Exploring Fluorescence Spectra of Apple Juice and Their Connection to Quality Parameters by Chemometrics

Piet Seiden,*,† Rasmus Bro,‡ Leif Poll,‡ and Lars Munck‡

Department of Agricultural Sciences, Royal Veterinary and Agricultural University, Thorvaldsensvej 40, DK 1871 Frederiksberg C, Denmark, and Department of Dairy and Food Science, Royal Veterinary and Agricultural University, Rolighedsvej 30, DK 1958 Frederiksberg C, Denmark

Fluorescence spectra of apple juice were recorded with a view to evaluating the information content in relation to picking date of the apples and possible correlation to traditional harvest indices. The data analysis was performed by using chemometric methods (PCA, PLS, SIMCA). It was shown that the fluorescence spectra correlated with the content of soluble solids in the apple juice and the two apple varieties could be correctly classified by their spectra.

Keywords: *Apple maturity; chemometrics; fluorescence spectroscopy; multivariate analysis*

INTRODUCTION

Methods for determining the optimum harvest time for apples have been a research subject for many years. For a review, see Kingston (1993). Most of the methods are based on measurements of titratable acidity, soluble solids, starch content, internal ethylene concentrations, or textural properties. Much of the work done has focused on obtaining a set of practical guidelines regarding the conditions for optimum yield with respect to the expected storage life of the apples involved (Streif, 1983; Herregods and Goffings, 1993). However, it is often concluded that the available methods are either limited in scope or unable to explain the variance found in overall quality (Douglas, 1983). Among the many factors known to influence apple maturity and hence harvest time are climate, nutrition of the trees, time from anthesis, position of the apple on the tree (effects of neighbors, light/shade, upper/lower branch), and variety. All this seems to indicate that a generalized model for estimating optimum picking time must be based on a very large set of measurements to account for all of these effects. This leads to costly and unwieldy methods that are unlikely to lead to a practical analysis. To counter this situation, one needs a set of screening analyses that can be calibrated to conventional data by multivariate analysis, thereby exploring hitherto unknown correlations which may be exploited for control of picking time and the chemical parameters related to quality and variety. An example of a successful application of this strategy is seen in the widespread use of near-infrared reflectance (NIR) and near-infrared transmittance (NIT) spectroscopy in the food industry (Williams and Norris, 1987). Fluorescence spectroscopy is another method yielding multivariate data with special advantages because of its high degree of specificity and sensitivity. Methods for handling the large number of variables found in data such as fluorescence spectra are now being made available through the evolving discipline of explorative data analysis/chemometrics. In this paper we report on the correlation between fluorescence spectra of apple juice and fruit quality parameters as determined through the use of multivariate statistical methods.

MATERIALS AND METHODS

Apples (*Malus domesticus* L. cv. Jonagold and Elstar) were grown at the University's experimental orchard at Høje Tåstrup. For each variety, 60 apples were picked at random from three trees growing in the same row. Picking was done on the morning of the day of analysis on 7 different days throughout the harvest period: the 23rd and 29th of September and the 4th, 7th, 11th, 14th, and 18th of October.

At the laboratory 20 apples were cut into pieces and juiced one by one in a manual household juicer (Vitamina, Westmark, Germany). Ten milliliters of the juice of each apple was pasteurized by immersing a test tube holding the juice in a 95 °C water bath for 2 min, resulting in a temperature of 90 °C in the tube. The pasteurized juice was stored at −28 °C for less than 2 months prior to recording of the fluorescence spectrum. An aliquot of unpasteurized juice was used for measurement of soluble solids and titrable acids. Soluble solids were measured on a refractometer (RFM 330, Bellingham & Stanley Ltd., England). All measurements were done in duplicate. Titrable acids were determined by titration of the juice with 0.1 N NaOH to an endpoint at pH 8.1 on a Mettler DL 21 titrator.

Fluorescence emission spectra were recorded on a Perkin-Elmer LS 50B luminescence spectrometer. Samples were prepared by filtering the pasteurized juice through glass wool, and the filtrate was diluted 1:10. Each sample was scanned at two excitation wavelengths, 315 (A) and 265 (B) nm, with the highest excitation wavelength first to minimize photodecomposition of the sample. Emission values were recorded in the interval 275−560 nm in 0.5 nm increments. The slit width was 10 nm for both monochromators.

