FOOD ADULTERATION

Class Definition and Mixture Class Definition by Means of Construction of Convex Hull Boundaries: Application to Analysis for Animal Fat Adulteration

ANNE THIELEMANS
Vrije Universiteit Brussel, Farmaceutisch Instituut, Laarbeeklaan 103, B-1090 Brussel, Belgium
HUBERT DE BRABANDER
University of Ghent, Veterinary Faculty, Laboratory of Chemical Analysis of Food from Animal Origin, Casinoplein 24, B-9000 Gent, Belgium
DESIRE L. MASSART
Vrije Universiteit Brussel, Farmaceutisch Instituut, Laarbeeklaan 103, B-1090 Brussel, Belgium

A visual classification technique based on the construction of convex hull boundaries in combination with a principal component analysis is described. This combined technique was evaluated in the situation in which a distinction has to be made between 2 pure animal fat classes and the corresponding mixture class. In the first instance, a principal component analysis is carried out to ensure the 2-dimensional and thus visual aspect of the technique. Convex hulls are then constructed in the 2-dimensional principal component plane to delimit the boundaries of the different classes to be defined. The effectiveness of the constructed hull boundaries in the definition of class-membership was investigated by means of the classification of different simulated test samples. The results show that, at least for the tested applications, the technique is valid, although some false positive classifications occur. The detection of outliers especially seemed to pose problems. Therefore, some propositions are made of how to refine the developed hull technique to enhance the classification results.

The estimation of food adulteration may often pose serious problems in food quality control. Entire or partial substitutions of the original product with an allied product of (mostly) inferior and cheaper quality are sometimes difficult to detect. A first necessity in trying to solve this substitution problem is to verify whether the 2 kinds or classes of products, i.e., the original products on the one hand and the adulterant on the other hand, can be distinguished in an efficient way and whether mixtures of these 2 products can also be recognized. The solution of this problem is 2-fold. In the first instance, one has to select and analyze those parameters that are supposed to discriminate the 2 possible classes. In a second step, it will be necessary to apply multivariate mathematical techniques to define class characteristics on the basis of the determined parameter. Those class characteristics in turn will allow us to define class-membership of unknown samples.

The scope of this article is restricted to the description and evaluation of a possible mathematical way of solving the "substitution" or "mixture" problem in the supposition that the necessary analytical measurements have already been carried out and discriminating data are available. The technique in question makes use of the construction of 2-dimensional convex hull boundaries for the delimitation of the different classes or, in other words, for the definition of class characteristics. In comparison to other techniques that might be used for solving the same problem, emphasis should be put on the graphical and thus visual aspect of this convex hull technique.

With supervised learning techniques such as, for instance, SIMCA (1) and UNEQ (2), it is possible to develop separate mathematical models for each of the classes to be defined (i.e., the 2 pure classes and the mixture classes) and to define class-membership by fitting each test object to each of the developed models. Linear discriminant analysis (LDA) (3) creates linear boundaries to distinguish different possible classes; objects are classified according to their position with respect to those boundaries. Factor analytical techniques such as target transformation factor analysis, which is used, among others, to assign the influence of different emission sources in a pollution pattern (4) or partial least squares (5) might be more precise. However, in all applications of multivariate mathematical techniques, one starts by representing the data visually, i.e., in 2 dimensions, using techniques such as principal components analysis. This is necessary because one likes to have a visual idea of the shape of the developed class-models, the position of the different classes in relation to each other and of test objects in relation to the different classes, the presence of outliers, the homogeneity of the classes, etc. We reasoned that since this technique is used anyway, it would be useful to enhance it by including the possibility of classifying pure samples and, more important, distinguishing mixture samples. The convex hull method seemed to us to be a valid alternative to reach this goal. It was tested on the detection and classification of adulteration of fat samples from different animal species.

Experimental

Data

A data set concerning the fatty acid composition of animal fat samples of different origin was taken as a starting point to evaluate the usefulness of the developed hull technique in the problem of class and mixture class definition. The original data set was made available by H. De Brabander. It lists the percent content of 7 fatty acids incorporated at the 2-position of triacylglycerol of 21 pork fat samples, 20 hen fat samples, 15 beef fat samples, and 14 horse fat samples. The fatty acids taken into consideration are myristic acid (C14:0), palmitic acid (C16:0), palmitoleic acid (C16:1), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2), and linolenic acid (C18:3).

