Classification of Commercial Orange Juice Types by Pattern Recognition Involving Volatile Constituents Quantified by Gas Chromatography

Philip E. Shaw,* Béla S. Buslig,[†] and Manuel G. Moshonas

Citrus and Subtropical Products Laboratory, U.S. Department of Agriculture, Agricultural Research Service, South Atlantic Area, 600 Avenue S N.W., Winter Haven, Florida 33881

Four types of commercial orange juices were compared by use of a multivariate computer pattern recognition program as either fresh squeezed, pasteurized not from concentrate, single-strength juice reconstituted from concentrate, or aseptically packaged single-strength juice from concentrate. Nineteen volatile constituents of each juice were quantified by headspace gas chromatography for these multivariate analyses. Elimination of one to five components considered least important generally led to poorer classification. The mean value and range of values quantified for each component in each juice classification are tabulated.

INTRODUCTION

Orange is the most popular of the fruit beverage flavors (Anonymous, 1983), and orange beverages are sold in a wide variety of containers and under varying time and temperature storage conditions. In addition, orange is the most delicate and complex citrus flavor (Shaw, 1991). Many efforts have been made to categorize orange products by sensory and/or quantitative analytical information to help identify and categorize various orange juice products as to quality, cultivar, and country or place of origin.

In recent years, computerized pattern recognition programs have become available which make it possible to categorize different samples of a food by considering many variables that can be measured, often in a single analytical determination (Massart et al., 1988). Some of these programs have been used to categorize orange juices and natural orange flavor fractions. Trace metal analysis of citrus juices was used to categorize orange juice by place of origin (Bayer et al., 1980; Page, 1986; Nikdel et al., 1988) and to distinguish blends of orange and grapefruit juice from 100% orange juice (Nikdel and Fishback, 1989). Early work on correlating orange juice flavor scores vs chemical and physical analysis data relied on various combinations of simple and multiple linear regression approaches (Fellers and Buslig, 1971, 1972, 1973; Attaway and Carter, 1971; Attaway et al., 1972; Carter et al., 1975). The series of multiple regression equations generated underscored the complexity of individual component contributions to the perception of flavor. Some of the same data were utilized by Rouseff and Nagy (1987), who correlated 34 quality factors quantified in orange juice with flavor scores and employed a multivariate pattern recognition program to show 5 of the instrumentally determined factors to be major quality indicators for orange juice. Nagy et al. (1992) correlated sensory evaluation of 16 natural aqueous orange aroma samples with 32 volatile constituents quantified by gas chromatography (GC). Two orange essence oil varieties were also classified by pattern recognition techniques (Mayfield et al., 1986).

In the present study pattern recognition was used to categorize commercial orange juice samples into four types

---

† Present address: Florida Department of Citrus, Winter Haven, FL.

Table I. Juice Samples Evaluated

sample no.[a]	type of juice sample[b]	type of packaging
1-4	pasteurized NFC[c]	flexible gable-top
5-7	pasteurized NFC[c]	flexible gable-top
8-10	pasteurized NFC[c]	flexible gable-top
11-14	pasteurized NFC[c]	flexible gable-top
15-17	pasteurized NFC[c]	flexible gable-top
18-21	fresh-squeezed[d]	plastic bottle
22-25	fresh-squeezed[d]	plastic bottle
26-29	fresh-squeezed[d]	plastic bottle
30-33	unpasteurized[e]	plastic bottle
34-36	single strength from concentrate	glass
37-39	single strength from concentrate	flexible gable-top
40-42	frozen concentrate[f]	paperboard with metal ends
43-45	frozen concentrate[f]	paperboard with metal ends
46-48	frozen concentrate[f]	paperboard
49-51	asepctic pack concentrate[f,g]	laminated multilayer
52-54	canned	tin can
55-57	aseptic pack[g]	laminated multilayer
58-60	aseptic pack[g]	laminated multilayer

[a] Each replicate of each type of juice sample individually numbered for identification in Figures 1-3. [b] Commercially packaged for retail market unless otherwise noted. [c] NFC, not from concentrate. [d] Commercially packaged unpasteurized juices purchased on April 27 (no. 18-21), April 17 (no. 22-25), and April 20 (no. 26-29). [e] Fresh unpasteurized Valencia juice commercially extracted on April 3, 1992. [f] Reconstituted to 11.8 °Brix juice prior to analysis. [g] Packaged aseptically in 250-mL rectangular flexible carton.

based on principal component analysis of 19 volatile constituents quantified by GC.

