the triterpenoids in almond hulls are important in considering the resistance of different almond varieties to the navel orangeworm.

LITERATURE CITED

Beuchat, L. R., Worthington, R. E., J. Food Technol. 13, 355 (1978).

Curtis, R. K., Barnes, M. M., J. Econ. Entomol. 70, 395 (1977).

Kimoto, W. I., Gaddis, A. M., J. Am. Oil Chem. Soc. 51, 307 (1974).

Noller, C. R. "Chemistry of Organic Compounds", W. B. Saunders Co., Philadelphia, (1957), p 255.

Schultz, T. H., Flath, R. A., Mon, T. R., J. Agric. Food Chem. 19, 1060 (1971).

Soderstrom, E. L., USDA, Fresno, CA, unpublished work, 1978.

Steiner, L. F., J. Econ. Entomol. 45, 241 (1952).

Takei, Y., Shimada, K., Watanabe, S., Yamanishi, T., Agric. Biol. Chem. (Jpn.) 38, 645 (1974).

Takei, Y., Yamanishi, T., Agric. Biol. Chem. (Jpn.) 38, 2329 (1974).

Teranishi, R., Lundin, R. E., McFadden, W. H., Mon, T. R., Schultz, T. H., Stevens, K. L., Wasserman, J., J. Agric. Food Chem. 14, 447 (1966).

Warthen, J. D. Science and Education Administration, Agricultural Reviews and Manuals, Northeastern Series, No. 4, April 1979.

Pattern Recognition Analysis of Gas Chromatographic Data. Geographic Classification of Wines of *Vitis vinifera* cv. Pinot Noir from France and the United States

Wing-On Kwan and Bruce R. Kowalski*

Organic compositions of 40 wines of *Vitis vinifera* cv. Pinot Noir from France and the United States were analyzed by glass capillary gas chromatography. Classifications of these wines according to their geographic origins were achieved by applying pattern recognition techniques to the chemical data. Key components for the classifications were then identified by gas chromatography–mass spectrometry. Two neutral compounds, 1-hexanol and cyclohexane, were found to be most important for the classification of French and American Pinot Noirs, while *p*-hydroxybenzaldehyde and 2-phenylethanol were major components for distinguishing California Pinot Noirs from those produced in the Pacific Northwest region.

The use of conventional gas chromatography has allowed the identification of many volatile components in wine (Amerine, 1954; Webb and Muller, 1972). Latest advances in glass capillary gas chromatography enable even finer separation of components in such a complex mixture, but to identify every volatile component can be a tiresome, and in some cases, unnecessary task. The best approach is to identify only those components which are related to the specific properties of wine and then make the identification. Before this can be achieved, relevant information must be ferreted out from a massive amount of gas chromatographic data. Pattern recognition techniques, which have been proven successful in enological research (Kwan and Kowalski, 1978; Kwan et al., 1979), are suited to this type of multivariate analysis.

Volatile constituents of many varieties of *Vitis vinifera* have been studied (Webb and Noble, 1976; Sakato et al., 1975; Brander, 1974; Webb et al., 1969; Van Wyk et al., 1967), but rarely were attempts made to correlate these organic profiles to various properties of the wine. In this study, volatile constituents of wines of *Vitis vinifera* cv. Pinot Noir from France and the United States were analyzed by glass capillary gas chromatography. Pattern recognition techniques were then used to analyze these data. Key components which related to the geographic origins of the wine samples were selected. Identification of these key components were achieved with the use of gas

chromatography–mass spectrometry.

EXPERIMENTAL SECTION

Description of the 40 wines of *Vitis vinifera* cv. Pinot Noir from France and the United States is given in Table I.

Organic Component Extraction. Two hundred milliliters of Pinot Noir wine was extracted with 500 mL of methylene chloride in a separatory funnel. The emulsion which formed upon vigorous agitation of wine and solvent was broken down by centrifugation. Free acids in the wine extract were separated by extraction with 100 mL of 5% Na$_2$CO$_3$ solution. The aqueous phase was back-extracted with 100 mL of methylene chloride to recover nonacidic components. The back extract was combined with the bulk of the "neutral" fraction and dried over anhydrous Na$_2$SO$_4$. The combined neutral fraction was then concentrated to 2 mL by rotary evaporation at 0 °C and stored in a vial for gas chromatographic analysis.