The analysis procedure is briefly described as follows: Triacylglycerols are extracted and isolated from meat tissue samples, or isolated from fat tissue samples by homogenization, melting, and filtering. The fatty acid composition in position 2 of the triacylglycerols is determined by *(1)* the reaction of pancreatic lipase on the triacylglycerols, *(2)* the separation of the reaction products by thin-layer chromatography, *(3)* the transesterification of the resulting 2-monoacylglycerols with sodium methylate, and *(4)* the quantitative gas chromatographic analysis of the resulting fatty acids. A more detailed description of the materials and methodology used for this fatty acid analysis can be found in ref. 6.

It is known that the fatty acid pattern and the relative

Received June 15, 1987. Accepted August 4, 1988.

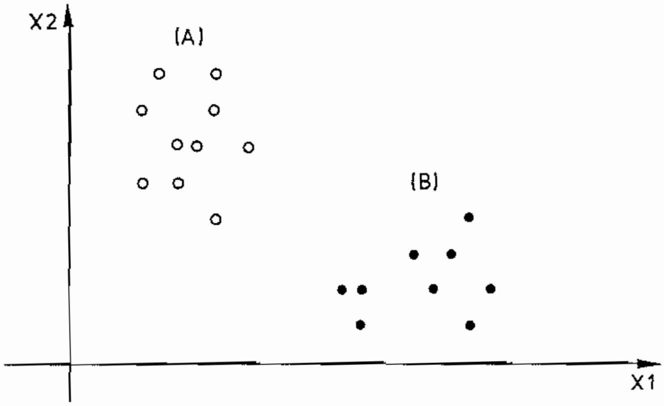

Figure 1a. Representation of objects of 2 different classes, A and B, in 2-dimensional plane defined by parameters X1 and X2.

distribution of fatty acids within the triacylglycerols is to some extent species-specific. However, because of differences in feeding regime and anatomical location of the fat, fatty acid contents may also vary within one animal species (6). Nevertheless, this data set seemed to be a suitable working sample, since it is our intention to evaluate the usefulness of the convex hull technique in defining class-membership rather than to determine whether the available data are the most discriminating in the distinction between the different animal fat species.

Philosophy of the Applied Hull Technique

Consider the situation where objects or samples known to belong to either class A or class B are characterized by their measurement values for 2 parameters X1 and X2. Both kinds of samples can be represented as points in a 2-dimensional plane defined by the 2 parameters, the coordinates of the objects being their measurement values for X1 and X2 (Figure 1a). A possible way of defining class characteristics in this 2-dimensional situation is to delimit the area around each of the 2 possible classes by constructing class boundaries. Class boundaries can be determined by computing the convex hull around the cloud of points representing the samples of a specific class. This procedure corresponds with successively connecting the most extreme points of each separate class with each other (Figure 1b). The convex hulls then serve as the criterion for the classification of unknown samples, a classification that is, however, not probabilistic since the convex hull technique can be seen as a nonparametric method. In fact, class-models may be described in 2 ways, that is, by parametric or by nonparametric methods.

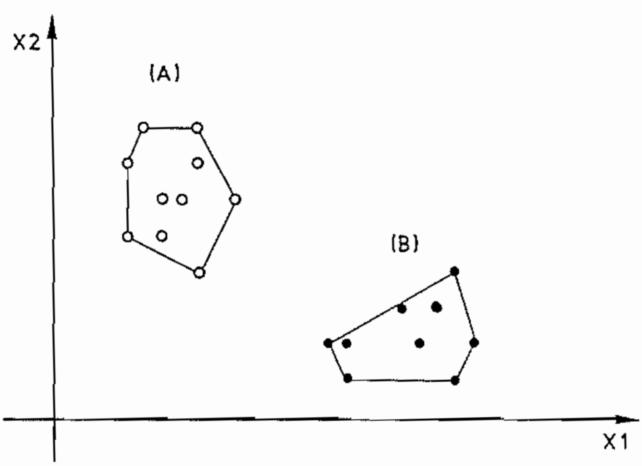

Figure 1b. Determination of class boundaries by construction of 2-dimensional convex hulls around each class.

