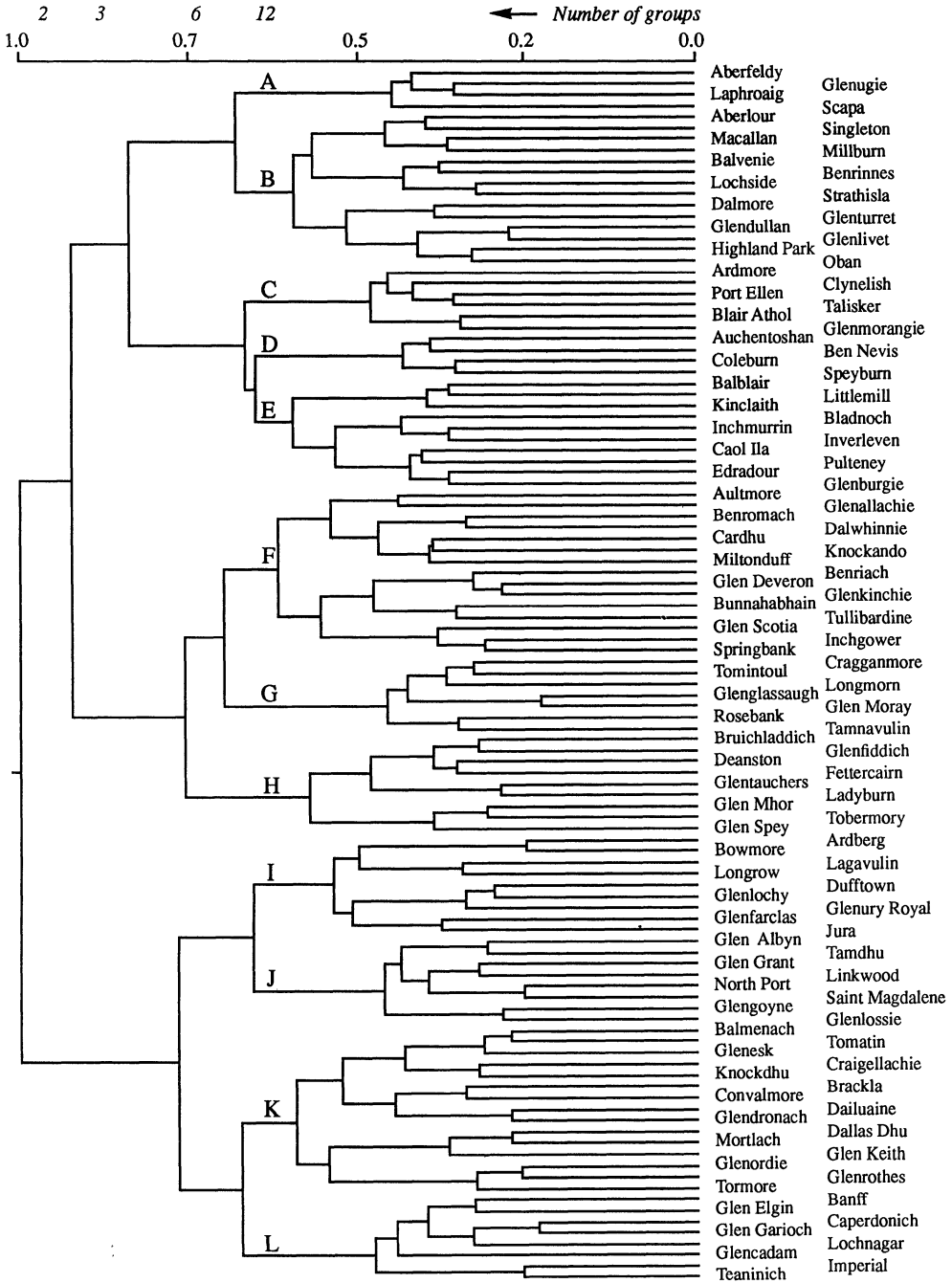


Fig. 1. Dendrogram representing the minimum variance hierarchical classification of single-malt Scotch whiskies: two scales are provided at the top of the graph—the number of groups formed by cutting the dendrogram vertically at the given points and the fusion distances of the hierarchical classification (represented by vertical segments in the dendrogram); the vertical order of the whiskies is partly arbitrary—swapping the branches of a dendrogram does not change the corresponding cophenetic matrix (the 12 groups detailed in Appendix A are labelled A–L here)

and tested against a random distribution obtained by repeatedly permuting at random the rows and corresponding columns of one of the matrices, and recomputing the statistic. The null hypothesis being tested is that the two matrices under study are no more similar than they would be by chance assignment of the labels (whisky names) to the rows and corresponding columns. Mantel statistics may turn out significant even if they have values that would seem very low for a Pearson product-moment correlation between two variables.

The cophenetic matrix representing the dendrogram was also statistically compared with the geographic and regional model matrices, by using the double-permutation test (DPT) proposed by Lapointe and Legendre (1990, 1991, 1992). This method has been developed to compare either two dendrograms obtained independently for the same objects or a dendrogram to some other independently obtained distance matrix of the same objects; dendrograms are represented by cophenetic matrices in the computations. As in the Mantel test, a measure of resemblance is first computed between the values in the two upper (or lower) triangular parts of the square symmetric distance matrices under comparison; as our test statistic, we chose again the matrix correlation coefficient, or standardized Mantel statistic, which is commonly used for such a purpose (Sokal and Rohlf, 1962; Lapointe and Legendre, 1992). In the DPT, the cophenetic matrix is subjected to a double permutation independently involving the fusion levels and the positions of the objects; when two dendrograms are being compared (as in Section 6), both are independently permuted. The measure of resemblance is recomputed during each cycle to construct a null distribution of the reference statistic. The null hypothesis being tested by the DPT is that the two dendrograms under comparison (or the dendrogram and the distance matrix) are no more similar than randomly generated dendrograms with the same number of objects, random topology and random labels would be. The DPT is conceptually the best way to compare dendrograms because the reference distribution of the statistic is obtained from repeated randomizations involving all components of the dendrograms (Lapointe, 1990). The Mantel test and the DPT may produce different probabilities, the difference lying only in the topology of the dendrogram, which is taken into account in the DPT randomization procedure, but not in the Mantel test; the statistic used is the same, though.

