Food Chemistry **19** (1986) 225–234

Application of Linear Discriminant Analysis to Different Proteolysis Parameters for Assessing the Ripening of Manchego Cheese

G. Santa-María & M. Ramos

Instituto de Fermentaciones Industriales (CSIC),
Juan de la Cierva, 3, 28003 Madrid, Spain

&

J. A. Ordoñez

Departamento de Higiene y Microbiologia de los Alimentos,
Facultad de Veterinaria, Universidad Complutense, Madrid, Spain

(Received: 19 July, 1985)

ABSTRACT

Different proteolysis parameters, the nitrogenous fractions and the breakdown of caseins, were determined for Manchego cheeses at different stages of ripening. Linear discriminant analysis was applied to these parameters to ascertain the degree of ripening. Two discriminant functions enabling 100% correct classification of the cheeses into fresh, medium ripe and aged were found.

INTRODUCTION

The extent of proteolysis has been widely used as an indicator of the degree of ripening of cheese. A review of different methods of proteolysis analysis, including fractionated precipitation, electrophoresis, electrofocusing and separation by size, has recently been published (Pelissier,

Food Chemistry 0308-8146/86/$03·50

1984). In a study on proteolysis and flavour development in Cheddar cheese, Aston *et al.* (1983) found high, very significant correlations between the proteolysis indices, PTA-soluble amino nitrogen and TCA-soluble tyrosine levels and flavour development. They also recorded correlations between indices of proteolysis and the estimated age of cheese.

Different mathematical methods of data analysis have been applied in chemistry and food science. Pattern recognition has been proven successful in a wide variety of chemical problems, and a general review of its theory and application was published by Kowalski (1975). Factorial analysis was applied by Symonds & Cantagrel (1982) to differentiate among wines from five varieties of grape. Navarro *et al.* (1984) studied the application of different multivariate analysis methods in detecting adulterated citric juices.

Several researchers have used stepwise discriminant analysis in assessing the properties of beers, grape brandies and varietal wines (Brown & Clapperton, 1978; Schreier & Reiner, 1979; Cabezudo *et al.*, 1983).

Linear discriminant analysis was used by Smeyers-Verbeke *et al.* (1977) to identify different milk samples from gas chromatography data for fatty acids in twenty samples each of cow's, sheep's and goat's milk.

Recently, Pham & Nakai (1984) applied stepwise discriminant analysis to aqueous extracts from Cheddar cheese analyzed by HPLC to classify them according to ripening time.

In the present study, linear discriminant analysis was applied to different proteolysis indices in order to classify Manchego cheese according to ripening time.

MATERIALS AND METHODS

Five batches of cheese were made from sheep's milk employing normal procedures (Román Piñana, 1975). Samples were taken for analysis at different stages (1, 22, 30, 60, 90 and 120 days) of ripening.

Total nitrogen (TN), soluble nitrogen (SN), non-protein nitrogen (NPN), peptidic nitrogen (PN), amino-acid nitrogen ($N\text{-}NH_2$), tyrosine (TYR), tryptophan (TRP) and casein contents were analyzed as described by Ramos *et al.* (1981).

Statistical linear discriminant analysis

This method is based on the calculation of the discriminant function, defined as a linear combination of the parameters evaluated:

$$Z = \sum_{i=1}^{n} a_i x_i$$

where Z is the discriminant score, x_i represents the analytical parameters and a_i are the weighting coefficients obtained by maximizing the ratio of between-class to within-class variance (Cuadras, 1981).

The discriminant score is calculated by solving the system:

$$Z\ (x) = (M_A - M_B)'\ S^{-1}\ x$$

where M_A and M_B are the mean vectors and S is the sum of the within-class scatter matrices defined:

$$S = S_A + S_B$$

where S_A and S_B are variance–covariance matrices of the samples.

The discriminant function having the least classification error is calculated by means of the sequence described in a previous paper (Santa-Maria & Diez, 1984). We start with a two variable function and, after this, the number of variables is increased, one at a time, until adding of the next one does not imply a reduction of the classification error.

The discriminating threshold (Z_0) was used to classify the samples according to the method of Moret *et al.* (1980). The classification error of this procedure can be estimated by calculating the probability that $Z_A > Z_0$ or $Z_B < Z_0$ (assuming $Z_A < Z_B$).