Parametric techniques, i.e., techniques that define class-models or class boundaries on the basis of statistical parameters derived from the underlying distribution of the class samples (e.g., UNEQ, LDA . . .), have the advantage that the classification of test samples can be expressed in a probabilistic way. Each test object can be assigned a certain probability of belongingness to a specific class. A disadvantage is that the techniques are based on the assumption that the samples are bi- or multivariate normally distributed. Moreover, LDA also assumes equal within-group variance. These conditions are certainly not always fulfilled. If the underlying conditions are not fully satisfied, the classification rules derived from the technique are non-optimal and, consequently, classification results will not always be reliable. Nonparametric methods, on the contrary, do not make any assumptions with regard to the underlying distribution of the class samples, but the disadvantage of these methods is that the classification decision is perhaps too "clear-cut." A new sample is assigned to a specific class when it falls inside its class boundaries (in our case, the convex hull); otherwise it is considered to be an outlier with respect to that class.

Defining the convex hull around a 2-dimensional cloud of points is the same as finding the most extreme points in the corresponding class of objects. These extreme points form the vertices of the convex hull. A step-wise procedure for determining those vertices is summarized hereafter and is illustrated in Figure 2.

(1) Determine the centroid (c) of the class points. The 2 coordinates of the centroid are calculated as the class mean of each of the 2 respective parameters (Figure 2a). *(2)* Find the class point situated at maximal distance (d_{max}) of the centroid. This point defines the first vertex of the convex hull (v_1) (Figure 2a). *(3)* Determine the second vertex (v_2) by selecting the class point that forms a maximal angle (θ_{max1}) with the line that connects the centroid with the first vertex (Figure 2b). *(4)* Define consecutive hull vertices ($v_3 \ldots v_n$) according to the same principle, namely, by selecting the class point that forms a maximal angle with the line segment that connects the 2 former vertices (Figure 2c). *(5)* Repeat step 4 until the last vertex defined coincides with the first vertex ($v_n = v_1$). The construction of the convex hull around the cloud of points of a particular class is then completed (Figure 2d).

The procedure for defining class-membership of new samples with respect to the constructed hull boundaries can also be described in a step-wise manner (see also Figure 3):

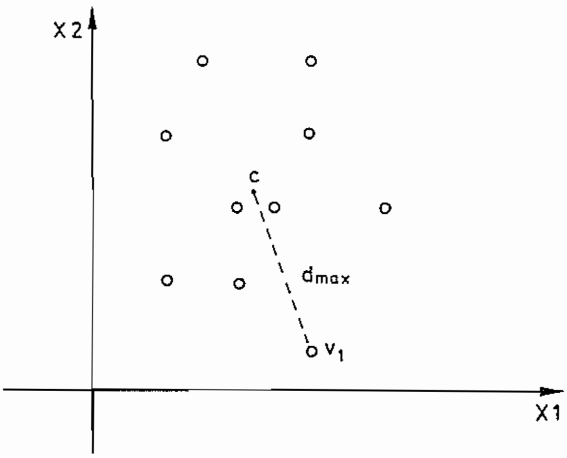

Figure 2a. Step-wise procedure for construction of convex hulls: Determination of centroid (c) and first vertex (v_1) by selecting class point with maximal distance (d_{max}) to centroid.

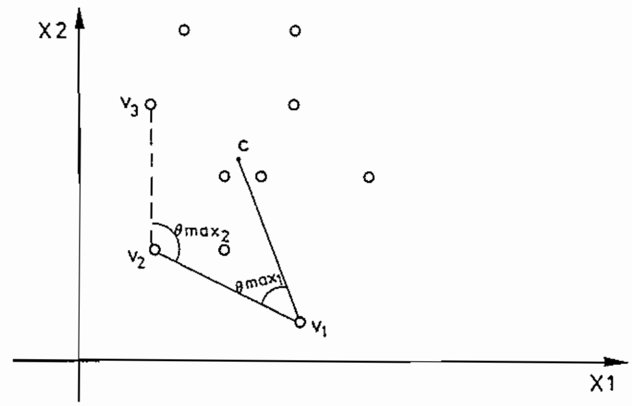

Figure 2c. Determination of consecutive vertices (v_3 . . .) by selecting class point that forms maximal angle (θ_{max2} . . .) with line segment that connects the 2 former vertices.

(1) Find the vertex to which the new point (T) is closest (v_c). *(2)* Determine the "direction" of the new point with respect to the closest vertex. More precisely, verify whether the new point lies in the wedge defined by vertices v_c and v_{c+1}. If this is the case, set v_d equal to v_{c+1}. Otherwise, that is, if the test point lies in the wedge defined by vertices v_c and v_{c-1}, set v_d equal to v_{c-1}. *(3)* Define whether the new point is (with respect to the line segment connecting vertices v_c and v_d) situated toward the centroid or away from it, or, in other words, whether the new point lies inside the triangle (c, v_c, v_d) or outside this triangle. When the new point lies inside this triangle, it is considered to be a member of the class under consideration. A description of an analogous procedure for defining convex hulls and for defining convex inclusion of new objects can be found in ref. 7.