MATERIALS AND METHODS

Juice Samples. Eighteen commercial juice samples were used, and each analysis was replicated three or four times as outlined in Table I. Five juice samples were of pure premium pasteurized juice not from concentrate (samples 1-17 in Table I). Four samples of fresh juice (not pasteurized) were studied, including three "fresh-squeezed" commercial samples with a 17-day expiration date from time of juice extraction (Florida Department of Citrus, 1992) and one fresh Valencia orange juice sample taken directly from a local commercial juice extraction line (samples 18-33). Seven samples from orange concentrate included three from frozen concentrated orange juice (FCOJ), one aseptically packed concentrate reconstituted for analysis, and three single-strength juices from concentrate, one each packaged in glass, fiberboard carton, and tin can (samples 34-54). Two aseptically packed samples were single-strength juice from concentrate packaged in flexible multilayered 250-mL cartons (samples 55-60). All commercially packaged samples

810 *J. Agric. Food Chem.*, Vol. 41, No. 5, 1993

Shaw et al.

Table II. Amounts of Volatile Compounds (Parts per Million) in Commercial Orange Juice Types

	juice type[a]							
	fresh[b]		pasteurized[c]		from concentrate[d]		aseptic pack[e]	
compound	mean	range	mean	range	mean	range	mean	range
aldehydes								
acetaldehyde	8.4B	2.3–11.6	7.6B	5.8–9.7	3.3C	0.72–6.71	11.5A	11.0–12.2
hexanal	0.16B	0.01–0.29	0.26A	0.18–0.32	0.07C	0–0.13	0.07C	0.04–0.10
octanal	0.16B	0.01–0.38	0.42A	0.20–0.83	0.48A	0.13–0.83	0.22B	0.19–0.24
decanal	0.13C	0–0.35	0.20C	0–0.48	0.35B	0–0.74	1.57A	1.40–1.73
esters								
ethyl acetate	0.61A	0.060–1.81	0.20B	0.14–0.32	0.07B	0–0.13	0.19B	0.17–0.23
methyl butyrate	0.03A	0–0.11	0.02AB	0–0.07	0.006B	0–0.02	0.01B	0–0.02
ethyl butyrate	0.49A	0.23–0.72	0.51A	0.30–0.89	0.17B	0.01–0.33	0.09B	0.04–0.14
alcohols								
methanol	109A	67–190	46B	13–64	5.4C	0–22	21C	6–37
ethanol	782A	282–1420	898A	348–1150	244B	14–642	673A	468–825
2-methyl-1-propanol	0.03A	0–0.05	0.045A	0.019–0.084	0.027A	0–0.11	0.047A	0–0.20
hexanol	0.65A	0.11–1.28	0.034B	0–0.074	0B	0	0B	0
cis-3-hexenol	0.42B	0.08–0.70	0.45B	0.05–1.15	0.36B	0–0.69	1.88A	1.73–2.14
trans-2-hexenol	0.06A	0–0.36	0.002B	0–0.014	0.018B	0–0.14	0B	0
linalool	0.81B	0–1.55	1.01B	0.01–3.37	0.72B	0–1.22	4.8A	3.7–6.06
α-terpineol	0.45AB	0–1.47	0.75A	0–2.49	0.10B	0–1.92	0.81A	0.05–1.96
hydrocarbons								
α-pinene	0.45BC	0.20–1.32	0.71B	0.16–1.98	1.49A	0.23–5.03	0.01C	0–0.06
sabinene	0.14A	0–0.37	0.01B	0–0.058	0.06B	0–0.25	0B	0
γ-terpinene	0.08B	0.055–0.12	0.11B	0–0.20	0.19A	0.04–0.53	0.06B	0.04–0.08
valencene	2.92A	0.62–10.1	2.76A	0.63–14.8	0.39B	0–2.48	0.18B	0–0.67