The pH of the Na$_2$CO$_3$ solution containing organic acids was adjusted to a value of 1.0 by adding 6 N H$_2$SO$_4$. The organic acids were then separated from the aqueous solution by extracting with 200 mL of diethyl ether. The ether phase was dried over anhydrous Na$_2$SO$_4$ and concentrated by rotary evaporation to a volume of 2 mL at 0 °C, which was then stored in a vial for future gas chromatographic analysis.

Gas Chromatographic Analysis. A Hewlett-Packard 5840A gas chromatograph with a flame ionization detector was used. The instrument was fitted with a 30-m SE-30 glass capillary column having an inner diameter of 0.25

Laboratory for Chemometrics, Chemistry Department, University of Washington, Seattle, Washington 98195.

Pattern Recognition Analysis of GC Data

J. Agric. Food Chem., Vol. 28, No. 2, 1980 **357**

Table I. Description of 40 Wine Samples

French Pinot Noir

Drouhin, Beaune-Greves	1972[a]
Drouhin, Volnay-Clos des Chenes	1972
Gevrey-Chambertin, Close du Chaptire Brocard	1972
Givry, Domain Baron Therard	1969
	1970
	1971
	1972
Grand Eschezeaux (Domain Romanee Conti)	1962
	1971
Jaffelin Bourgogne du Chaptire	1972
Ponnelle, Chambolle-Musigny	1972
Ponnelle, Chassagne-Montrachet	1972

American Pinot Noir

I. California	
David Bruce (lot 1)	1973[b]
Freemark Abbey	1970
Lords & Elwood	1973
Louis Martini	1973
Robert Mondavi	1973[b]
Simi	1973
Stonegate	1975
II. Pacific Northwest	
Washington	
Associate Vintners	1966-
	1969
	1970[c]
	1971-
	1974
Ste Michelle	1968
Oregon	
Erath	1975
Eyrie	1974
Ponzi	1975
Salishan	1974

[a] Three duplicate bottles. [b] Two duplicate bottles. [c] Four duplicate bottles.

mm. The gas chromatograph was operated in a splitless mode. Sample injection was accomplished with a Hamilton syringe, and injection volume was 3 μL for each sample. Detector and injector temperatures were set at 350 and 150 °C, respectively. Oven temperature was set at 40 °C and maintained for 5 min before it was programmed linearly to 275 °C at 10 °C/min. For each wine sample, two gas chromatograms were obtained. One was the separation of the acidic components and the other, neutral components. Gas chromatographic output included a gas chromatogram, peak retention times, integrated peak areas, and percentage peak areas, which were ratios between individual peak areas and total integrated peak area.

The ether extract, containing organic acids, was silylated by reacting with 1 mL of silylating reagent BSTFA + 1% TMCS [bis(trimethylsilyl)trifluoroacetamide + 1% trimethylchlorosilane] (Regis Chemical Co., Morton Grove, IL) at 80 °C for 5 min. The organic acids would then be converted into their appropriate esters which were volatile enough to be analyzed by gas chromatography.

Pattern Recognition Analysis. Using peak retention times as references, only gas chromatographic peaks which were present in all wine samples, were chosen. A total of 137 peaks were selected for each wine sample. These included 47 peaks from the gas chromatogram of the organic acids and 90 from that of neutral components. Each wine sample was, therefore, considered as an assembly of variables; each peak selected was a variable and its percentage integrated peak area was the magnitude of the particular variable. These 137 variables were called "features", and together they formed a "data vector" which represented a wine sample. When these data vectors,

belonging to the same group such as geographic origin, were analyzed, the group was termed a "category". Pattern recognition techniques used in this study were described in detail by Harper et al. (1977) and Kwan and Kowalski (1978).

Identification of key organic components, selected by pattern recognition techniques for geographic classifications of wine samples, was done commercially by gas chromatography-mass spectrometry (Energy Resources Co., Cambridge, MA). Samples were analyzed by gas chromatography under the same experimental conditions as described earlier. Mass spectral analyses of selected peaks were performed. A computer search of matching mass spectra was done for the identification.

RESULTS AND DISCUSSION

Classification based on geographic origin was performed on the data set. There were 137 features for each data vector. Interpretation of the classification and feature selection results was done with caution as a large number of features was used. If fewer features were used, the results obtained would be more reliable. Usually, it would be ideal to maintain a data vector/feature ratio at a value higher than 3. This could only be achieved in the present case by chosing features which contained most discriminatory information for the classification. SELECT (Kowalski and Bender, 1976) was designed for such a task, and its criterion for selection was the variance weights (Kowalski and Bender, 1972) of the features. Selection of a small number of key features offered another advantage besides increasing the reliability of mathematical classification results. Two-dimensional plots of these key features, one vs. the other, allowed visual examination of the data set.