The discriminating threshold is calculated in such a way that the standardized distances from Z_0 to $\bar{Z}_A$ and from Z_0 to $\bar{Z}_B$ are equal:

$$Z_0 = (SD_A\ \bar{Z}_B + SD_B\ \bar{Z}_A)/(SD_A + SD_B)$$

where SD_A and SD_B are the standard deviations of the Z values obtained for groups A and B, respectively.

For an unknown cheese sample, p, Z_p is calculated using:

$$Z_p = \sum_{i=1}^{n} a_i\ x_{pi}$$

The sample is classified as 'A' (or 'B') if Z_p is $< Z_0$ (or $> Z_0$) (on the hypothesis that $\bar{Z}_A < \bar{Z}_B$).

A program developed by the authors and run on an Olivetti M-20 computer was used to perform the calculations. The listing of that program may be obtained from the authors.

RESULTS AND DISCUSSION

Linear discriminant analysis was applied to assess the age of cheeses using eighteen different proteolysis parameters. Table 1 shows the five discriminant functions, the discriminating threshold and the classification errors for the different groups.

TABLE 1

Weighting Coefficients, Discriminating Threshold (Z_0) and Classification Error (E) for the Cheeses Using the Respective Discriminant Functions (Z) (Subscripts Indicate Ripening Time in Days)

	Discriminant function				
	$Z_{D1,D22}$	$Z_{D22,D30}$	$Z_{D30,D60}$	$Z_{D60,D90}$	$Z_{D90,D120}$
Total nitrogen (TN)	—	—	—	—	—
Soluble nitrogen (SN)	0·464	0·645	—	0·397	—
Non-protein nitrogen (NPN)	—	0·184	−0·494	1·398	—
Peptidic nitrogen (PN)	—	—	0·324	0·174	−0·382
Amino-acid nitrogen ($N\text{-}NH_2$)	—	—	—	—	—
Tyrosine (TYR)	1·380	−1·461	—	0·226	0·690
Tryptophan (TRP)	−0·903	−5·011	—	—	2·502
SN/TN × 100	—	—	—	—	—
NPN/TN × 100	—	0·602	—	−0·155	—
NNH_2/TN × 100	—	—	—	—	—
NPN/SN	−0·020	—	—	—	—
$N\text{-}NH_2$/SN × 100	—	—	−0·047	—	−0·020
PN/SN × 100	0·390	0·404	—	0·128	—
Polypeptides (%)	0·280	0·033	0·216	0·034	—
β-casein	—	−0·085	0·074	0·078	—
α_s-casein	—	—	—	0·015	0·096
α_sI-casein	0·264	−0·019	—	—	—
Residual α_s-casein (%)	—	−0·098	0·035	—	—
Discriminating threshold (Z_0)	23·90	11·20	5·99	22·77	4·43
Classification error (E)	7×10^{-5}	4×10^{-15}	3×10^{-3}	5×10^{-3}	0·356