[a] Means in the same row followed by a common letter are not significantly different ($P > 0.05$); values for three or four replicates of each sample were used in these calculations. [b] Four unpasteurized juice samples: three commercial "fresh-squeezed" samples and one commercially extracted and finished Valencia juice. [c] Five pure premium pasteurized juice samples, not from concentrate. [d] Seven samples, three from frozen concentrated orange juice, one from aseptically packaged concentrate reconstituted for analysis, and three single-strength juices from concentrate. [e] Two aseptically packaged juices in 250-mL cartons.

were purchased at local markets except for one aseptically packaged single-strength juice which was obtained directly from a processing line and the aseptically packaged concentrate which was provided by the processor, and each was stored at –18 °C until analyzed.

Headspace GC Analysis of Juice. A 2-mL sample of juice in a 10-mL vial sealed with a crimptop cap with TFE/silicone septum seal was equilibrated for 15 min at 80 °C in a Model HS-6 headspace sampler attached to a Perkin-Elmer Model 8500 GC with an FID detector. A 0.53 mm × 30 m polar Durowax column with 1.0-μm film thickness (J&W Scientific, Folsom, CA) was used with a helium carrier gas at a head pressure of 6.0 psi (81 cm/s linear gas velocity). Injection conditions for the headspace sampler were 0.5-min vial pressurization time followed by 0.02-min injection time. The GC was temperature programmed at 40 °C isothermally for 6 min and then raised at 6 °C/min to 180 °C. The FID amplifier range was set for high sensitivity, and the detector temperature was 250 °C. All determinations were carried out in triplicate or quadruplicate. Individual compounds were identified by comparison of retention times with those of standards and by enrichment of juice with authentic samples. Concentrations for the 19 compounds were calculated with regression equations, determined by injecting five different concentrations of each compound added to a juice base to obtain a peak height calibration curve as described previously (Nisperos-Carriedo and Shaw, 1990). The juice base was prepared by reconstitution to 11.8 °Brix from concentrated juice (pumpout) from an evaporator that contained no added flavor fractions.

Statistical Analyses. Significant differences between means in Table II were determined by analysis of variance (ANOVA) using the General Linear Model (GLM) procedure, a program package of the Statistical Analysis System (SAS Institute Inc., Cary NC). Specific differences between means in Table II were determined by least significant difference (LSD) at a 95% confidence level.

The EinSight Version 3.0 data analysis and multivariate pattern recognition program from Infometrix, Inc., Seattle, WA, was used in this study. This program can preprocess data by the following methods: autoscale, mean center, variance scale, or range scale. In this study, autoscale preprocessing was used, since it is a preferred method when large variations in relative

Table III. Eigenvector Report from Autoscaled Data

eigenvector	variance	% variation	% cumulative variation
1	280.2	25.0	25.0
2	221.8	19.8	44.8
3	148.6	13.3	58.0
4	132.1	11.8	69.8
5	83.3	7.4	77.2

magnitudes of measurements are present, such as found here (Nikdel et al., 1988; Massart et al., 1988).

RESULTS AND DISCUSSION

Commercially produced fresh and processed orange juice samples were analyzed by headspace GC and the quantities of 19 compounds determined by this method are summarized in Table II, in four categories of juices. Table II lists average values and ranges of voltaile flavor compounds present in various juice products. Statistical comparison by ANOVA indicated that no single compound could be used to differentiate these four juice categories since no significant difference in means for any one single compound was found in all four categories of juices. In addition, a cursory examination of these data showed no other obvious correlations (Nikdel et al., 1988). Since 19 variables are measured in a single GC determination, a multivariate pattern recognition approach involving all of the 19 variables should be more effective in recognizing differences among the samples analyzed (Massart et al., 1988).