Separation of wine samples from France and the United States was first performed. The utility of each feature in separating these wine samples was assessed by calculating for each feature the multicategory variance weights which were the geometric averages of the inter-intra category weights. The first feature chosen by SELECT to be most important was peak 108, which belonged to the neutral component portion. It was later identified by gas chromatography-mass spectrometry to be 1-hexanol, from which the remaining 136 features were decorrelated. The decorrelated features were reweighted, and the feature with the highest weight then became the second select feature. It was identified as cyclohexane, which was also one of the neutral components in the wine extract. There were a few features which possessed variance weights higher than cyclohexane, but they were not selected because they contained similar information as 1-hexanol for the classification. When they were decorrelated from 1-hexanol, their variance weights became less than that of cyclohexane. The variance of 1-hexanol was 2.499 which was much higher than 1.440 of cyclohexane. The big difference in variance weight between these two features showed that 1-hexanol contained much more information than cyclohexane for the classification of French and American Pinot Noirs. Features, chosen by SELECT after these two features, had much lower variance weights, indicating that they contained insignificant amount of addition information for the classification of Pinot Noirs from France and the United States.

Three classification methods, LEAST (Harper et al., 1977), SIMCA (Wold, 1976), and KNN (Cover and Hart, 1967), were then applied to the data, using the first two select features, namely 1-hexanol and cyclohexane. Table II shows the classification results of these three methods. LEAST and SIMCA both attained results with only one misclassification out of a total of 40 wine samples. KNN was less successful

358   J. Agric. Food Chem., Vol. 28, No. 2, 1980

Kwan, Kowalski

Table II. SELECT, LEAST, SIMCA, and KNN Results for Geographic Classifications

	US/France	Pacific Northwest/ California
pattern ratio	26:14	17:9
1st SELECT feature	1-hexanol (2.499)[a]	p-hydroxy-benzaldehyde (1.826)
2nd SELECT feature	cyclohexane (1.440)	2-phenylethanol (1.496)
LEAST (two features)	1[b]/40[c]	2/26
SIMCA (two features)	1/40	5/26
KNN (two features)	4/40	6/26

[a] Variance weight.   [b] Misclassification.   [c] Total number of patterns.
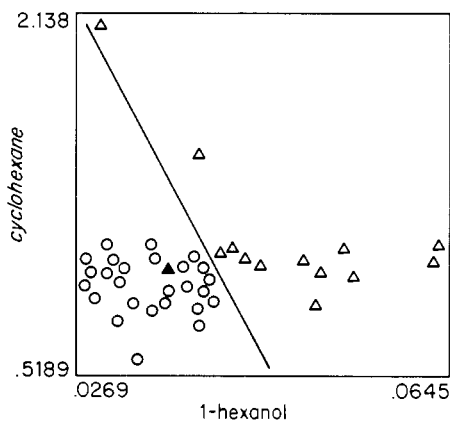
Figure 1. Plot of first SELECT feature, 1-hexanol (peak 108) vs. second SELECT feature, cyclohexane (peak 48) in peak percentage area for classification of French (△) and American (○) Pinot Noirs (blackened symbol = misclassified sample).

with four misclassifications. A large number of additional select features was required before classification results could be improved to 100%. This was due to the fact that features selected after 1-hexanol and cyclohexane contained very little additional information for classification. In the case for LEAST, improvement could only be made on one misclassification out of a total of 40 samples. Therefore, it was concluded that by using just the first two select features, satisfactory results in classification of the wine samples were obtained.

The data were then examined by a two-dimensional plot with normalized peak areas of 1-hexanol versus those of cyclohexane shown in Figure 1. One French sample was misclassified, confirming the results obtained by LEAST and SIMCA. Judging from the extent of scattering of French samples along both axes, the result suggested that greater variations in the levels of both 1-hexanol and cyclohexane were observed in French Pinot Noirs compared to those in American Pinot Noirs.