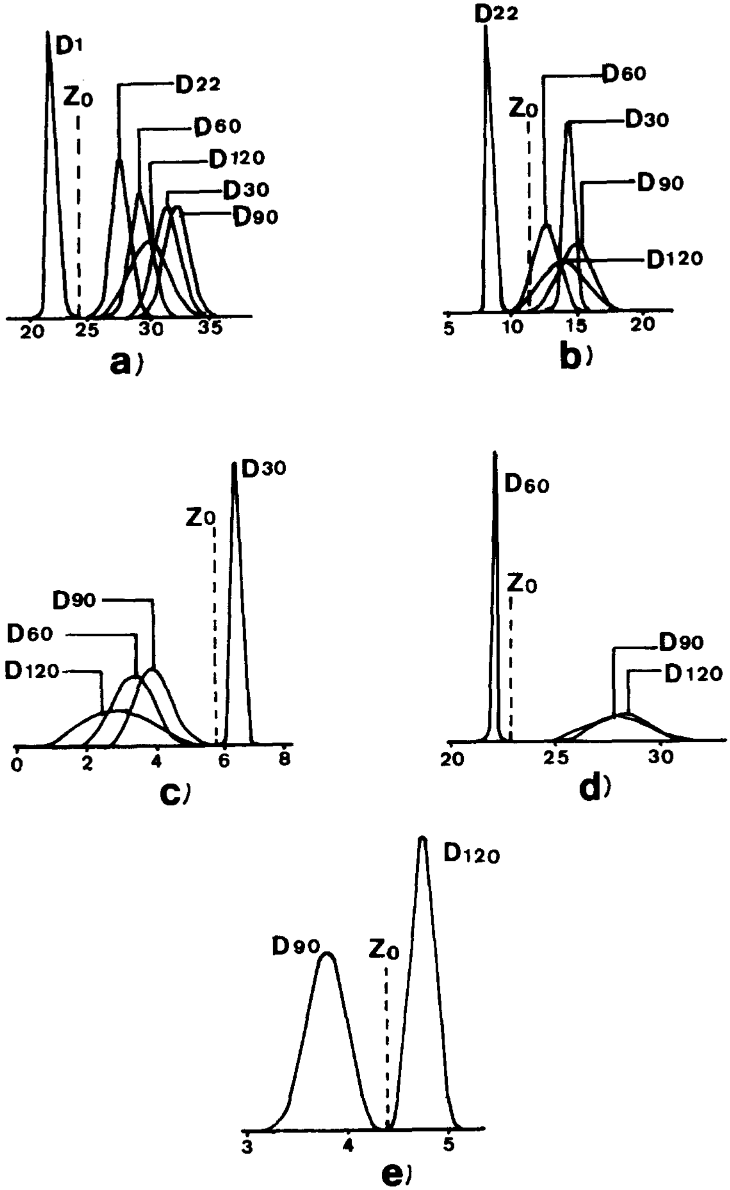

Fig. 1. Normal distributions of the discriminant scores for cheeses grouped by ripening time: 1 day (D_1), 22 days (D_{22}), 30 days (D_{30}), 60 days (D_{60}), 90 days (D_{90}) and 120 days (D_{120}). (a) Function $Z_{D1, D22}$. (b) Function $Z_{D22, D30}$. (c) Function $Z_{D30, D60}$. (d) Function $Z_{D60, D90}$. (e) Function $Z_{D90, D120}$. Z_0 = discriminating threshold.

TABLE 2

Weighting Coefficients Discriminating Threshold (Z_0) and Classification Error (E) for Discriminant Functions Applied to 1-Day Old (F), Medium Ripe (N: 22-, 33- and 60-Day Old), and Aged (M: 90- and 120-Day Old) Cheeses

	Discriminant function	
	$Z_{F,N}$	$Z_{N,M}$
Total nitrogen (TN)	−0·426	1·730
Soluble nitrogen (SN)	4·292	1·028
Non-protein nitrogen (NPN)	0·355	−15.961
Peptidic nitrogen (PN)	−4·068	−8·030
Amino-acid nitrogen ($N\text{-}NH_2$)	−4·449	0·476
Tyrosine (TYR)	5·187	0·887
Tryptophan (TRP)	−5·009	3·487
SN/TN × 100	−1·169	0·145
NPN/TN × 100	0·115	6·788
NNH_2/TN × 100	2·506	−0·978
NPN/SN × 100	0·306	−0·916
$N\text{-}NH_2$/SN × 100	−0·122	0·903
PN/SN × 100	1·567	1·617
Polypeptides (%)	0·146	1·008
β -casein	−0·058	0·233
α_s-casein	−1·328	0·802
α_sI-casein	0·528	1·369
Residual α_s -casein (%)	−0·202	−0·340
Discriminating threshold (Z_0)	18·83	152·95
Classification error (E)	2×10^{-7}	1×10^{-3}

The classification errors were of the order of 10^{-3} or less, except in the case of discriminant function $Z_{D90,D120}$. Completely correct classification of the samples was achieved in all cases, as indicated in Table 1.

The discriminant function with the lowest classification error was not always the function calculated using the largest number of parameters or the function calculated exclusively on the basis of those parameters with the highest statistical significance (Santa-María & Diez, 1984).

Of the eighteen parameters studied, total nitrogen (TN), amino-acid nitrogen ($N\text{-}NH_2$) and the ratios SN/TN × 100 and $N\text{-}NH_2$/TN × 100 were not selected for use in any of the five discriminant functions. In contrast, tyrosine (TYR) and the polypeptides were used in four of the functions and soluble nitrogen (SN) in three.