A computer multivariate pattern recognition program which considers most or all of the 19 volatile compounds quantified was used to differentiate the four categories of orange juice shown in Table II. Principal component analysis of 19 compounds quantified by headspace GC afforded the eigenvector report shown in Table III. Eigenvectors 1 and 2 represent 45% of the total variance for these 19 compounds. As described by Nikdel and

Classification of Commercial Orange Juice Types

J. Agric. Food Chem., Vol. 41, No. 5, 1993   **811**

Table IV. Principal Component Loadings Autoscaled Data

| variable	loading 1	variable	loading 2	variable	loading 3
ethyl acetate	0.361078	*cis*-3-hexenol	0.388887	hexanal	0.493630
methyl butyrate	0.337852	linalool	0.381779	ethyl butyrate	0.380153
valencene	0.324008	ethanol	0.376715	octanal	0.349594
ethyl butyrate	0.315764	acetaldehyde	0.3723335	hexanol	-0.312923
trans-2-hexenol	0.312118	α-pinene	-0.338202	ethanol	0.284959
hexanol	0.293149	α-terpineol	0.265903	*trans*-2-hexenol	-0.260631
methanol	0.272859	decanal	0.263688	2-methyl-1-propanol	0.206338
decanal	-0.265331	γ-terpinene	-0.261115	linalool	-0.193269
linalool	-0.222216	methanol	0.191467	ethyl acetate	-0.190756
cis-3-hexenol	-0.214736	hexanal	0.188115	sabinene	0.177465
octanal	-0.194837	sabinene	0.077326	*cis*-3-hexenol	-0.153625
acetaldehyde	0.185368	2-methyl-1-propanol	0.077269	decanal	-0.152429
γ-terpinene	-0.143219	octanal	-0.065345	methanol	-0.120360
ethanol	0.092499	ethyl butyrate	0.062495	α-pinene	0.117971
hexanal	0.085968	*trans*-2-hexenol	-0.052728	α-terpineol	-0.105861
α-terpineol	-0.081146	valencene	-0.044464	acetaldehyde	0.077282
sabinene	0.076923	hexanol	0.043328	γ-terpinene	-0.035781
α-pinene	-0.057471	ethyl acetate	0.023423	methyl butyrate	0.031165
2-methyl-1-propanol	-0.050022	methyl butyrate	0.017551	valencene	0.007962
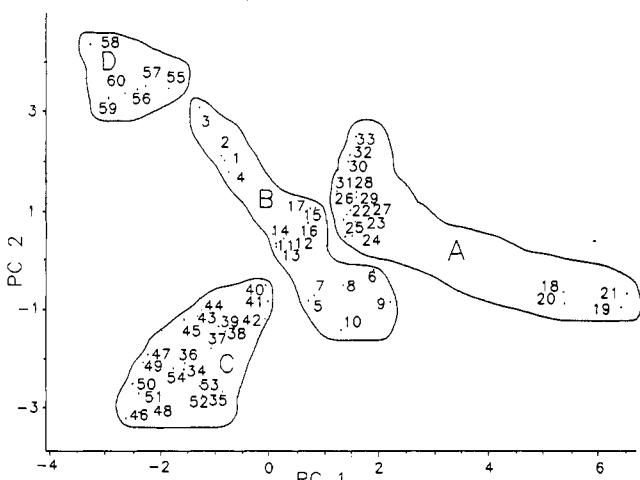
Figure 1. Two-dimensional plot of scores for eigenvector 1 (PC 1) vs eigenvector 2 (PC 2) from 19 compounds to categorize processed orange juice samples. Sample categories circled are (A) fresh squeezed, not pasteurized; (B) pasteurized not from concentrate; (C) reconstituted from concentrate; (D) single-strength aseptically packaged from concentrate.