Mixture class boundaries can also be described in terms of convex hulls. The procedure described hereafter is restricted to the situations where it can be assumed that the response of a linear combination of material from class A and from class B is equal to the linear combination of the responses of the material from both classes. This rules out situations where the combination of material from 2 classes interacts synergistically. If there is no interaction between the material from both classes, it can be stated that the mixture of a sample (a) from class A with a sample (b) from class B results in a new sample (c) with a composition intermediate to both original samples. Consequently, on the X1 vs X2 graph, the new mixture sample is situated on the line that connects both original samples (Figure 4a). This is true for all mixtures of A-samples with B-samples in every possible proportion. One observes that each of the possible "mixture lines" falls within the boundaries of a convex hull constructed around the total of points, that is, considering the points of class A and the points of class B as a whole (Figure 4b). Mixture class boundaries can thus be defined as the convex hull around the total of points belonging to both pure classes. The same procedure as described above can be applied to determine its hull vertices (i.e., the most extreme points) and to define class-membership.

The area occupied by the mixture class comprises thus the area occupied by class A, the area occupied by class B, plus the area between these 2 classes. Hence, since the mixture class includes both pure classes, it will not always be possible to distinguish a mixture (adulterated sample) and a pure sample. For instance, object d in Figure 4a can be a pure A-sample but it can just as well be a mixture sample obtained by mixing a large proportion of an A-sample with a small proportion of a B-sample. The distinction between a pure or a mixed sample is even less obvious as the 2 classes become more similar (small between-class distance) and as the classes become strongly heterogeneous (large within-class distance).

Evaluation Procedure

In testing the convex hull technique, we considered the 2-dimensional situation in which a distinction has to be made between 2 "pure" classes and the corresponding mixture class.

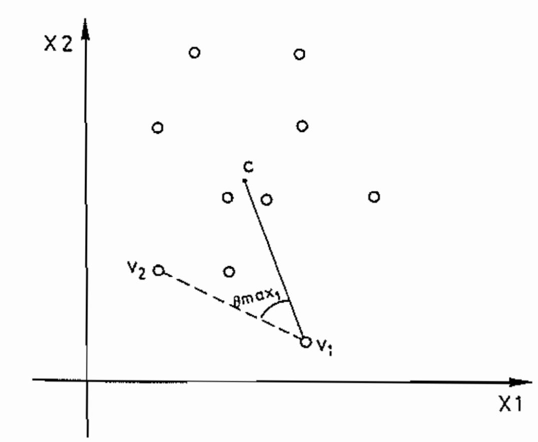

Figure 2b. Determination of second vertex (v_2) by selecting class point that forms maximal angle (θ_{max1}) with line that connects centroid with first vertex.

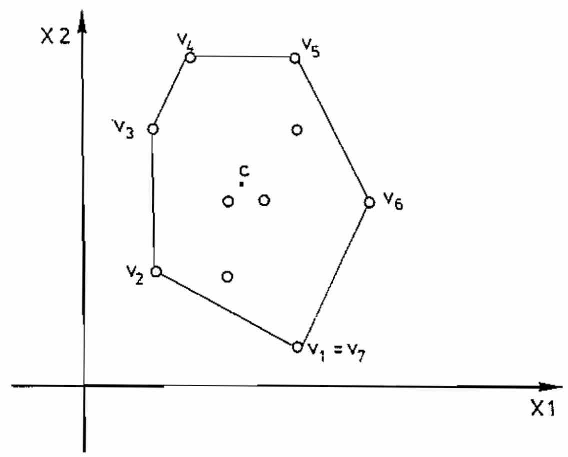

Figure 2d. Complete convex hull around points of a class.

