The application of near infrared spectroscopy to the identification of fruit pulps

I.J. Wesley and A.E.J. McGill

Department of Agriculture and Food Studies, University of Plymouth, Newton Abbot, Devon, TQ12 6NQ, UK.

Introduction

Fruit pulps are used in a variety of products including jams, fruit flavoured yoghurts and ice creams. Pure, single fruit pulps command a high price in the market place, whilst mixed fruit pulps have a low financial value. At present, the only method for authenticating fruit pulps is time-consuming chemical analysis. A fast screening method for authenticating fruit pulps would be advantageous to both importers and end-users.

Near infrared (NIR) spectroscopy has been used for many years in the analysis of agricultural and food products. There is a considerable amount of published material on the analysis of fruit juices by near infrared spectroscopy, but little on the analysis of fruit pulps and purees. A recent publication by Defernez *et al.*[1] discusses the use of Fourier transform mid-infrared spectroscopy and chemometrics for the identification of pure fruit purees.

Experimental

Samples of fruit pulps were obtained from Fen Fruits Ltd, a contract pulping company. The following pulps were obtained—English Strawberry, Spanish Strawberry, Ukrainian Strawberry, English Yellow Egg Plum, English Burbank Plum, Rhubarb, Apple and Raspberry. The pulps were preserved in 0.1% sulphur dioxide solution and were stored in a refrigerator. A set of samples of different dilutions was prepared from each pulp and their NIR spectra recorded. In total, 159 pure samples were prepared. The dry matter of each sample was calculated by measuring the weight loss due to drying in an oven thermostatted at 65°C. These values were used to develop a calibration for dry matter content across a wide range of pure fruit pulps of different dry matters. Typical NIR spectra of different fruit pulps are shown in Figure 1.

NIR spectra were recorded using a NIRSystems 6500 extended range NIR spectrometer fitted with a sample transport accessory. A standard 10 mm quartz cuvette was used to hold the sample and the spectra were recorded in reflection over the range 400–2400 nm. Each sample was scanned until three scans with an RMS difference of less than 5000 had been obtained. These were averaged to produce the final spectrum of each sample. The sample was stirred gently between each scan.

Calibration for dry matter

A calibration for dry matter was developed using ISIS Near Infrared Software (Infrasoft International). Thirty spectra were extracted from the data set and kept aside as a validation set. The remaining 129 samples (the calibration set) were used to develop the calibration. A second derivative function was applied to the spectra using a 12 point gap and a 10 point smoothing

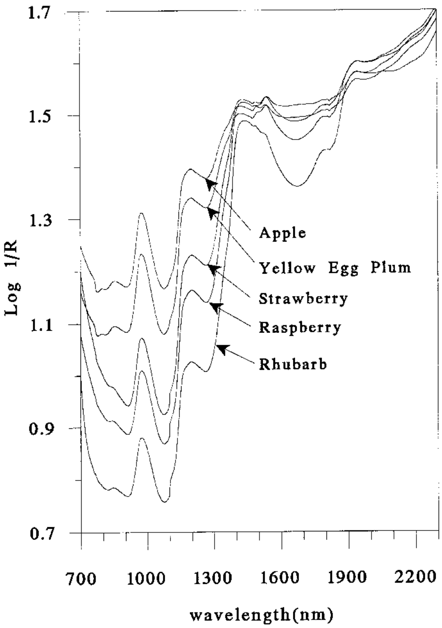

Figure 1. Near infrared spectra of pure fruit pulps.