The results of all comparison tests between the Scotch classification and the geographic distance matrix are presented in Table 1. Those between the classification and the regional model are summarized in Table 2. Mantel tests were computed in

TABLE 1  
*Results of the comparison tests between the cophenetic matrix representing the Scotch classification and the geographic distance matrix†*

<i>Test</i>	<i>Dendrogram</i>	<i>2 groups</i>	<i>3 groups</i>	<i>6 groups</i>	<i>12 groups</i>
Mantel	0.031‡	-0.027§	0.007§	0.065‡	0.064‡
DPT	0.031§§				

†Mantel: standardized Mantel statistic and associated probabilities; dendrogram refers to the cophenetic matrix representing the whole classification (Fig. 1); groups are defined in Fig. 1 and in the text.

‡ $p \leq 0.001$ .

§ $p > 0.05$ .

§§ $0.001 < p \leq 0.01$ .



TABLE 2

*Results of the comparison tests between the cophenetic matrix representing the Scotch classification and the three-region regional model†*

<i>Test</i>	<i>Dendrogram</i>	<i>2 groups</i>	<i>3 groups</i>	<i>6 groups</i>	<i>12 groups</i>
Mantel	0.042‡	-0.025§	0.019§	0.066‡	0.076‡
DPT	0.042§§				

†The terms are as defined in Table 1.

‡ $p \leq 0.001$ .

§ $p > 0.05$ .

§§ $0.001 < p \leq 0.01$ .

all cases, whereas the DPT was only used when dealing with the cophenetic matrix corresponding to the Scotch dendrogram; these two tests would have produced identical results when computed on binary distance matrices, for which there is no topology.

We observe from the results of these tests that the cophenetic matrix is significantly correlated with both the regional model and the geographic distances. This is true with the Mantel test as well as with the DPT. Likewise, the regional model and geographic distances also produced identical results when compared with different partitions of the Scotch dendrogram. Although the partitions in two or three groups are not correlated with geography or with regions, we obtained significant results when dealing with six or 12 groups. Geographic information therefore seems more relevant to define smaller clusters of closely related Scotches, rather than large groups containing many single malts that share few characteristics and come from several regions of Scotland.

## 5. Regional Classification of Pure Malt Scotch Whiskies

Given the significant effect of geography on Scotch characteristics, a second classification was computed by using geographic information as a constraint to obtain regional clusters of Scotch whiskies. A clustering with spatial contiguity constraint (Legendre, 1987) was computed from the 68 Scotch characteristics by using a spatially constrained version of the  $k$ -means method (MacQueen, 1967; Anderberg, 1973). The purpose was to divide Scotland into optimally defined organoleptic regions, in which the distillery products satisfy the criterion of organoleptic homogeneity (minimum variance, computed from the 68 variables) *and* form a contiguous subdivision of Scotland. The  $k$ -means method is not a hierarchical clustering method, but rather a partitioning algorithm that produces groups minimizing the sum of within-group variances or, which is equivalent, the sum of squared distances to each cluster's centroid. So the criterion being minimized is the same as in Ward's (1963) hierarchical minimum variance method used above. The user decides how many groups  $k$  she or he wants to obtain from the program. In its spatially constrained form, objects (distilleries) are allowed to cluster only if they form a continuous area on the map, as defined by a list of links indicating which pairs of objects are neighbours on the map; this list of neighbourhood links was established by using the Delaunay triangulation method, from the list of  $(X, Y)$  geographic co-ordinates of the distilleries. So, in this analysis, even if two Scotches happened to have exactly the same

organoleptic characteristics, the spatial contiguity constraint would prevent them from clustering if they are produced in different parts of Scotland.

The resulting classification, presented in Appendix B, describes 11 groups of Scotches. They are also illustrated in Figs 2 and 3, which show that the groups formed by spatially constrained  $k$ -means clustering are compact and non-overlapping in space. We tried several values for the parameter  $k$ , and 11 is the largest number of groups that did not result in the formation of singleton clusters, which we felt were undesirable in a synoptic classification (Williams *et al.*, 1971). This regional classification should not be used as a reference guide for choosing organoleptically different Scotches, but rather as a tool for Scotch fans to select distinct single malts coming from different regions of Scotland.

## 6. Comparison of Feature Type Classifications

The purpose of this section is to answer our third question and to find out whether the various feature types are congruent in the classifications that they lead to. For each type—colour (14 features), nose (12 features), body (8 features), palate (15 features) and finish (19 features)—we computed a new distance matrix over the 109 single-malt Scotches and derived a dendrogram and the corresponding cophenetic matrix. Then we compared all ( $5 \times 4/2 = 10$ ) pairs of raw data matrices, distance matrices and cophenetic matrices to assess the congruence of the five features.

Comparing raw data matrices related to the same objects (here, the whiskies) is usually done by canonical correlation analysis. The difficulty with the data for the present study lies in their asymmetry, i.e. the fact that 0s do not have the same meaning as 1s and double 0s should somehow be excluded from the comparisons; this was the reason for using the Jaccard coefficient in a previous section, instead of one of the symmetrical resemblance coefficients. The problem of 0s was dealt with, in the present analyses, by first running the five data sets through correspondence analysis, a multivariate ordination method which is explained in depth in many classical textbooks (e.g. Nishisato (1980), van Rijkevorsel and de Leeuw (1988) and ter Braak (1987)). The characteristic of interest to us here is that correspondence analysis preserves the  $\chi^2$ -distance among objects, which, like the Jaccard coefficient, is an asymmetrical resemblance coefficient (Legendre and Legendre, 1983). The resulting ordination axes were then used as the new data tables among which canonical correlations were computed; the redundancy coefficients, which measure the proportion of variance of each of the matrices which is accounted for by the canonical correlations, as well as the permutational significance of the redundancy coefficients, were recorded.

Distance matrices were compared by using Mantel tests, whereas the cophenetic matrices corresponding to dendrograms were compared through the DPT, as in Section 4. Separate tests were performed for each of the 10 pairwise comparisons involving the five feature types, and probabilities obtained under the randomization models were noted.

Computations of distance matrices, cluster analyses and Mantel tests were carried out using the 'R' package (Legendre and Vaudor, 1991). Comparisons of dendrograms were performed by using a program developed by Lapointe, whereas a program written by Legendre was used to compute the correspondence and canonical correlation analyses with permutational tests.