The normal distributions of the discriminant scores (Z) for each cheese

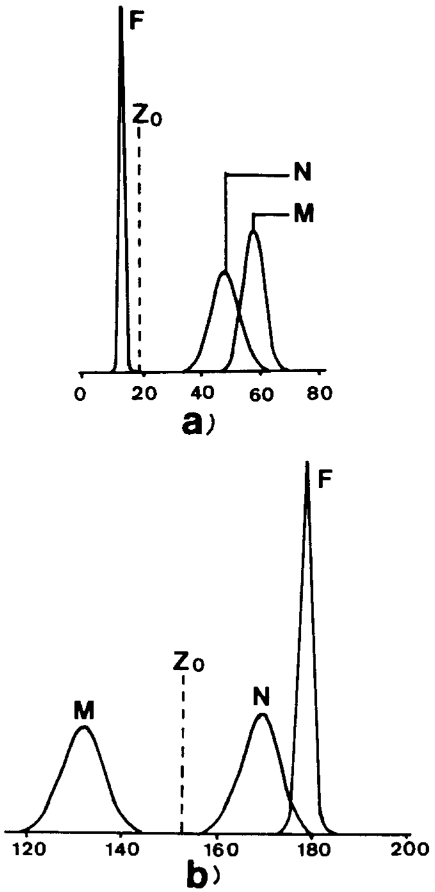

Fig. 2. Normal distributions of the discriminating scores for the fresh (F). medium ripe (N) and aged (M) cheese groupings. (a) Function $Z_{F,N}$. (b) Function $Z_{N,M}$. Z_0 = discriminating threshold.

sample and its respective function, given in Table 1, are plotted in Fig. 1. Plots of the corresponding discriminating thresholds (Z_0) are also shown.

Cheeses can be identified on the basis of their ripening times by successively applying the functions $Z_{D1,D22}$, $Z_{D22,D30}$, $Z_{D30,D60}$, $Z_{D60,D90}$ and $Z_{D90,D120}$.

The sole difficulty encountered was for cheeses ripened for 60 and 120 days, which were correctly classified using the function $Z_{D22,D30}$ in only 95% of all attempts (Fig. 1(b)).

Applying Student's t-test to the different parameters evaluated indicated that the smallest differences existed between 30- and 60-day-old

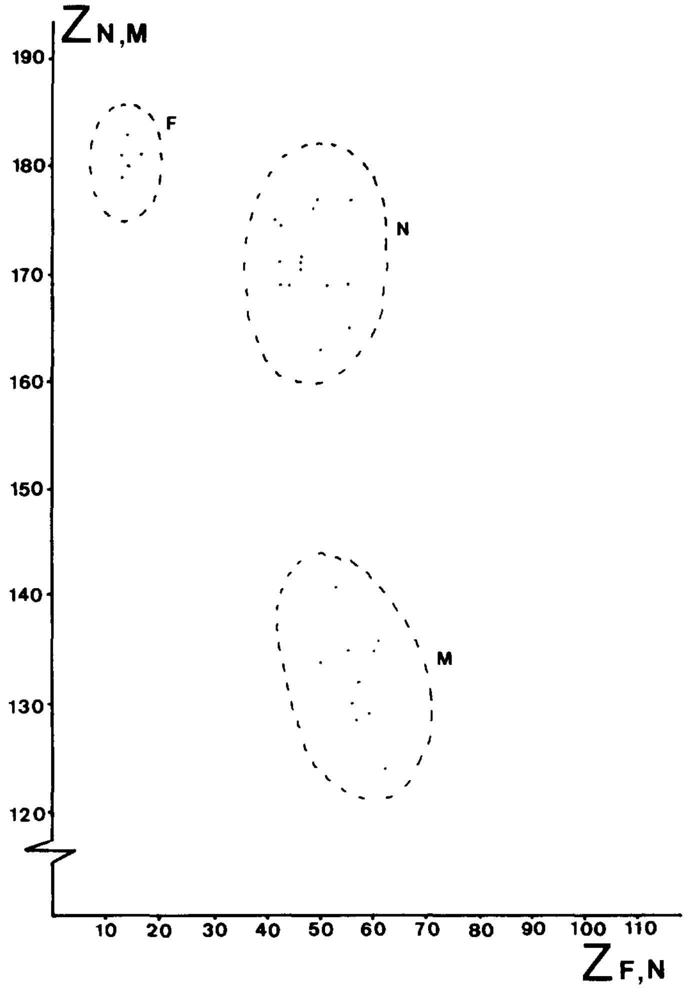

Fig. 3. Distributions of samples of fresh (F), medium ripe (N) and aged (M) cheeses by their discriminant scores ($Z_{F,N}$ and $Z_{N,M}$).

cheeses (D_{30} and D_{60}) and between 90- and 120-day-old cheeses (D_{90} and D_{120}). The largest differences were found between 60- and 90-day-old cheeses (D_{60} and D_{90}).