A short qualitative description of the multivariate analytic methods used will be given in the following. For a more detailed description, the reader is referred to the literature (Martens and Næs, 1989; Thomas, 1994). Principal component analysis (PCA) is used for exploring the correlated variations in many variables simultaneously. This decomposition technique replaces the original (many) variables with a lower number of latent variables or principal components. The principal components are constructed so that they essentially describe the same variation as the original variables but in a more condensed way. For example, instead of hundreds of variables as in the case of spectroscopic data, one can visualize and explore two to five latent variables. As only the essential information is kept and noise is reduced, the interpretation and numerical stability are much better than when using the original data. For the classification of apple varieties, a PCA model is estimated for each class. The PCA model defines the

* To whom correspondence should be addressed [fax (+45) 35 28 34 68; e-mail Piet.Seiden@pop.foodsci.kvl.dk].
† Department of Agricultural Sciences.
‡ Department of Dairy and Food Science.

S0021-8561(96)00179-3 CCC: $12.00

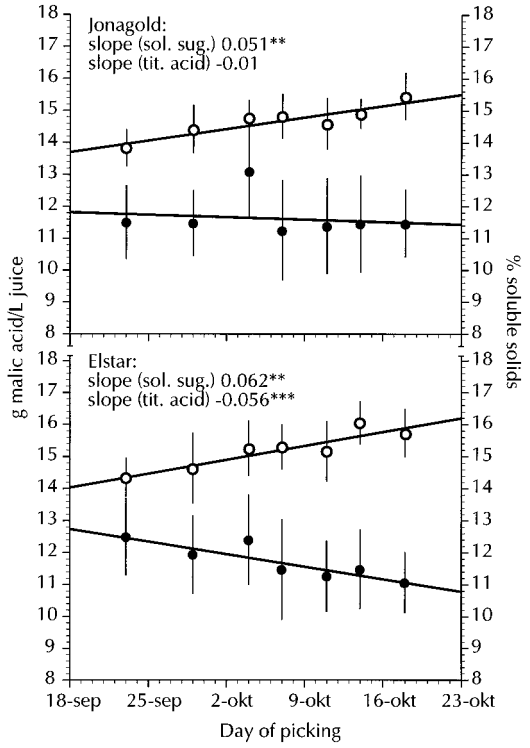

Figure 1. Soluble solids (open symbols) and titrable acid (solid symbols) in apple juice at different picking days. Soluble solids is given as percent sucrose. Titrable acid is given as malic acid equivalents. The measurements are means of 20 samples (apples) per day, with ±1 SD indicated and a regression line fitted to the data. The slopes of the regression lines are shown in the insets. The asterisks denote the significance level for the hypothesis that the lines are nonhorizontal.

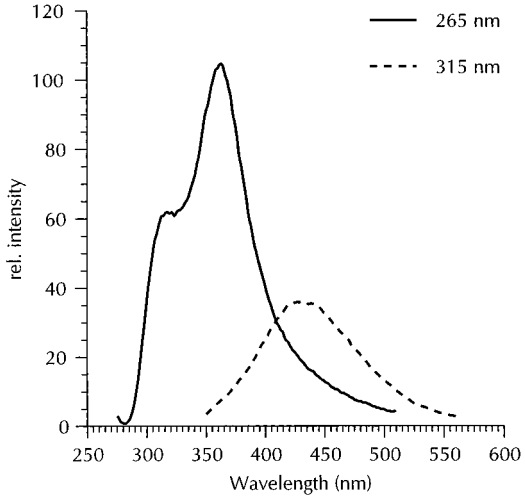

Figure 2. Emission spectra of apple juice at the two excitation wavelengths used. The spectra shown are from Jonagold sample 11 picked on Oct 4, 1994.