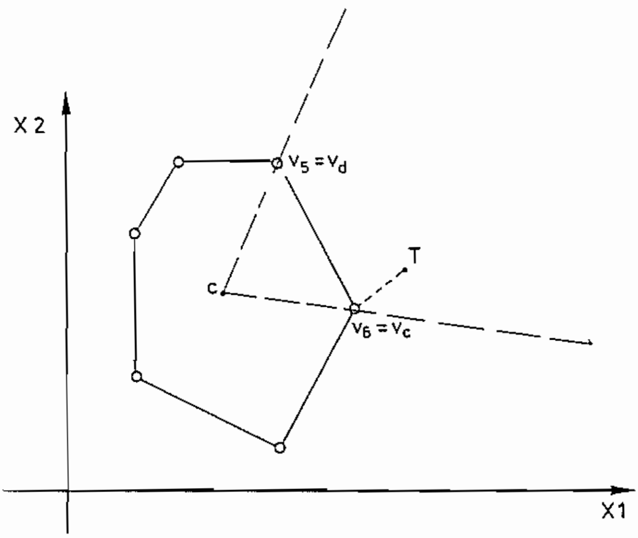

Figure 3. Definition of class-membership: T is closest to v_6, thus $v_6 = v_c$; T lies in wedge defined by vertices v_6 ($=v_c$) and v_5 ($=v_{c-1}$), thus $v_5 = v_d$; T falls outside triangle (c, v_c, v_d). Conclusion: T is outlier.

Starting from the total data set consisting of 4 different animal species (classes), every possible combination of 2 classes was considered. Convex hulls were constructed around each class (and mixture class) in each possible combination considering, respectively, the 21 pork, 20 chicken, 15 beef, or 14 horse fat samples from the original data set as representative samples.

Since in the original data set 7 parameters were determined (i.e., fatty acids incorporated at the 2-position of the triacylglycerols), it was necessary to reduce the number of dimensions before calculating the 2-dimensional class boundaries. We used principal component analysis to do this. Principal component analysis (PCA) can be seen as a visual dimension reduction technique with the objective of representing multidimensional data into a 2-dimensional space without losing too much of the original information residing in the data set. This goal is achieved by the computation of new variables (i.e., principal components or PCs) as orthogonal linear combinations of the original variables. The principal components are constructed in such a way that the first one explains more variation than the second, the second explains more than the third, etc. The dimension reduction results thus from the fact that most of the original variation in the data set is retained by the first few PCs. Hence, by carrying out a preliminary PC analysis on each subset of 2 classes, it becomes possible to represent the 2 classes into a 2-dimensional plane defined by the first 2 PCs and to construct the convex hulls around each class considering these 2 PCs as the new dimensions. New objects that have to be classified should then be projected in the corresponding PC plane.

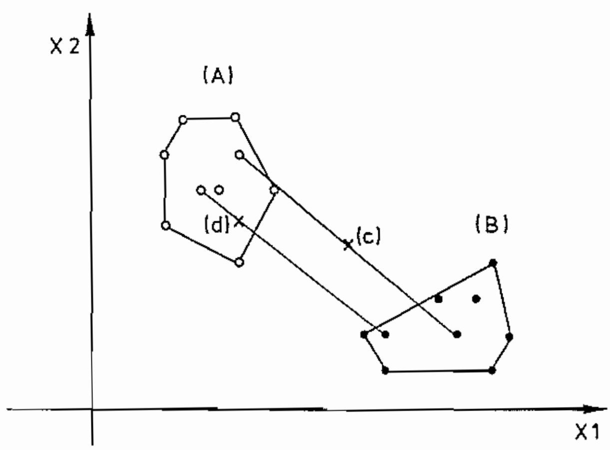

Figure 4a. Mixture samples (c) and (d) are situated on line that connects 2 original samples of which they are composed.

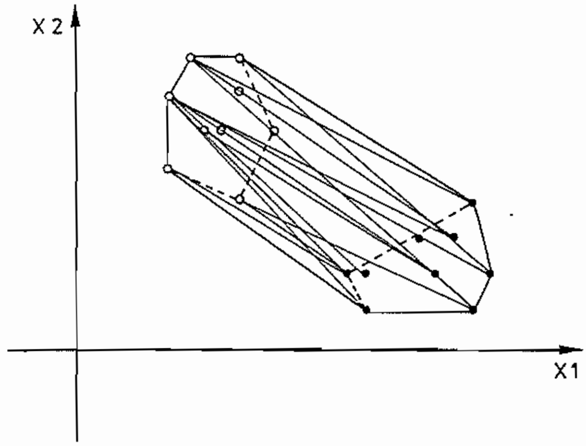

Figure 4b. Mixture class boundaries are defined as convex hull around total of points. (All possible mixture lines fall within these boundaries).