Classification of samples from two wine regions within the United States was attempted next. Again, SELECT was used to choose important features from the 137 gas chromatographic peaks. The variance weights of the first few select features showed a similar pattern as that observed for the previous classification. Peak 115 and peak 127 were selected as the first two uncorrelated key features. They were identified as p-hydroxybenzaldehyde and 2-phenylethanol with respective variance weights of 1.826 and 1.496. Other select features had variance weights much lower than these two, indicating that they contained relatively insignificant addition information for this classification problem. Table II shows that when the first two select features were used, LEAST attained a result with only two
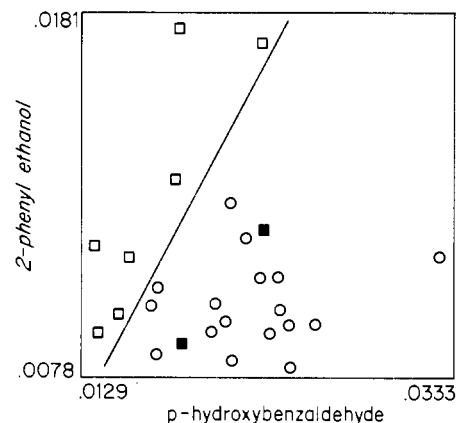

Figure 2. Plot of first SELECT feature, p-hydroxybenzaldehyde (peak 115), vs. second SELECT feature, 2-phenylethanol (peak 127) in peak percentage areas for classification of Pinot Noirs from California (□) and the Pacific Northwest region (○) (blackened symbol = misclassified sample).

misclassifications out of a total of 26 samples. SIMCA and KNN were less successful with five and six misclassifications, respectively. Again, addition of a large number of select features other than the first two was necessary to make any significant improvement on the results.

Classification results using all of the features and SIMCA and KNN for classification produced essentially the same results as those given above. These two methods are less sensitive to the problems associated with low sample-to-measurement ratios.

In Figure 2, the normalized peak areas of p-hydroxybenzaldehyde were plotted against those of 2-phenylethanol to allow examination of the distribution of American samples on this two-dimensional plot. Again, two California samples were misclassified, agreeing with the results obtained from LEAST. The California samples showed greater variations in levels of both p-hydroxybenzaldehyde and 2-phenylethanol compared to those from the Pacific Northwest region.

It was interesting to note that all four key features used in the classifications of American and French Pinot Noirs, and California and Pacific Northwestern Pinot Noirs, were neutral compounds. These four compounds have been isolated and identified in many wine studies (Webb et al., 1967, 1969; Stern et al., 1967). Organic acids seemed to play a lesser role in terms of geographic origin of this set of wine samples.

CONCLUSION

Pattern recognition techniques are able to extract useful information from a massive amount of data. Information was used to relate chemical compositions of wines to their geographic origin. Use of two neutral organic compounds, 1-hexanol and cyclohexane, allowed the classification of Pinot Noirs from France and the United States. Classification of Pinot Noirs from the state of California and the Pacific Northwest region of the United States was made possible by using two features, p-hydroxybenzaldehyde and 2-phenylethanol. Two-dimensional plots of levels of these organic compounds, one vs. the other, were sufficient to solve the classification problem. This approach has been generalized and used to relate chemical composition of wine to other properties such as sensory quality. Results of this study would appear on future publication by our laboratory.

daKoven for valuable observations.

LITERATURE CITED

Amerine, M. A., Adv. Food Res. 5, 353 (1954).
Brander, C. F., Am. J. Enol. Viticult. 25, 13 (1974).
Cover, T. M., Hart, P. E., IEEE Trans. Inform. Theory, 21 (1967).
Harper, A. M., Duewer, D. L., Kowalski, B. R., Fasching, J. L., "ARTHUR and Experimental Data Analysis: The Heuristic Use of a Polyalgorithm", ACS Symposium Ser. No. 52, American Chemical Society, Washington, DC, 1977.
Kowalski, B. R., Bender, C. F., J. Am. Chem. Soc. 94, 5632 (1972).
Kowalski, B. R., Bender, C. F., J. Pattern Recogn. 8, 1 (1976).
Kwan, W. O., Kowalski, B. R., J. Food Sci. 43, 1320 (1978).
Kwan, W. O., Kowalski, B. R., Skogerboe, R. K., J. Agric. Food Chem. 27, 1321 (1979).
Sakato, K. H., Hoekman, M., Kepner, R. E., Webb, A. D., Muller, C. J., Am. J. Enol. Viticult. 26, 70 (1975).
Stern, D. J., Lee, A., McFadden, W. H., Stevens, K. L., J. Agric. Food Chem. 15, 1100 (1967).
Van Wyk, C. J., Webb, A. D., Kepner, R. E., J. Food Sci. 32, 660 (1967).
Webb, A. D., Kepner, R. E., Maggiora, L., Am. J. Enol. Viticult. 18, 190 (1967).
Webb, A. D., Kepner, R. E., Maggiora, L., Am. J. Enol. Viticult. 20, 16 (1969).
Webb, A. D., Muller, C. J., Adv. Appl. Microbiol. 15, 75 (1972).
Webb, A. D., Noble, A. C., Biotech. Bioeng. 18, 939 (1976).
Wold, S., J. Pattern Recogn. 8, 127 (1976).