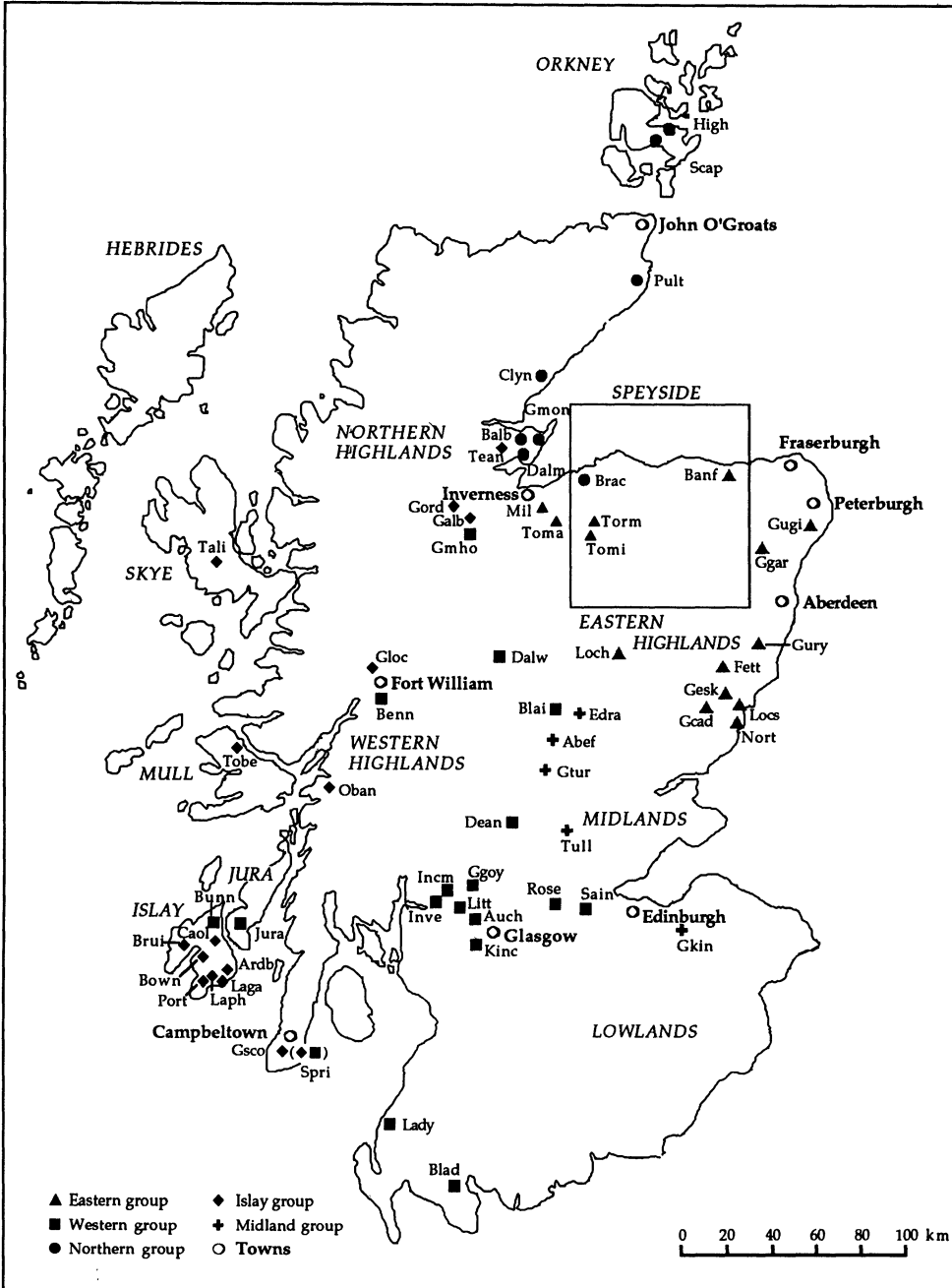


Fig. 2. Map of Scotland showing the positions of the Scotch distilleries, divided into 11 groups (symbols) in the regional classification of single-malt whiskies (Appendix B) (the six Speyside groups are deferred to Fig. 3): distillery names are represented by four-letter abbreviations (see Fig. 3); the names of regions and of some major cities are also indicated—notice that two Scotches in the present study come from the Springbank distillery; Springbank pertains to the western group whereas Longrow is a member of the Islay group