location and dispersion of the given class. These PCA models can be used to classify new samples. This is called a SIMCA classification (soft independent modeling of class analogy) (Frank and Lanteri, 1989; Wold et al., 1983). To predict to which class a new sample belongs, the distance of the sample to the different models is calculated, and the model that is closest to the sample constitutes the most likely class. Partial least-squares regression (PLS) is used for regression analysis. This method has several advantageous features. As PCA, PLS also handles correlated variables by decomposing the original variables into a set of orthogonal latent variables, thereby also reducing the noise. The PLS model also gives information as to whether the calibration samples are homogeneous or if outliers are present which reduce the quality of the model. A

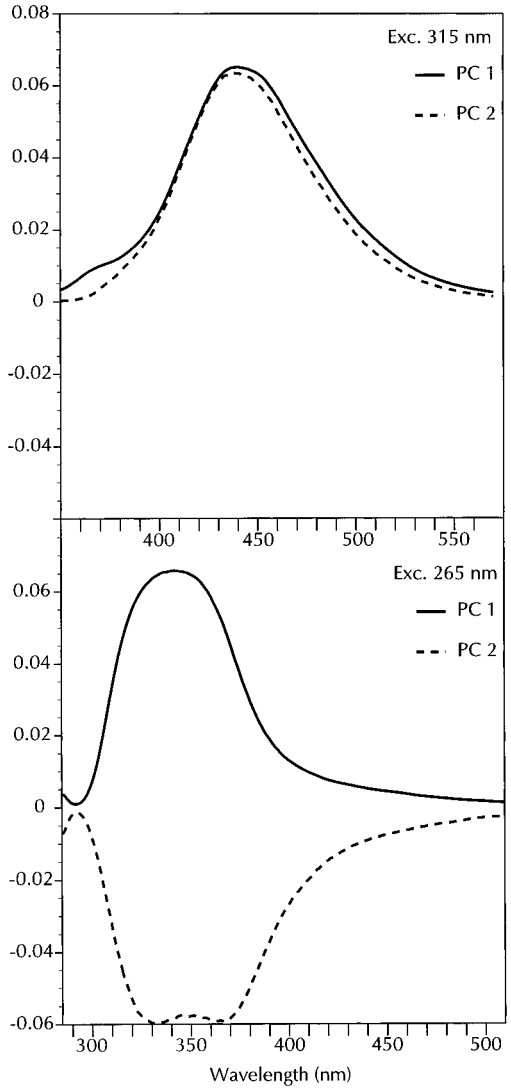

Figure 3. Loading plot of the first two loading vectors from a PCA on the complete fluorescence data set. As the emission spectrum is measured at two different excitation wavelengths, each sample is characterized by two emission spectra. In the PCA analysis these are treated simultaneously as one "spectrum" by concatenating the two spectra. For interpretation purposes the two loading vectors from the two-component PCA model are split into the first part corresponding to excitation at 315 nm (topmost) and the second part corresponding to excitation at 265 nm (bottom). The first loading vector, for example, is the two solid line curves concatenated into one curve.

very important aspect of PLS and other multivariate techniques is that interferences in the samples need not be eliminated, as long as they are also present in the calibration samples.

The multivariate statistical analyses were performed using the software package Unscrambler (Camo A/S, Norway). Variance analysis and trend line estimation were done using SAS 6.10 (SAS Institute Inc., Cary, NC).

RESULTS

Traditional Harvest Indices. The increase in soluble solids and decrease in titratable acidity during the harvest period are shown in Figure 1. For titratable acidity a significant change during the picking period could be observed only for Elstar. Significant changes in the soluble solids fraction occurred throughout the picking period for Elstar as well as for Jonagold. However, the large intraday variation led to difficulties