Figure 2. Three-dimensional plot of scores for eigenvectors 1 (x axis), 2 (y axis), and 3 (z axis) from 19 compounds to categorize processed orange juice samples using categories as listed in Figure 1.

Fishback (1989) for an earlier version of this EinSight pattern recognition program, eigenvectors for principal components represent more variance than that found for any single component. A graphic representation of the scores for eigenvectors 1 and 2 is shown in Figure 1. This is a two-dimensional representation of the 19-dimensional data. Eigenvector 1 (x axis) represents the axis of greatest variance through the measurement space, and eigenvector 2 (y axis) is the axis of second greatest variance, which is perpendicular to the first axis (Nikdel et al., 1988). Each numbered point in Figure 1 represents a two-dimensional plot of principal components 1 and 2 for an individual headspace GC run for a given juice sample. For example, points 18-21 represent the two-dimensional plots of four replicate analyses of a single fresh-squeezed juice sample. The four juice categories separated in Figure 1 show no overlapping values, and all of the pure premium juice samples not from concentrate fall between the fresh-squeezed juice samples, not pasteurized, and the juices reconstituted from concentrate.

Since the EinSight program is not capable of selecting which of the 19 components are most important in differentiating samples, several trial-and-error methods were tried to eliminate variables and yet retain or improve the separation among the four categories. The least
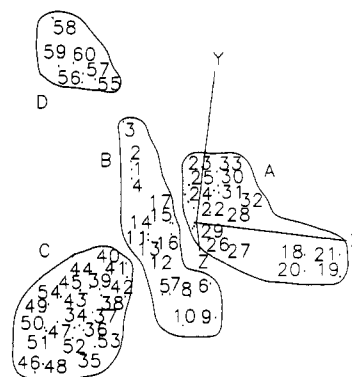
significant compound in each of the first eigenvectors was selected from the loadings report shown in Table IV. The loadings report lists the 19 variables in decreasing order of importance for that eigenvector. Elimination of the quantitative data on these two volatile compounds (2-methyl-1-propanol and methyl butyrate) afforded no improvement in separation of the four categories.

Other attempts at retaining or improving separation of the four categories resulted in less separation. These attempts included elimination of values for the following: ethanol; 2-methyl-1-propanol; 2-methyl-1-propanol and sabinene; 2-methyl-1-propanol, methyl butyrate, and valencene; 2-methyl-1-propanol, methyl butyrate, α-terpineol, and valencene; sabinene, hexanol, and *trans*-2-hexenol; 2-methyl-1-propanol, α-terpineol, valencene, linalool, and γ-terpinene; and finally α-pinene, sabinene, γ-terpinene, and linalool. These compounds either were undetected in many of juice samples (sabinene, hexanol, and *trans*-2-hexenal), showed low levels of importance in the loadings reports for eigenvectors 1 and 2, or were largely contributed by peel oil which might vary considerably because of processing conditions (α-pinene, sabinene, γ-terpinene, and linalool).

A three-dimensional plot of eigenvectors 1-3 could be displayed with EinSight, also. Eigenvectors 1-3 represent 58% of the total variance (Table III). As shown in Figure 2, the separation of classes was no better than with eigenvectors 1 and 2 in Figure 1. The effect of the third dimension (z axis for eigenvector 3) was not readily apparent, either on the computer monitor or in this two-
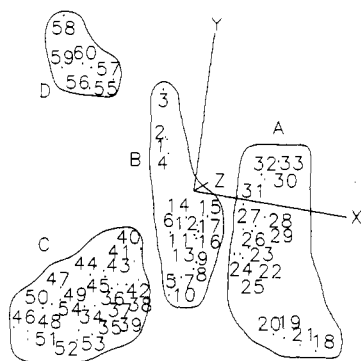
Figure 3. Three-dimensional plot of scores for eigenvectors 1-3 from 16 compounds to categorize processed orange juice samples using categories as listed in Figure 1.
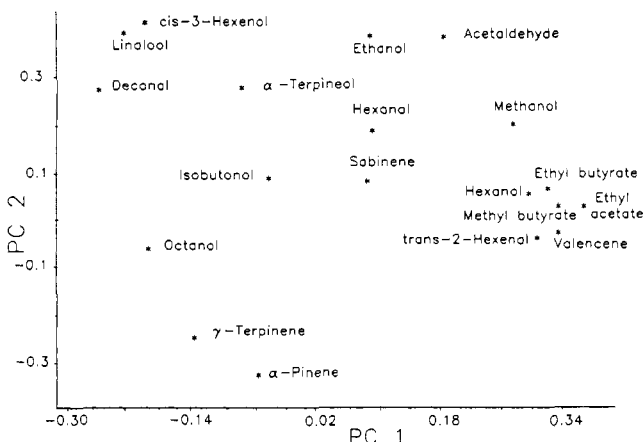