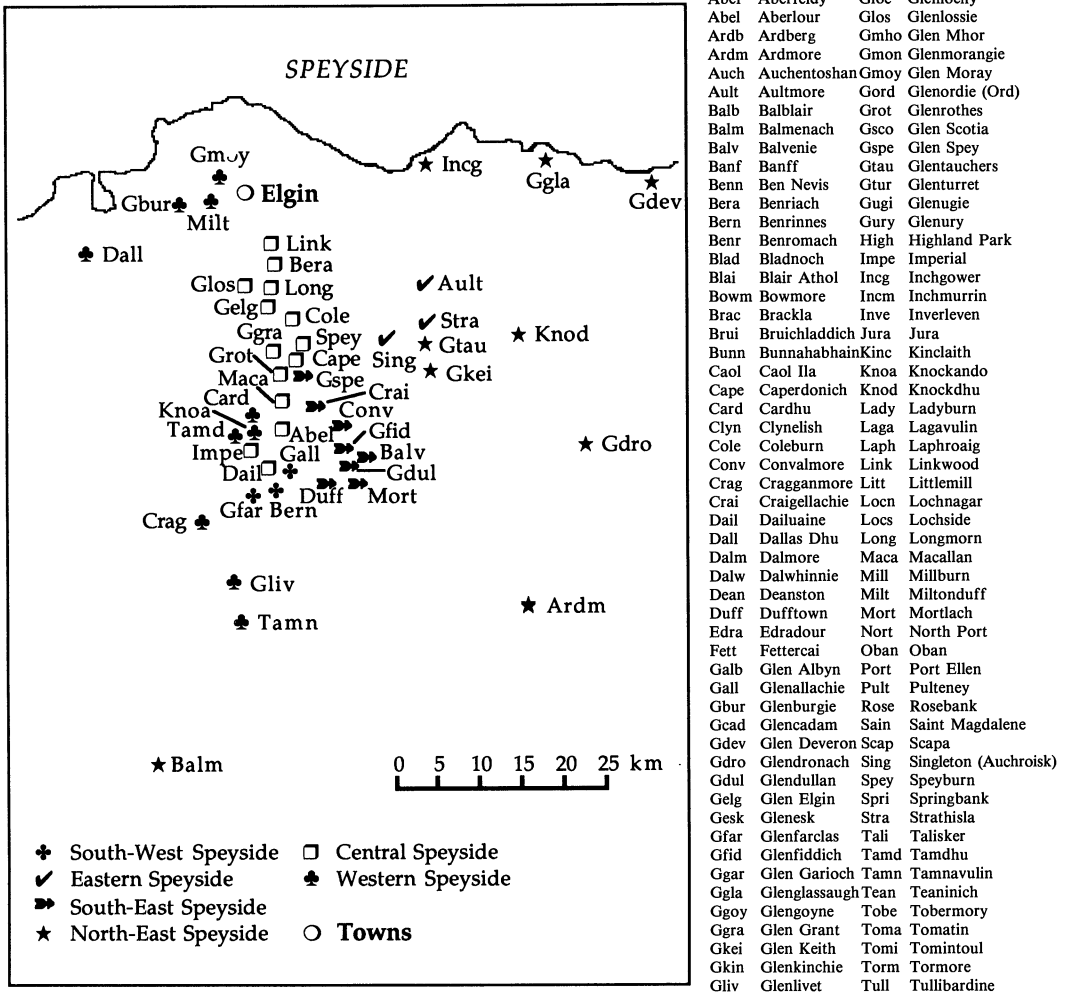


Fig. 3. Map of the Speyside region showing six of the 11 groups (symbols) of Scotch distilleries of the regional classification of single-malt whiskies (Appendix B) (the names of regions and of some major cities are also indicated) and abbreviations and full names of the distilleries

Results of all comparison tests from the raw data tables, distance matrices and dendrograms are presented in Tables 3, 4 and 5 respectively. The significant and non-significant results from these three tables are summarized, in Fig. 4, by graphs connecting the significantly related (i.e. the not statistically different) feature types. We can see from these results that the finish characteristics are always left unconnected and so differ from all other features, at all three comparison levels (Figs 4(a), 4(b) and 4(c)). Other type comparisons sometimes differ, depending on the test. Four tests are significant when comparing the raw data tables, five tests are significant with distance matrices and five other tests are significant when dealing with cophenetic matrices. Of all these significant results, only three comparisons are congruent at all levels: palate–nose; nose–colour; colour–body. The colour–palate relationship is

TABLE 3

*Canonical correlation results of the comparisons among the five feature types, based on the raw data tables†*

	<i>Nose</i>	<i>Body</i>	<i>Palate</i>	<i>Finish</i>
Colour	0.147‡	0.121‡	0.151§	0.152§
Nose		0.121‡	0.164§§	0.153§
Body			0.105§	0.116§
Palate				0.129§

†The figures are the averages of the *A:B* and *B:A* redundancy coefficients; the significance of the *A:B* (matrix 1:matrix 2) and the *B:A* redundancy coefficients was always the same.

‡ $0.001 < p \leq 0.01$ .

§ $p > 0.05$ .

§§ $p \leq 0.001$ .

TABLE 4

*Mantel test results of the comparisons among the five feature types, represented by distance matrices†*

	<i>Nose</i>	<i>Body</i>	<i>Palate</i>	<i>Finish</i>
Colour	0.032‡	0.067§	0.027§§	-0.011§§
Nose		0.042‡	0.074‡	0.009§§
Body			0.054‡	0.018§§
Palate				-0.012§§

†The figures are standardized Mantel statistics.

‡ $0.001 < p \leq 0.01$ .

§ $p \leq 0.001$ .

§§ $p > 0.05$ .

TABLE 5

*DPT results of the comparisons among the five feature types, based on the cophenetic matrices†*

	<i>Nose</i>	<i>Body</i>	<i>Palate</i>	<i>Finish</i>
Colour	0.048‡	0.064§	0.046‡	-0.025§§
Nose		0.021§§	0.071§	-0.001§§
Body			0.051§	0.008§§
Palate				0.002§§

†The figures are standardized Mantel statistics.

‡ $0.001 < p \leq 0.01$ .

§ $p \leq 0.001$ .

§§ $p > 0.05$ .

only significant when comparing cophenetic matrices, whereas the nose-body test was not found significant with dendrograms, even though it was significant with raw data tables and distance matrices. However, the body-palate tests were significant when dealing with raw data or distances, but not when comparing dendrograms. Differences in statistical methods and/or transformation of data from one level to