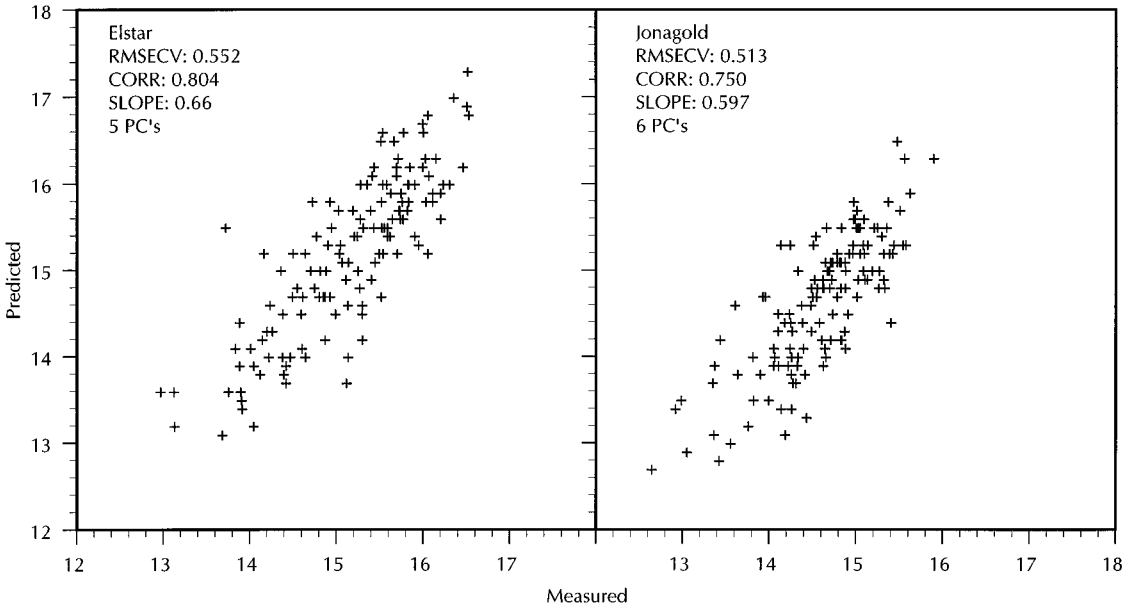

Figure 4. Prediction of soluble solids from fluorescence data. The measured value for soluble solids (in percent sucrose) is plotted against the value predicted by the PLS model based on the fluorescence spectra. Statistics for the model fit and the number of latent variables used in the models are shown above the plots. RMSECV, root mean square error of cross-validation; CORR, correlation between predicted and measured values; SLOPE, slope of regression line; PC, number of latent variables utilized in the model.

in discriminating between the individual days, as can be seen in Figure 1. It was not possible to discern between the varieties on the basis of the soluble solids or titratable acidity data.

Fluorescence Spectra. On the basis of preliminary experiments, apple juice was found to exhibit a fair amount of fluorescence (Figure 2). Excitation wavelengths of 315 and 265 nm were chosen, as they yielded the richest spectra. The obtained spectra were analyzed by PCA after removal of the parts dominated by Rayleigh scattering effects, using the method described by Noergaard (1995). From the loading plot of the two first principal components (Figure 3) the wavelengths responsible for the main variation in the data can be found as the maxima. On the first loading vector the areas around (ex/em) 315/440 nm and 265/350 nm have the highest loadings, while for the second loading vector the area from 330 to 370 nm (ex 265 nm) has high loadings. The loading vectors are mathematical constructs but related to the underlying phenomena (spectra). The areas of high loadings indicate areas where the pure spectra of the analytes producing the fluorescence may have peaks. The bump on the second loading vector is an indication that there are actually two main contributors to the measured fluorescence spectra and that their spectra are highly overlapping.

Correlation to Known Harvest Indices. PLS was used to attempt to develop models between the most significant spectral principal components and soluble solids, titratable acidity, time of picking, and variety. The fluorescence spectra were found to exhibit a positive correlation with soluble solids in both varieties. For each variety a PLS model using mean centered fluorescence spectra as independent variables and soluble solids as dependent variable gave rise to five and six component models, respectively (Figure 4). These models were found by using a four-segmented cross-validated procedure, as a full cross-validation overestimates the predictive power of the model when there are many samples in the calibration set (Esbensen et al., 1994). Two samples from the Elstar extracts and one sample from the Jonagold extracts were eliminated