Figure 4. Two-dimensional plot of loadings for eigenvector 1 (PC 1) vs eigenvector 2 (PC 2) for 19 compounds showing correlations between variables and principal components.

dimensional printout. The EinSight program does not include a three-dimensional printout.

No improvement in separation of the four categories was achieved by addition or elimination of the same series of variables described above for eigenvectors 1 and 2, except for one series. Thus, elimination of 2-methyl-1-propanol, methyl butyrate, and valencene afforded improvement in class separation (Figure 3) by visual inspection. These were the three variables ranked lowest in eigenvectors 1-3 as shown in Table III.

A graphic display of loadings for principal components 1 and 2 is shown in Figure 4. This plot shows that methyl butyrate, ethyl butyrate, and ethyl acetate score high on the first principal component and are very close together, indicating they provide similar information (Massart et al., 1988). This is in keeping with the belief that these volatile esters all contribute to the desirable fresh "fruity" note in good quality orange juice (Shaw, 1991). Similarly, cis-3-hexenol and linalool rate high on the second principal component, are very close together on the plot, and are both known to contribute fresh top notes to fruity and flowery aromas (Bauer et al., 1990; Arctander, 1969). It seems easier to visualize such correlations in the plot (Figure 4) than in Table IV, although the same basic information is contained in both.

Many of the compounds ranked highest in loadings 1 and 2 are known to be important to fresh flavor top notes in good quality juice (Schreier et al., 1977; Shaw, 1991; Nisperos-Carriedo and Shaw, 1990). In addition to the three esters discussed above, these include acetaldehyde and ethanol. Relatively lower amounts of these compounds in more highly processed juice such as those

reconstituted from concentrate would help explain observed differences in juice categories.

This study shows that a relatively simple headspace GC technique has potential for classifying commercial freshsqueezed not pasteurized, pasteurized not from concentrate, and reconstituted orange juices from concentrate by computerized pattern recognition techniques. Aseptically packaged reconstituted juice from concentrate may have a different and unique profile of flavor components, as well. Oil compounds, as well as juice compounds, make important contributions to this classification. In limited trials to exclude specific compounds, at least 16 of the 19 compounds quantified by headspace GC were necessary for adequate classification of all samples.

LITERATURE CITED

Anonymous. Sophisiticated listing delineates flavor component reactions. Food Process. 1983, 44 (7), 30–32.

Arctander, S. Perfume and Flavor Chemicals; Arctander: Montclair, NJ, 1969; Vol. I and II.

Attaway, J. A.; Carter, R. D. Some new analytical indicators of processed orange juice quality. Proc. Fla. State Hortic. Soc. 1971, 84, 200–205.

Attaway, J. A.; Barron, R. W.; Blair, J. C.; Buslig, B. S.; Carter, R. D.; Dougherty, M. H.; Fellers, P. J.; Fisher, J. F.; Hill, E. C.; Huggart, R. L.; Petrus, D. R.; Ting, S. V.; Rouse, A. H. Some new analytical indicators of processed orange juice quality, 1971–72. Proc. Fla. State Hortic. Soc. 1972, 85, 192–203.