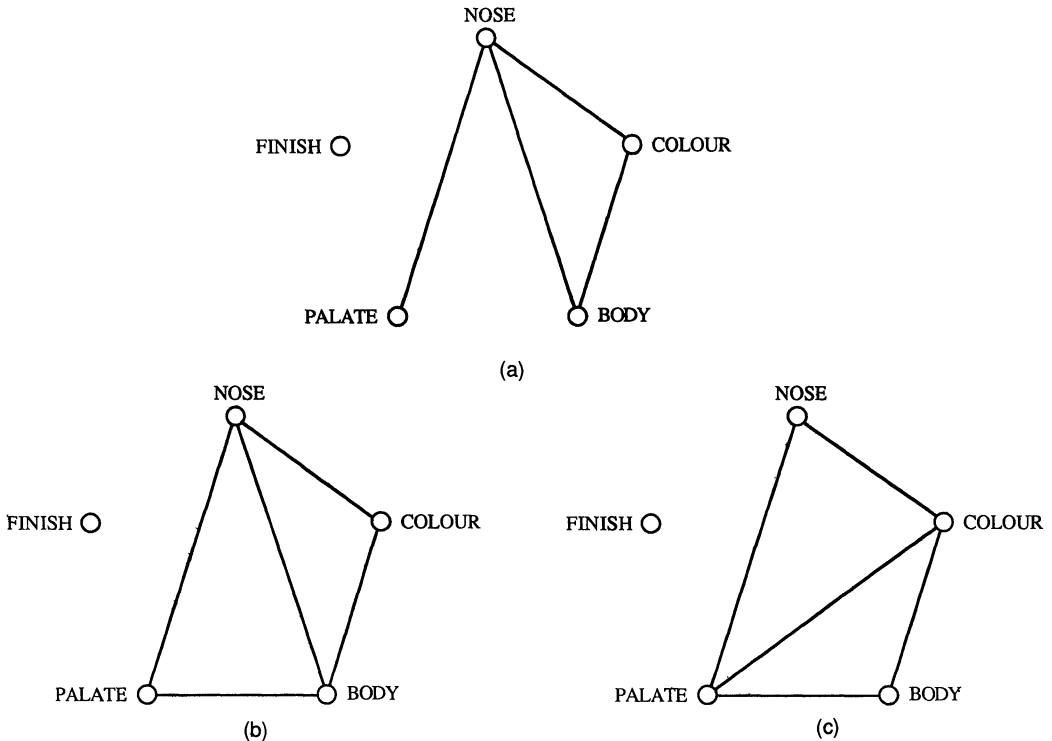


Fig. 4. Graphs presenting the results of all comparison tests among the five characteristic feature types from (a) the raw data tables, (b) distance matrices and (c) dendrograms: nodes connected by an edge correspond to feature types that are not statistically different; feature types that are statistically unrelated ( $p > 0.05$  in Tables 3–5) are left unconnected

another must be responsible for discrepancies in some of these results; some information is lost when moving from raw data to a distance matrix using one particular distance function, and information is lost again along the path from distances to dendrogram, at which point the ultrametric property is acquired.

## 7. Discussion

What makes a Scotch whisky so special? How do we distinguish single malts from one another? Why do Highlands Scotches taste different from Lowlands whiskies? We have addressed some of these questions, of interest to all Scotch lovers, by using a specific statistical framework. First, we were interested in deriving a classification of single-malt whiskies to characterize the various groups of Scotches. Then, this classification was correlated to geographic variables to assess the effect of the geographic features on Scotch taste. The relationship with geographic features was demonstrated, supporting the hypothesis that whiskies are affected not only by distillery secrets and traditions but also by factors dependent on region such as water, soil, microclimate, temperature and even air quality. This led us to produce a second classification where the map of Scotland is divided into regions that are relatively homogeneous in terms of the organoleptic properties of the Scotch whiskies that they

produce. Finally, the congruence of appearance, smell and taste characteristics was evaluated from our categorical variables. However, we were not only interested in answering these important epicurean questions; we also tried to address the following mathematical problems, discussed below. What are the effects of spatial constraints on clustering results? How does canonical correlation analysis relate to the Mantel test and the DPT? What is the agreement between raw data, distance matrix and dendrogram comparisons? In this paper, we were interested both in the methodology and in practical answers to our questions about Scotch whiskies.

Let us compare the two classifications that were constructed to depict the similarities between the 109 distillery products considered in this study. One is a simple hierarchical classification derived from the organoleptic features only, whereas the second is a space-constrained classification also based on organoleptic properties but emphasizing geographic proximities. The clusters in each of these classifications (Appendixes A and B) may differ depending on whether or not we consider spatial information in the clustering process. Indeed, a simple examination of these classifications shows that all groups are very different. The contingency table (Table 6) comparing the unconstrained and constrained classifications illustrates the poor agreement among the clusters obtained in the two cases. It implies that the spatial constraint modifies the classification structure based on organoleptic features: single malts produced in different regions may be as similar as or more similar than Scotches from neighbouring distilleries. Differences in distillation processes, the shape of the still, types of barley or yeast and maturation (duration, pretreatment of the casks, etc.), may affect Scotch characteristics in addition to geographic information. Whereas Appendix A defines groups of Scotches with similar organoleptic characteristics, Appendix B describes contiguous regions of Scotland where single malts possess similar organoleptic features.

The next step consisted of assessing whether different characteristic features were congruent in the classification that they led to, i.e. to evaluate whether colour, nose, body, palate and finish classifications could be used to compare Scotches in the same

TABLE 6

*Contingency table comparing the unconstrained (Appendix A, 12 columns) and spatially constrained (Appendix B, 11 rows) classifications of the 109 Scotches†*

	A.1	A.2	A.3	A.4	A.5	A.6	A.7	A.8	A.9	A.10	A.11	A.12	$\Sigma$
B.1	1	2					1	1	1	1	3	4	14
B.2			1	2	5	3	1	3	1	2			18
B.3	1	2	2		2	1					1		9
B.4	1	1	2		1	1		2	5	1	1	1	16
B.5	1	1			1	2							5
B.6		1				1			1				3
B.7		2				1							3
B.8		2						2	1		3		8
B.9			1			2	1	1			4		9
B.10		2		2		1	1			3	2	3	14
B.11		1			1	3	3	0		1	1		10
$\Sigma$	4	14	6	4	10	15	7	9	9	8	15	8	109

†The frequencies in the table are the numbers of whiskies in the group (column) of the first classification that are also in the group (row) of the second classification. The last column and row of the table present the row and column totals ( $\Sigma$ ).

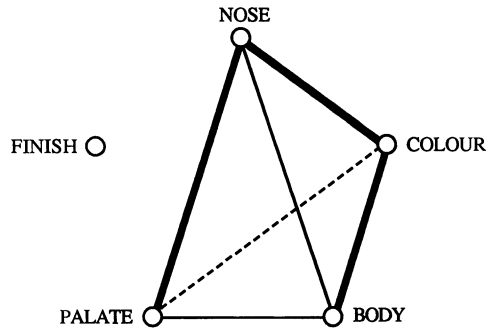


Fig. 5. Consensus graph summarizing the results presented in Fig. 4: bold edges depict congruent relationships at all three levels (raw data tables, distance matrices and dendrograms); standard edges are congruent at two levels, whereas a broken edge indicates a relationship which is significant at one level only; one of the feature types, finish, is left unconnected

way. Three different levels of comparison were considered: raw data matrices, distance matrices and dendrograms. We wanted to know whether statistical tests would produce different results depending on the comparison levels, i.e. we were interested in the following methodological question: are distance methods related to raw data comparison tests? Fig. 5 presents a consensus graph summarizing the three levels of results of Fig. 4. It appears from this graph that three out of six of the significant comparison tests were congruent for all three levels; two more tests were congruent for two levels, whereas a single comparison was significant at one level only. Considering also the not significant outcomes involving the finish characteristics, we obtain a strict consensus in which 70% of all tests are statistically identical at all three levels, leaving only three discrepancies among the 30 tests performed.