The choice of PCA as the dimension reduction method is made somewhat arbitrarily. There is no specific reason for preferring it above other factorial methods except that it is the most commonly used and therefore also the best known and best understood method. For reasons of visualization, only the first 2 PCs are retained as the basic dimensions for constructing 2-dimensional convex hulls. However, it should be mentioned that 2 PCs might not always be sufficient to approximate adequately the multidimensional data set. Consequently, classification results derived from convex hulls constructed in a PC plane that poorly represents multidimensional reality might be unreliable. Nevertheless, experience shows that on most occasions the 2 first PCs fairly approximate the real dispersion among the objects so that the constructed convex hulls can be considered as reliable parameters for defining class characteristics.

The method was thus tried in 6 different combinations, namely, in the definition of the pure and mixed classes of beef and horse fat, beef and pork fat, beef and chicken fat, horse and pork fat, horse and chicken fat, and, finally, pork and chicken fat. For each subset of 2 classes, the effectiveness of the hull boundaries in the definition of class-membership was evaluated by means of the classification of test objects that were generated by simulating—arithmetically—combinations of the genuine animal fat samples of the original data set. Each test set consists, more precisely, of 4 pure samples, 6 mixture samples, and 4 outliers. Pure samples were simulated by combining 2 randomly chosen samples of the corresponding pure class with each other; mixture samples are imitated by combining (in different proportions) randomly chosen samples of each of the 2 pure classes with each other; whereas the samples that are supposed to be outliers are arbitrarily chosen among the 2 remaining classes.

Results and Discussion

The graphical reproductions of the different classes, the constructed convex hulls, and the positions of the test samples with respect to those hull boundaries are represented in Figures 5 and 6 for, respectively, the beef-horse classification

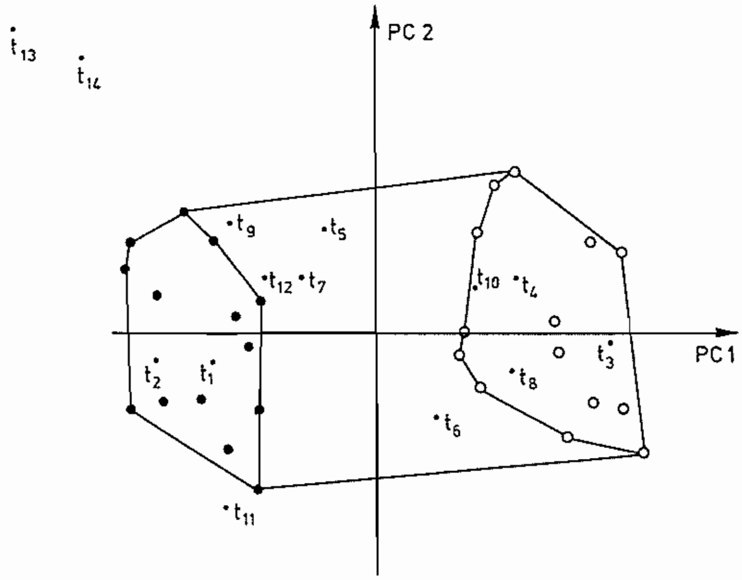

Figure 5. Beef-horse classification: 2-dimensional representation (PC1 vs PC2) of constructed convex hulls and position of test samples toward the hull boundaries. ● = beef samples; ○ = horse samples; t = test samples.

and the pork-chicken classification. The plots are PC1 vs PC2 plots. For reasons of classification, 3 different areas were determined on each of the 2-dimensional PC plots. The first area is defined as the area delimited by the convex hull boundaries constructed around the first pure class. It should comprehend all corresponding pure samples and possibly also some of the mixture samples (for reasons described above). The second area is defined as the area delimited by the convex hull around the second pure class. Again, it should include all corresponding pure samples and possibly some of the mixture samples. Finally, the third area is defined as the area comprehended by the convex hull around the complete mixture class minus the area comprehended by both pure classes. It defines an area in which only mixture samples should fall. Samples that fall outside these 3 areas are defined as outliers.

The detailed classification results obtained in each of the 2 abovementioned situations are represented in Tables 1 and 2. A positive sign means that the test object falls inside the corresponding area, a negative sign indicates that the test object falls outside it. Test objects that are "wrongly" classified are bracketed.