On the basis of these results, the cheeses were grouped as follows: 1-day old, fresh (F), five samples; medium ripe (N: D_{22}, D_{30}, D_{60}), fifteen samples and aged (M: D_{90}, D_{120}), ten samples. These groupings were felt to be of great practical interest, and discriminating analysis was applied.

Table 2 shows the values of the weighting coefficients, discriminating threshold and classification errors for functions $Z_{F,N}$ and $Z_{N,M}$.

Classification of these cheese samples using these two functions was completely correct, with errors of 2×10^{-7} for $Z_{F,N}$ and 1×10^{-3} for $Z_{N,M}$, as indicated in Fig. 2, which presents plots of the normal distributions of the discriminant scores for the three groups (*F*, *N* and *M*) and for each of the two discriminant functions.

The percentage of correct classifications employing this method was higher than that achieved by Pham & Nakai (1984) applying stepwise discriminant analysis to aqueous cheese extracts.

Using the two functions indicated above, cheeses can be classified by ripening time into fresh, medium ripe and aged. The results for the thirty cheese samples are plotted in Fig. 3 on a co-ordinate grid formed using the two discriminant functions, $Z_{F,N}$ and $Z_{N,M}$, as the axes.

It is true that all eighteen variables were used in the two functions but, since six of these were ratios while an additional five were obtained from electrophoretic analysis, it can be seen that Manchego cheeses can be classified according to ripening time by analysis of the nitrogenous fractions (TN, SN, NPN, PN, N-NH_2, TYR and TRP) and electrophoresis of the caseins.

REFERENCES

Aston, J. W., Durward, I. G. & Dulley, J. R. (1983). Proteolysis and flavour development in Cheddar cheese. *Austr. J. of Dairy Techn.*, **38**, 55–59.

Brown, D. G. & Clapperton, J. F. (1978). Discriminant analysis of sensory and instrumental data on beer, *J. Inst. Brew.*, **84**, 318–23.

Cabezudo, M. D., Herraiz, M. & Gorostiza, E. F. (1983). On the main analytical characteristics for solving enological problems, *Proc. Biochem*, **18**, 17–22.

Cuadras, C. M. (1981). *Metodos de análisis multivariante*, Ed. Euni Bar, Barcelona.

Kowalski, B. R. (1975). Measurement analysis by pattern recognition. *Anal. Chem.*, **47**, 1152A–62A.

Moret, I., Scarponi, G., Capodaglio, G., Zanin, S., Canaiani, G. & Toniolo, A. (1980). Analytical parameters in the characterization of three Venetian wines. Application of the statistical linear discriminant analysis. *Am. J. Enol. Vitic.*, **31**, 245–9.

Navarro, J. L., Aristoy, M. & Izquierdo, L. (1984). El análisis estadístico en la detección de adulteraciones en zumos cítricos, *Rev. Agroq. Tecnol. Alim.* **24**, 48–58.

Pelissier, J. P. (1984). Proteolysis of caseins, *Sciences des Aliment*, **4**, 1–35.
Pham, A. M. & Nakai, S. (1984). Application of stepwise discriminant analysis to high pressure liquid chromatography profiles of water extracts for judging ripening of Cheddar cheese. *J. Dairy Sci.*, **67**, 1390–6.
Ramos, M. Barneto, R. & Ordoñez, J. A. (1981). Evaluation of a specific starter for Manchego cheese production. *Milchwiss.*, **36**, 528–32.
Román Pinaña, M. (1975). Microbial flora of Spanish Manchego cheese. I. Development during manufacture and ripening. *Le lait*, **547**, 401–13.
Santa-Maria, G. & Diez, C. (1984). Application de l'analyse multivariante a l'etude des composes phenoliques des vins. *Bul. C. R. Groupe Polyphenols*, **11**, 285–300.
Schreier, P. & Reiner, L. (1979). Characterisation and differentiation of grape brandies by multiple discriminant analysis. *J. Sc. Food Agric.*, **30**, 319–27.
Smeyers-Verbeke, J., Massart, D. L. & Coomans, D. (1977). Application of linear discriminant analysis to the differentiation of pure milk from different species and mixtures. *J. Ass. Off Anal. Chem.*, **60**, 1382–5.
Symonds, P. & Cantagrel, R. (1982). Application de l'analyse discriminante a la différenciation des vins. *Ann. Fals. Exp. Chim.*, **75**, 63–74.