It is difficult to understand why the body–palate relationship appeared after transforming the raw data tables into distance matrices. The other two discrepancies between dendrogram and distance matrix comparisons can be explained, however. Two distinct causes may be invoked: the permutation test or the clustering method. The permutational method is not the same when comparing dendrograms and when dealing with distance matrices, as explained in a previous section. Contrary to the Mantel test, the DPT generates random dendrograms with random topologies, in addition to the permuted positions of the objects. The null hypotheses of these tests thus differ, as described earlier.

However, given that most of the tests computed with dendrograms are congruent with those performed on distance matrices, we have to look at some other explanation to interpret the discrepancies in the colour–palate and the nose–body comparisons. The problem seems to be with the clustering method, which is known to transform the initial distance values into ultrametric values, with a loss of information in the process. If the ultrametric representation in the form of a dendrogram is very remote from actual distances, comparison tests involving ultrametric (cophenetic) matrices may differ from distance matrix comparisons. Therefore, we looked at the stress values depicting the remoteness between the initial distances and the ultrametric distances recovered by the classifications, for all five feature types. This value is computed as the sum of the squared differences between the values in the cophenetic similarity matrix and in the original similarity matrix. Interestingly, we found that the stress



values for the palate-colour ( $d = 232.3$ ) and nose-body ( $d = 248.1$ ) comparisons were the two most remote pairs from their initial distance matrices, among all 10 pairwise comparisons (average, 168.17; standard deviation, 56.82). The fact that the body-nose comparison was not significant only when comparing dendrograms and that the colour-palate comparison was the only significant test that was not congruent with the raw data or the distance matrix are consequences of the ultrametric transformation that is necessary to obtain a representation of distances in the form of a dendrogram.

What we found, nevertheless, is that most features lead to congruent results, except for the finish characteristic that always differs from the other features. The other four categories (colour, body, nose and palate) tell virtually the same story; a taster could use any of these feature types to define, compare and recognize single malts. However, we believe that finish is an important feature, if not the most important for whiskies. In the light of such results, we may thus question the method used by the taster to characterize single malts. Were the whiskies swallowed or spitted? The second alternative would imply that the finish characteristics could not be fully detected. Only when swallowing an alcoholic product could one totally capture the aftertaste. So, is the strange case of the finish classification a consequence of the taster's method, or is it really a different feature that leads to clusters of Scotches that are not obtained with other categorical features? Jackson explains that 'some professional blenders work only with their nose, not finding it necessary to let the whisky pass their lips' (Jackson (1989), p. 230); this could be read as an indication that smell is the most important feature to distinguish single malts. However, we are led to believe from our analyses that finish should be equally weighted as a selection criterion. In any case, single malts must be swallowed.

### Acknowledgement

We are grateful to Isabelle Lachance who, through patience and skill, drew the distillery maps (Figs 2 and 3).

### Appendix A: Classification of Scotch Whiskies

The 12 groups resulting from Ward's (1963) minimum variance clustering of Scotches are described as follows. Group names (A-L) are as in Fig. 1.

*Scotches*: a list of the whiskies in the group

*Average characteristics*: the dominant characteristics for members of this group; the colour, nose, body, palate and finish characteristics are separated by semicolons

*Distilleries and Scotches average rating*: average scores given by Jackson (1989) to these distilleries (from 1 to 5) and Scotches (from 0 to 100), separated by a semicolon

*The best of its class*, according to Jackson (1989), among those selected for our study: Scotch name, (region), age and score for that Scotch

#### A.1. Group A

*Scotches*: Aberfeldy, Glenugie, Laphroaig, Scapa

*Average characteristics*: full gold; fruity, salty; medium; oily, salty, sherry; dry

*Distillery and Scotch average rating*: 2.75; 75.3

*The best of its class*: Laphroaig (Islay), 10 years, 86

### A.2. Group B

*Scotches:* Aberlour, Balvenie, Benrinnes, Dalmore, Glendullan, Glenlivet, Glenturret, Highland Park, Lochside, Macallan, Millburn, Oban, Singleton of Auchroisk, Strathisla

*Average characteristics:* amber; sweet, sherry; medium, smooth; dry, sweet; long

*Distillery and Scotch average rating:* 3.86; 80.0

*The best of its class:* Highland Park (Orkneys), 12 years, 90

### A.3. Group C

*Scotches:* Ardmore, Blair Athol, Clynelish, Glenmorangie, Port Ellen, Talisker

*Average characteristics:* pale gold; sweet, salty; medium to full, oily; sweet, spicy; big, long, spicy

*Distillery and Scotch average rating:* 3.50; 78.5

*The best of its class:* Talisker (Skye), 10 years, 90

### A.4. Group D

*Scotches:* Auchentoshan, Ben Nevis, Coleburn, Speyburn

*Average characteristics:* pale gold; fruity, grassy; oily; sweet, fruity; sweet, quick

*Distillery and Scotch average rating:* 2.75; 69.5

*The best of its class:* Auchentoshan (Lowlands), 10 years, 85

### A.5. Group E

*Scotches:* Balblair, Bladnoch, Caol Ila, Edradour, Glenburgie, Inchmurrin, Inverleven, Kinclaith, Littlemill, Pulteney

*Average characteristics:* pale wyne, gold; fruity, peaty; light; sweet, spicy; fruity

*Distillery and Scotch average rating:* 3.00; 74.8

*The best of its class:* Bladnoch (Lowlands), 8 years, 85

### A.6. Group F

*Scotches:* Aultmore, Benriach, Benromach, Bunnahabhain, Cardhu, Dalwhinnie, Glenallachie, Glen Deveron, Glenkinchie, Glen Scotia, Inchgower, Knockando, Miltouduff, Springbank, Tullibardine

*Average characteristics:* gold; aromatic; medium, smooth, light; sweet; sweet

*Distillery and Scotch average rating:* 3.27; 76.5

*The best of its class:* Springbank (Campbeltown), 15 years, 88

### A.7. Group G

*Scotches:* Cragganmore, Glenglassaugh, Glen Moray, Longmorn, Rosebank, Tamnavulin, Tomintoul