Determination of Thiolacetates and Some Other Volatile Sulfur Compounds in Alcoholic Beverages

Olavi A. Leppänen,* John Denslow, and Pentti P. Ronkainen

Methyl and ethyl thiolacetates (S-methyl and S-ethyl esters of ethanethioic acid) were identified and determined for the first time in beers and wines. The average content of methyl thiolacetate was 17 µg/L in Finnish medium strength beer and 25 µg/L in Finnish strong beer. The respective average contents for white and red wines were 7 and 11 µg/L. In most samples the concentration of ethyl thiolacetate was about 10% that of methyl thiolacetate. The strong beers contained significantly more methyl and ethyl thiolacetates than the medium beers from the same brewery, but the content of methyl sulfide was not significantly different. Carbon disulfide was identified for the first time in wines, in concentrations of up to 10 µg/L. The contents of thiolacetates were quantified down to 1 µg/L by adsorbing the head-space volatiles of samples on porous polymer. Because thiolacetates can be hydrolyzed to give free thiols, these esters are a potential source of off-flavors in alcoholic beverages despite their own high sensory thresholds.

Direct head-space methods are rapid and simple but are often not sensitive enough for the analysis of low concentrations of sulfur compounds in alcoholic beverages. The methyl sulfide in beer samples has been determined from the head-space down to 2–10 µg/L (Jansen et al., 1971; Anderson and Howard, 1974; Takahashi et al., 1978), but the concentrations of other sulfur compounds in the head-space of beers and wines are commonly near or below the detection limits of the flame photometric detector.

Liquid–liquid extraction has been used to concentrate the less volatile sulfur compounds, notably 3-(methylthio)-1-propanol and 3-(methylthio)propyl acetate, in beers and wines (Schreier et al., 1974a,b; Nakajima and Narziss, 1978).

Pickett et al. (1976) developed a method of trapping head-space volatiles at low temperatures (below –46 °C) under reduced pressure. With this method they identified some volatile sulfur compounds in British lagers and ales and determined the methyl sulfide contents.

Schreier et al. (1976) used on-line enrichment to concentrate the more volatile sulfur compounds of beer. The method involves sweeping the volatiles from the sample in a stream of nitrogen and condensing them in a cold trap connected directly to the chromatographic column. They determined the contents of methyl sulfide and disulfide in several types of beer and found no characteristic differences between them.

Activated carbon (Clark and Cronin, 1975) and porous polymers (Jennings et al., 1972; Bertuccioli and Montedoro, 1974; Williams et al., 1978) have proved suitable for concentrating the volatiles from foodstuffs and alcoholic beverages. An advantage of the polymers is their relative indifference to polar compounds; they selectively adsorb relatively more of the less polar compounds than the more abundant water and alcohols in the head-space of alcoholic beverages.

We here report the content of volatile sulfur compounds, some of which have not previously been identified, in various beers and wines. The volatiles are first adsorbed from the head-space in a small tube of Chromosorb 101 and then desorbed directly onto the chromatographic column.

EXPERIMENTAL SECTION

Apparatus and Materials. A Varian 1400 gas chromatograph fitted with a flame photometric detector (Melpar Inc.) was used for analysis. It was connected to a Hewlett-Packard 3352B laboratory computer, which integrated and named the peaks. A 4 m × 2 mm FEP (fluorinated ethylene propylene copolymer) column packed with 12% DC-200 on Chromosorb W (AWS, 60–80 mesh) was used for the quantitative work. The nitrogen carrier gas flow was 15 mL/min and the oven temperature was programmed from 40 to 150 °C at 8 °C/min. The 120 ×