A general description of the results obtained in each of the 6 situations is given below:

Hull boundaries.—In each of the 6 combinations, the 2 pure classes are totally separated from each other. There is no overlap between the 2 corresponding convex hulls. However, one observes that in some situations (e.g., pork-chicken

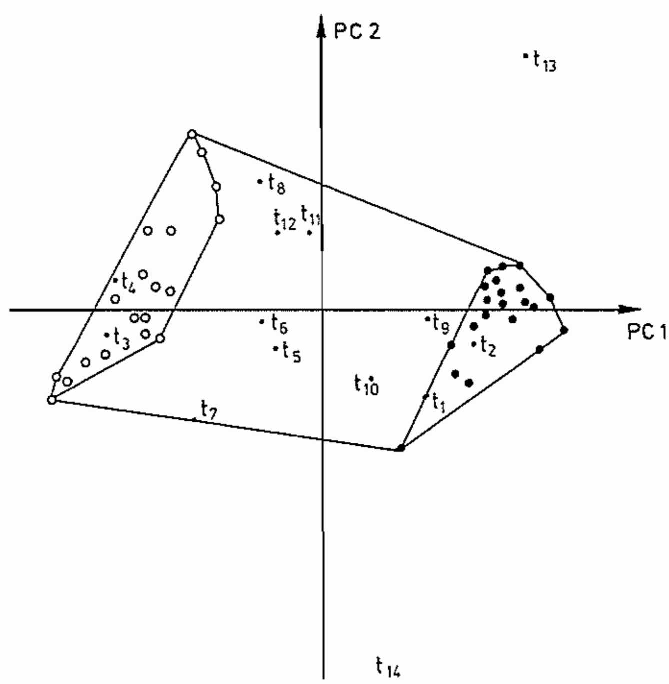

Figure 6. Pork-chicken classification: 2-dimensional representation (PC1 vs PC2) of constructed convex hulls and position of test samples toward hull boundaries. ● = pork samples; ○ = chicken samples; t = test samples.

Table 1. Areas of classification for beef, horse, and mixtures[a]

Test objects[b]	Area I: beef class	Area II: horse class	Area III: mixture class
T1 = Pure beef	+	–	–
T2 = Pure beef	+	–	–
T3 = Pure horse	–	+	–
T4 = Pure horse	–	+	–
T5 = Beef/horse: 1/1	–	–	+
T6 = Beef/horse: 1/1	–	–	+
T7 = Beef/horse: 2/1	–	–	+
T8 = Beef/horse: 1/2	–	[+]	–
T9 = Beef/horse: 3/1	–	–	+
T10 = Beef/horse: 1/3	–	[+]	–
T11 = Chicken	–	–	–
T12 = Chicken	–	–	[+]
T13 = Pork	–	–	–
T14 = Pork	–	–	–

[a] + = test object falls inside test area; – = test object falls outside test area; brackets = "wrong" classification.

[b] Simulation of mixture samples: Beef/horse:1/3, for example, is mixture sample consisting of 25% beef fat and 75% horse fat. It was simulated (arithmetically) by adding fatty acid percentages of one beef sample and 3 horse samples (randomly chosen among samples of original data set) and dividing resulting value by 4. Same remark accounts for all other mixture samples.

Table 2. Areas of classification for pork, chicken, and mixtures[a]

Test objects[b]	Area I: pork class	Area II: chicken class	Area III: mixture class
T1 = Pure pork	+	–	–
T2 = Pure pork	+	–	–
T3 = Pure chicken	–	+	–
T4 = Pure chicken	–	+	–
T5 = Pork/chicken: 1/1	–	–	+
T6 = Pork/chicken: 1/1	–	–	+
T7 = Pork/chicken: 2/1	–	–	+
T8 = Pork/chicken: 1/2	–	–	+
T9 = Pork/chicken: 3/1	–	–	+
T10 = Pork/chicken: 1/3	–	–	+
T11 = Beef	–	–	[+]
T12 = Beef	–	–	[+]
T13 = Horse	–	–	–
T14 = Horse	–	–	–

[a,b] See Table 1.

(see Figure 6)), the classes are rather heterogeneous resulting in convex hulls of an abnormal or stretched-out shape. It can be expected that with such widely spread hull boundaries, the classification results will not always be optimal.

Classification results.—(1) In each of the 6 situations, the 4 pure samples are correctly classified into the corresponding pure class.