*Average characteristics:* gold, full gold; grassy; smooth, light; grassy; dry

*Distillery and Scotch average rating:* 3.43; 79.1

*The best of its class:* Cragganmore (Speyside), 12 years, 90

### A.8. Group H

*Scotches:* Bruichladdich, Deanston, Fettercairn, Glenfiddich, Glen Mhor, Glen Spey, Glentauchers, Ladyburn, Tobermory

*Average characteristics:* white wyne, pale; sweet; smooth, light; sweet, dry, fruity, smoky; dry, light

*Distillery and Scotch average rating:* 2.44; 69.2

*The best of its class:* Bruichladdich (Islay), 10 years, 76

**A.9. Group I**

*Scotches:* Ardbeg, Bowmore, Dufftown, Glenfarclas, Glenlochy, Glenury Royal, Jura, Lagavulin, Longrow (Springbank distillery)

*Average characteristics:* gold, full gold, bronze; dry, peaty; medium, light, firm; dry, smoky, sweet; salty

*Distillery and scotch average rating:* 3.67; 79.9

*The best of its class:* Longrow (Campbeltown), 14 years, 90

**A.10. Group J**

*Scotches:* Glen Albyn, Glengoyne, Glen Grant, Glenlossie, Linkwood, North Port, Saint Magdalene, Tamdhu

*Average characteristics:* full gold; dry, peaty, sherry; light to medium, round; sweet; dry

*Distillery and Scotch average rating:* 2.88; 72.8

*The best of its class:* Linkwood (Speyside), 12 years, 83

**A.11. Group K**

*Scotches:* Balmenach, Brackla, Convalmore, Craigellachie, Dailuaine, Dallas Dhu, Glendronach, Glenesk, Glen Keith, Glenordie (Glen Ord), Glenrothes, Knockdhu, Mortlach, Tomatin, Tormore

*Average characteristics:* gold, full gold; sweet, dry, peaty; medium to full; sweet, dry; dry, long

*Distillery and Scotch average rating:* 3.07; 74.0

*The best of its class:* Dallas Dhu (Speyside), 1971, 85

**A.12. Group L**

*Scotches:* Banff, Caperdonich, Glencadam, Glen Elgin, Glen Garioch, Imperial, Lochnagar, Teaninich

*Average characteristics:* full gold; aromatic, peaty; medium; sweet, smoky; smoky

*Distillery and Scotch average rating:* 3.00; 73.3

*The best of its class:* Lochnagar (eastern Highlands), 12 years, 80

**Appendix B: Regional Classification of Scotch Whiskies**

The 11 groups resulting from spatially constrained  $k$ -means clustering of Scotches are described as follows; see also Figs 2 and 3. The descriptions (below) are presented as in Appendix A.

**B.1. Eastern Group: Spey District and Eastern Highlands**

*Scotches:* Banff, Fettercairn, Glencadam, Glenesk, Glen Garioch, Glenugie, Glenury Royal, Lochnagar, Lochside, Millburn, North Port, Tomatin, Tomintoul, Tormore

*Average characteristics:* full gold; peaty and aromatic; medium and smooth; malty, sweet, dry and smoky; dry

*Distillery and Scotch average rating:* 2.86; 72.3

*The best of its class:* Lochnagar (eastern Highlands), 12 years, 80

**B.2. Western Group: Lowlands, Islands of Islay and Jura, and Western Highlands**

*Scotches:* Auchentoshan, Ben Nevis, Bladnoch, Blair Athol, Bunnahabhain, Dalwhinnie, Deanston, Glengoyne, Glen Mhor, Inchmurrin, Inverleven, Jura, Kinclaith, Ladyburn, Littlemill, Rosebank, Saint Magdalene, Springbank

*Average characteristics:* gold; dry and peaty; light; malty, sweet and fruity; dry  
*Distillery and Scotch average rating:* 2.89; 72.4  
*The best of its class:* Springbank (Campbeltown district), 15 years, 88

**B.3. Northern Group: Spey District, Orkney Islands and Northern Highlands**  
*Scotches:* Balblair, Benromach, Brackla, Clynelish, Dalmore, Glenmorangie, Highland Park, Pulteney, Scapa  
*Average characteristics:* amber; sweet, fresh, fruity and salty; medium; sweet, smooth, spicy, salty and fruity; fruity, dry and long  
*Distillery and Scotch average rating:* 3.56; 79.0  
*The best of its class:* Highland Park (Orkneys), 12 years, 90

**B.4. Islay Group: Islay, Mull and Skye Islands, Western Highlands, Campbeltown District**  
*Scotches:* Ardbeg, Bowmore, Bruichladdich, Caol Ila, Glen Albyn, Glenlochy, Glenordie (Glen Ord), Glen Scotia, Lagavulin, Laphroaig, Oban, Port Ellen, Longrow (Springbank distillery), Talisker, Teaninich, Tobermory  
*Average characteristics:* full gold; peaty and salty; light to medium and oily; dry, malty and smoky; big and spicy  
*Distillery and Scotch average rating:* 3.56; 79.0  
*The best of its class:* Lagavulin (Islay), 12 years, 89

**B.5. Midland Group: Midlands and Eastern Lowlands**  
*Scotches:* Aberfeldy, Edradour, Glenkinchie, Glenturret, Tullibardine  
*Average characteristics:* gold; sweet, aromatic, fruity and peaty; soft to medium, light but firm; clean and sweet; warm, fruity and spicy  
*Distillery and Scotch average rating:* 3.40; 75.8  
*The best of its class:* Edradour (Midlands), 10 years, 81

**B.6. South-west Speyside: West of Duff Town, East of Spey River**  
*Scotches:* Benrinnes, Glenallachie, Glenfarclas  
*Average characteristics:* bronze; aromatic, sweet and peaty; firm; dry; sweet, malty and long  
*Distillery and Scotch average rating:* 3.33; 80.0  
*The best of its class:* Glenfarclas (Speyside), 10 years, 86

**B.7. Eastern Speyside: North and West of Keith**  
*Scotches:* Aultmore, Singleton of Auchroisk, Strathisla  
*Average characteristics:* amber; sweet, fresh, fruity and salty; medium; malty, smooth, spicy, fruity and salty; dry, fruity and long  
*Distillery and Scotch average rating:* 3.56; 79.0  
*The best of its class:* Singleton (Speyside), 1975, 79