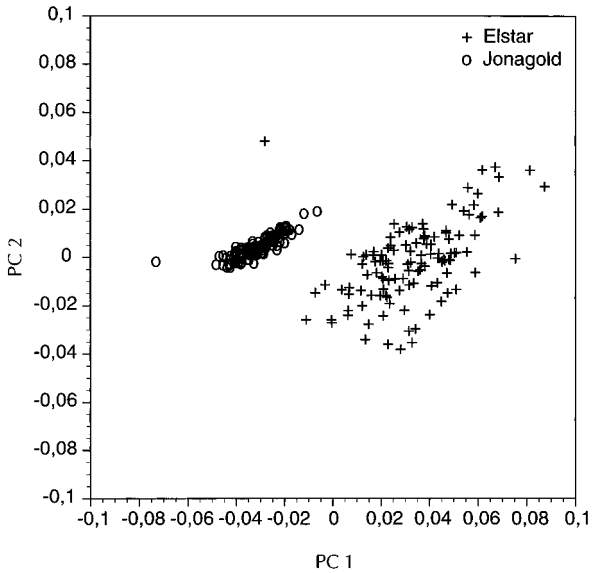

Figure 5. Score plot of the score values of the first two principal components obtained in a PCA on fluorescence data of both varieties.

from the data set in addition to four erroneous samples. This exclusion was not essential for the result; however, there were strong indications that some error was present in their soluble solids values. As the correlation between soluble solids and fluorescence was observed independently in both varieties, it was tested whether any distinction could be made between the two varieties on the basis of fluorescence alone. By developing a SIMCA model on half of the samples, it was possible to correctly classify the remaining half of the samples as either Jonagold or Elstar. In Figure 5 a score plot of the fluorescence spectra shows the separation of the two varieties into two distinct clusters. The variation seen in the fluorescence spectra was found to correlate poorly to the amount of titrable acids in the juice, and no model could be established here. Models derived from the fluorescence spectra were not able to predict the day of picking throughout the harvest period. When supple-

mented with the data for soluble solids and titratable acidity, it was only possible to establish separate models for picking days 1, 2, and 7 for Jonagold.

DISCUSSION

An important use of fluorescence information in relation to fruit maturation is seen in the measurement of delayed light in connection to chlorophyll fluorescence (Forbus and Dull, 1991; Gunasekaran et al., 1990). However, attractive as they are, these methods are not likely to find use in fruits with large color variations in the peel, such as apple. Reports on the study of emission spectra from fruits, fruit juices or extracts of fruit in relation to fruit maturity are scarce. Most concentrate on the attempts to find aging-induced pigments, lipofuscins, in apples (Knee, 1982), avocado (Meir et al., 1991), and bananas (Maguire and Haard, 1976). These studies use a single emission wavelength band in an attempt to distinguish a particular fluorophor. However, Sheehy and Roberts (1991) pointed out that inconsistencies in many lipofuscin studies regarding the emission wavelengths detected raise questions as to the origin of these signals. In this study we have not *a priori* established a specific analyte for which to analyze but instead employed multivariate statistics in an effort to uncover potential signs of correlation between apple fluorophore appearance and the maturation processes occurring in apples during the harvest season.

There are many inherent advantages to using multivariate techniques: (i) All variables can be used simultaneously, so that model selection does not have to be based on preconceived ideas about the data. (ii) A larger number of samples than variables is not necessary, as is the case in traditional statistical methods. (iii) Interference can be incorporated into the mathematical model, possibly avoiding tedious and time-consuming unit operations for cleaning up samples. (iv) Samples not appropriate for the model can automatically be detected as outliers (Beebe and Kowalski, 1987).

It is interesting to note that the fluorescence data enable the distinction between the varieties which could not be shown from the soluble solids or titratable acidity data alone. The soluble solids of apple juice are known to consist mainly of sugars, which in themselves do not exhibit fluorescence. However, it appears that the ripening process involves an increase in soluble, fluorescent compounds. Many of the naturally occurring compounds in apples exhibit fluorescence, e.g. tryptophan, phenolics, and nucleic acids. This makes it difficult to pinpoint the source of fluorescence without further analysis of the spectra, but the wavelength pair observed (315/440 nm) corresponds well to chlorogenic acid (320/430 nm) (Knee, 1982).

The observed correlation of fluorescence spectra with the increase in soluble solids indicates a possibility of modeling the progression in maturity with information obtained from the spectra. The inability to model the day of picking based on the obtained data is assumed to be partly due to the problem of high variability in apples picked on the same day for all variables studied. A model based on a broader set of parameters may be successful in accounting for all of the sources of systematic variation. The fluorescence spectra were found to yield valuable information but could not help to overcome the problem of large interapple variation.