Bauer, K.; Garbe, D.; Surburg, H. Common Fragrance and Flavor Materials, 2nd ed.; VCH Publishers: Cambridge, England, 1990; pp 9, 24.

Bayer, S.; McHard, J. A.; Winefornder, J. D. Determination of the geographical origins of frozen concentrated orange juice via pattern recognition. J. Agric. Food Chem. 1980, 28, 1306–1307.

Carter, R. D.; Buslig, B. S.; Cornell, J. A. Statistical inferences on processed orange juice quality from the 1972–73 and 1973–74 juice definition program. Proc. Fla. State Hortic. Soc. 1975, 88, 358–370.

Fellers, P. J.; Buslig, B. S. Correlation of objective measurements with subjective indices of quality and consumer preference. "Twenty-second Annual Citrus Processors Meeting"; Agricultural Research and Education Center, Lake Alfred, FL; AREC-LA Mimeo Report, 1971; 71–40, p 14.

Fellers, P. J.; Buslig, B. S. Prediction of orange juice flavor using a multiple regression equation. "Twenty-third Annual Citrus Processors Meeting"; Agricultural Research and Education Center, Lake Alfred, FL; AREC-LA Mimeo Report, 1972; 72–23, pp 9–10.

Fellers P. J.; Buslig, B. S. Objective prediction of FCOJ flavor—a two year regression analysis. "Twenty-fourth Annual Citrus Processors Meeting"; Agricultural Research and Education Center, Lake Alfred, FL; AREC-LA Mimeo Report, 1973; CS-1973-10, p 9.

Florida Department of Citrus. Official rules affecting the Florida citrus industry; Florida Department of Citrus: Lakeland, FL, 1992; Chapter 20-64.0081, Florida Administrative Code.

Massart, D. L.; Vandeginste, B. G. M.; Deming, S. N.; Michotte, Y.; Kaufman, L.Chemometrics—a textbook; Data Handling in Science and Technology; Elsevier: New York, 1988; Vol. 2, Chapter 20.

Mayfield, H. T.; Bertsch, W.; Mar, T. Application of chemometrics to the classification of orange essence oil varieties by GLC. J. High Resolut. Chromatogr. Chromatogr. Commun. 1986, 9 (2), 78–83.

Nagy, S.; Lin, J. C. C.; Klim, M. Application of pattern recognition techniques to sensory and gas chromatographic flavor profiles of natural orange aroma. Food Chem. 1993, in press.

Nikdel, S.; Fishback, V. Chemometric sleuthing: is it really orange juice? Sci. Comput. Automat. 1989, No. 8, 19–23.

Nikdel, S.; Nagy, S.; Attaway, J. A. Trace metals: defining geographical origin and detecting adulteration of orange juice. In *Adulteration of Fruit Juice Beverages*; Nagy, S., Attaway, J. A., Rhodes, M. E., Eds.; Dekker: New York, 1988; pp 81–105.

Nisperos-Carriedo, M. O.; Shaw, P. E. Comparison of volatile flavor components in fresh and processed orange juices. *J. Agric. Food Chem.* **1990**, *38*, 1048–1052.

Page, S. W. Pattern recognition methods for the determination of food composition. *Food Technol.* **1986**, *40* (11), 104–109.

Rouseff, R.; Nagy, S. A multivariate pattern recognition approach for identifying quality factors in Florida orange juice. In *Flavor Science and Technology*; Martens, M., Dalen, G. A., Russwurm, H., Jr., Eds.; Wiley: New York, 1987; pp 481–488.

Schreier, P.; Drawert, F.; Junker, A.; Mick, W. The quantitative composition of natural and technologically changes aromas of plants. II. Aroma compounds in oranges and their changes during juice processing. *Z. Lebensm. Unters. Forsch.* **1977**, *164*, 188–193.

Shaw, P. E. Fruits II. In *Volatile Compounds in Foods and Beverages*; Maarse, H., Ed.; Dekker: New York, 1991; pp 305–328.