**B.8. South-east Speyside: around Dufftown**  
*Scotches:* Balvenie, Convalmore, Craigellachie, Dufftown, Glendullan, Glenfiddich, Glen Spey, Mortlach  
*Average characteristics:* full gold and amber; sweet, aromatic and peaty; medium to full; sweet, dry, fruity and smoky; dry, oily and lingering  
*Distillery and Scotch average rating:* 3.13; 75.1  
*The best of its class:* Balvenie (Speyside), no age statement, 85

**B.9. North-east Speyside: Mostly around Keith**

*Scotches:* Ardmore, Balmenach, Glen Deveron, Glendronach, Glenglassaugh, Glen Keith, Glentauchers, Inchgower, Knockdhu

*Average characteristics:* gold; malty, dry, grassy and peaty; medium and smooth; dry and sweet; dry and long

*Distillery and Scotch average rating:* 2.67; 71.6

*The best of its class:* Glenglassaugh (Speyside), no age statement, 76

**B.10. Central Speyside: around Rothes**

*Scotches:* Aberlour, Benriach, Caperdonich, Coleburn, Dailuaine, Glen Elgin, Glen Grant, Glenlossie, Glenrothes, Imperial, Linkwood, Longmorn, Macallan, Speyburn

*Average characteristics:* full gold; malty, sweet, fruity and sherry character; light to medium; dry, sweet and fruity; dry, smooth and smoky

*Distillery and Scotch average rating:* 3.43; 76.8

*The best of its class:* Macallan (Speyside), 10 years, 87

**B.11. Western Speyside: West of the Elgin, Rothes and Dufftown**

*Scotches:* Cardhu, Cragganmore, Dallas Dhu, Glenburgie, Glenlivet, Glen Moray, Knockando, Milnorduff, Tamdhu, Tamnavulin

*Average characteristics:* pale gold; aromatic, grassy and peaty; light but firm; sweet, grassy and fruity; sweet, dry, grassy, smoky and long

*Distillery and Scotch average rating:* 3.30; 77.8

*The best of its class:* Cragganmore (Speyside), 12 years, 90

**References**

- Anderberg, M. R. (1973) *Cluster Analysis for Applications*. New York: Academic Press.
- ter Braak, C. J. F. (1987) Ordination. In *Data Analysis in Community and Landscape Ecology* (eds R. H. G. Jongman, C. J. F. ter Braak and O. F. R. van Tongeren), pp. 91–173. Wageningen: Centre for Agricultural Publishing and Documentation.
- Funk and Wagnalls (1950) *New "Standard" Dictionary of the English Language*. New York: Funk & Wagnalls.
- Hartigan, J. A. (1967) Representation of similarity matrices by trees. *J. Am. Statist. Ass.*, **62**, 1140–1158.
- Jaccard, P. (1901) Etude comparative de la distribution florale dans une portion des Alpes et du Jura. *Bull. Soc. Vaud. Sci. Nat.*, **37**, 547–579.
- Jackson, M. (1989) *Michael Jackson's Malt Whisky Companion: a Connoisseur's Guide to the Malt Whiskies of Scotland*. London: Dorling Kindersley.
- Johnson, S. C. (1967) Hierarchical clustering schemes. *Psychometrika*, **32**, 241–254.
- Lapointe, F.-J. (1990) La comparaison statistique de classifications hiérarchiques: évaluation de trois tests de comparaison par permutations. In *Recueil des Textes des Présentations du Colloque sur les Méthodes et Domaines d'Application de la Statistique 1990*, pp. 255–266. Québec: Bureau de la Statistique du Québec.
- Lapointe, F.-J. and Legendre, P. (1990) A statistical framework to test the consensus of two nested classifications. *Syst. Zool.*, **39**, 1–13.
- (1991) The generation of random ultrametric matrices representing dendrograms. *J. Classif.*, **8**, 177–200.
- (1992) Statistical significance of the matrix correlation coefficient for comparing independent phylogenetic trees. *Syst. Biol.*, **41**, 378–384.
- Legendre, L. and Legendre, P. (1983) *Numerical Ecology*. Amsterdam: Elsevier.
- Legendre, P. (1987) Constrained clustering. *Nato Adv. Stud. Inst. Ser.*, **G14**, 289–307.
- (1990) Quantitative methods and biogeographic analysis. *Nato Adv. Stud. Inst. Ser.*, **G22**, 9–34.
- Legendre, P. and Vaudor, A. (1991) *The R Package: Multidimensional Analysis, Spatial Analysis*. Montréal: Université de Montréal.

- MacQueen, J. (1967) Some methods for classification and analysis of multivariate observations. In *Proc. 5th Berkeley Symp. Mathematical Statistics and Probability* (eds L. M. Le Cam and J. Neyman), vol. 1, pp. 281–297. Berkeley: University of California Press.
- Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res.*, **27**, 209–220.
- Needham, J. (1974) *Science and Civilisation in China*, vol. 5, *Chemistry and Chemical Technology: Part II, Spagyric Discovery and Invention: Magistries of Gold and Immortality*. Cambridge: Cambridge University Press.
- Nishisato, S. (1980) *Analysis of Categorical Data: Dual Scaling and Its Applications*. Toronto: University of Toronto Press.
- van Rijkevorsel, J. L. A. and de Leeuw, J. (1988) *Component and Correspondence Analysis*. Chichester: Wiley.
- Sneath, P. H. A. and Sokal, R. R. (1973) *Numerical Taxonomy—the Principles and Practice of Numerical Classification*. San Francisco: Freeman.
- Sokal, R. R. and Rohlf, F. J. (1962) The comparison of dendrograms by objective methods. *Taxon*, **11**, 33–40.
- Temple, R. K. G. (1986) *China, Land of Discovery*. London: Multimedia.
- Ward, Jr, J. H. (1963) Hierarchical grouping to optimize an objective function. *J. Am. Statist. Ass.*, **58**, 236–244.
- Williams, W. T., Lance, G. N., Dale, M. B. and Clifford, H. T. (1971) Controversy concerning the criteria for taxonomic strategies. *Comput. J.*, **14**, 162–165.