ABBREVIATIONS USED

NIR, near-infrared reflectance spectroscopy. NIT, near-infrared transmittance spectroscopy; PLS, partial least squares regression; PCA; principal component analysis; SIMCA, soft independent modeling of class analogy.

ACKNOWLEDGMENT

The technical assistance of Peter Henriksen and Jette Skaarup Jensen is acknowledged.

LITERATURE CITED

Beebe, K. R.; Kowalski, B. R. An introduction to multivariate calibration and analysis. *Anal. Chem.* **1987**, *57*, 1007A−1017A.

Douglas, J. B. An evaluation of harvest indices for McIntosh apples in two orchards. *HortScience* **1983**, *18*, 216−218.

Esbensen, K.; Midtgaard, T.; Schönkopf, S. *Multivariate Analysis in Practice*; CAMO A/S: Trondheim, Norway, 1994.

Forbus, W. R., Jr.; Dull, G. G. Delayed light emission as an indicator of peach maturity. *J. Food Sci.* **1990**, *55*, 1581−1584.

Frank, I. E.; Lanteri, S. Classification models: discriminant analysis, SIMCA, CART. *Chemom. Intell. Lab. Syst.* **1989**, *5*, 247−256.

Gunasekaran, S.; Paulsen, M. R.; Shove, G. C. Optical methods for nondestructive quality evaluation of agricultural and biological materials. *J. Agric. Eng. Res.* **1985**, *32*, 209−241.

Herregods, M.; Goffings, G. Variability in maturity, quality and storage ability of Jonagold apples on a tree. *Acta Hortic.* **1993**, *326*, 59−64.

Kingston, C. M. Maturity indices for apple and pear. *Hortic. Rev.* **1993**, *15*, 407−432.

Knee, M. Attempted isolation of fluorescent ageing pigments from apple fruits. *J. Sci. Food Agric.* **1982**, *33*, 209−212.

Maguire, Y. P.; Haard, N. F. Isolation of lipofuscin-like fluorescent products from ripening banana fruit. *J. Food Sci.* **1976**, *41*, 551−554.

Martens, H.; Næs, T. *Multivariate Calibration;* Wiley: Chichester, U.K., 1989.

Meir, S.; Philosoph, H. S.; Zauberman, G.; Fuchs, Y.; Akerman, M.; Aharoni, N. Increased formation of fluorescent lipid peroxidation products in avocado peels precedes other signs of ripening. *J. Am. Soc. Hortic. Sci.* **1991**, *116*, 823−826.

Noergaard, L. A multivariate chemometric approach to fluorescence spectroscopy. *Talanta* **1995**, *42*, 1305−1324.

Sheehy, M. R. J.; Roberts, B. E. An alternative explanation for anomalies in "soluble lipofuscin" fluorescence data from insects, crustaceans, and other aquatic species. *Exp. Gerontol.* **1991**, *26*, 495−509.

Streif, J. Experiences with ripening tests for apples. *Acta Hortic.* **1983**, *138*, 63−67.

Thomas, E. V. A primer on multivariate calibration. *Anal. Chem.* **1994**, *66*, 795A−803A.

Williams, P.; Norris, K. *Near-Infrared Technology in the Agricultural and Food Industries;* American Association of Cereal Chemists: St. Paul, MN, 1987.

Wold, S.; Albano, C.; Dunn III, W. J.; Esbensen, K.; Hellberg, S.; Johansson, E.; Sjöström, M. Pattern recognition: finding and using regularities in multivariate data, I. In *Food Research and Data Analysis*; Martens, H., Russwurm, H., Eds.; Applied Science Publishers: London, 1983.

Received for review March 19, 1996. Revised manuscript received July 12, 1996. Accepted August 6, 1996.[⊗] The investigation was supported by grants from The Research Council for Agriculture and The Research Council for Technology to L.M.

JF960179Q

[⊗] Abstract published in *Advance ACS Abstracts,* September 15, 1996.