(2) Thirty-three of the total 36 simulated mixture samples can unambiguously be defined as such since they all fall inside area III, i.e., the area delimited by the boundaries of the complete mixture class minus both pure classes. Three mixtures (2 for the beef-horse classification, 1 for the beef-chicken classification) are, however, classified into the area of one of the corresponding pure classes. Since the complete mixture class cannot totally be distinguished from the pure classes, this result can be expected, especially for mixtures of pure samples in unequal proportions and/or for mixtures of extreme pure samples (i.e., samples situated near the class boundaries). This is the case for the 3 mixture samples that are classified into a pure class: For instance, the 2 mixture samples that fall inside the pure horse class consist of a large proportion of horse fat (see Table 1) and, moreover, the horse fat samples, of which the mixtures are partly composed, are situated near the rightmost horse class boundaries.

(3) In 5 of the 6 situations, one or 2 of the samples that are supposed to be outliers are found inside the mixture class boundaries. These wrong classification results might be due to the fact that the mixture class boundaries are too liberal and cover too large a space (which may happen when the classes are heterogeneous). Another possibility is that, perhaps, the starting parameters are not discriminating enough to make a distinction between more than 2 different animal fat species and their corresponding mixtures.

In summary, it can be noticed that the developed hull technique leads to some false positive but no false negative results. In other words, some test samples are classified *inside* specific hull boundaries while in fact they should not be, but there are no samples that fall *outside* specific boundaries when they should be classified inside them. Therefore, it could be stated that, at least in this application, the method is probably sensitive but perhaps not selective enough.

Outlier detection especially seems to pose a problem. Possibly, the method could be improved by incorporating more sophisticated calculations for solving the abovementioned problem. The calculation of residuals, such as proposed and incorporated in the SIMCA method (1), seemed to us a possible help in solving the outlier problem. The residuals can be seen as a measure for the distance of an object toward the calculated PC plane. If the residuals of an individual object are found to be too large compared to the "mean" residuals of the global matrix to the PC plane, the object in question can be defined as an outlier. Classes would then be 3-dimensional: the 2-dimensional convex hull sandwiched between 2 boundaries in the third dimension, the boundaries being determined by allowable residuals. Preliminary calculations carried out on this subject prove that, although the outlier problem is not entirely solved, the evaluation of these residuals in combination with the 2-dimensional convex hull technique is certainly worth further investigation.

Conclusion

The research work carried out until now is not complete since, for instance, the effect of mixture samples composed of a known animal fat with an unknown animal fat is not yet investigated. However, it seems clear that the convex hull technique in combination with a principal component analysis can be considered as a mathematically simple and easily interpretable visual classification method giving good results. It must not be seen as a miracle technique bringing solutions to all kinds of problems. We would rather present it as an easily applicable technique that can be used besides other methods but not to the exclusion of other methods.

Certainly, further refinements could be implemented. For instance, the construction of ellipses instead of convex hulls can be seen as an analogous but probabilistic method. Boundaries of the corresponding mixture class could then possibly be defined as the common tangents to the 2 pure elliptical classes or perhaps the mixture class could be described as an ellipse of which the foci are the centroids of both pure classes. Those possibilities, however, have not been investigated. The technique could also be made more selective by the incorporation of supplementary and more sophisticated calculations for outlier detection. To solve the problem of selectiveness and accuracy, one also could consider the possibility of constructing convex hulls in more than 2 dimensions, but then the advantage of mathematical simplicity and visuality disappears.

Acknowledgment

Anne Thielemans is research assistant of the National Fund for Scientific Research.

References

(1) Wold S., & Sjostrom, M. (1977) in *Chemometrics: Theory and Applications*, B. R. Kowalski (Ed.), Symposium Series No. 52, American Chemical Society, Washington, DC, pp. 243–281
(2) Derde, M. P., & Massart, D. L. (1986) *Anal. Chim. Acta* **184**, 33–51
(3) Coomans, D., Jonckheer, M., Massart, D. L., Broeckaert, I., & Blockx, P. (1978) *Anal. Chim. Acta* **103**, 409–415
(4) Hopke, P. K. (1985) *Receptor Modeling in Environmental Chemistry*, John Wiley & Sons, New York, NY, Chapter 7
(5) Wold, S., Albano, C., Dunn, W., III, Esbensen, K., Hellberg, S., Johansson, E., Lindberg, W., & Sjostrom, M. (1984) *Analusis* **12**, 477–485
(6) Verbeke, R., & De Brabander, H. (1980) in *Proceedings of the 26th European Meeting of Meat Research Workers*, Colorado Springs, Vol. 1, pp. 150–153
(7) Preparata, F. P., & Shamos, M. I. (1985) *Computational Geometry*, Springer, Berlin, West Germany, Chapter 3