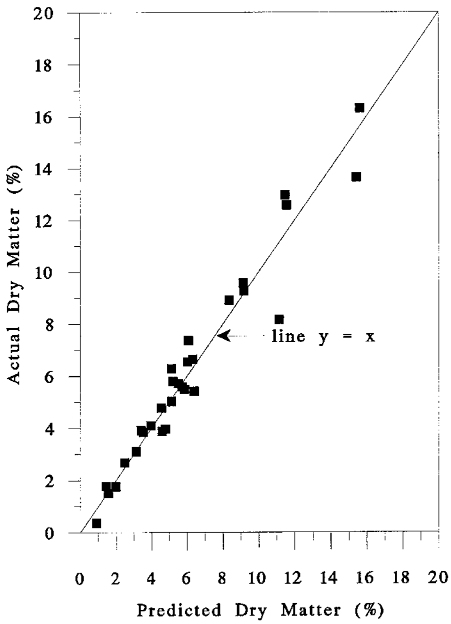

Figure 2. Actual dry matter vs predicted dry matter for external validation set of fruit pulps.

function. The second derivative function partially removes the effects of scattering due to different particle sizes. The spectral range used was 700–2400 nm at 10 mm intervals. Ten cross-validations were used and outliers were eliminated after each cross-validation cycle. A total of 116 spectra were used with a modified partial least squares regression to develop the final equation. The equation was validated using the 30 spectra extracted from the data set before calibration. The results of the cross-validation and the external validation are summarised in Table 1. Figure 2 illustrates the performance of the calibration on the external validation set.

Identification of fruit pulps

A model to authenticate fruit pulps was developed using Win-Discrim Software (K. Kemsley, Institute for Food Research, Norwich). There are two types of analysis possible—Discriminant Analysis and Soft Independent Modelling of Class Analogies (SIMCA). Both techniques rely on reducing the spectral data to a set of variables which explain the variance in the data—a technique known as principal components analysis (PCA). It is comparatively easy to develop a discriminant analysis model which can identify fruit pulps. However, this type of analysis relies on any unknown sample belonging to one of the groups in the original model. SIMCA analysis works on the principle of modelling the similarities within a group and is thus potentially much more useful as a screening technique. When this type of model is applied to an unknown sample, it simply tests whether the sample belongs to the group or not. All of the pure strawberry spectra (64 spectra) were extracted from the calibration set and used to develop a SIMCA model for the identification of pure strawberry. The spectra were first transformed using a second derivative function as before to remove the effects of scattering and then the first 15 principal components were calculated using

Table 1. Summary of cross-validation results.

	Cross-validation results	External validation results
Number of spectra	116	30
Mean of lab values	5.92	6.23
Standard deviation of lab values	3.39	3.83
Variance explained by model (R^2)	0.97	0.95
Slope of line of best fit	1.00	0.98
Standard error of cross-validation	0.56	N/A
Standard error of prediction	N/A	0.88

a covariance matrix method. Models were developed using both an F-Test on residuals method and a squared Mahalanobis distance method and tested using the validation set. An additional set of spectra were obtained from samples of strawberry puree adulterated with other fruit purees. This set of spectra was also used to test the models. The best model, obtained using squared Mahalanobis distance method using nine principal components, is shown in Figure 3.

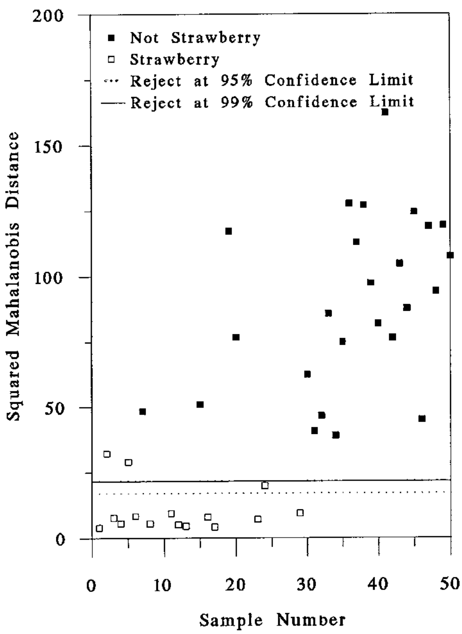

Figure 3. Prediction of fruit type. Samples below the reject limits are predicted as strawberry. Samples above the 1% reject line are rejected with a 1% chance that rejection is incorrect. Actual assignment is given by a marker style.

Results and discussion

The results of the calibration for dry matter show that it is clearly possible to predict the dry matter content of a fruit puree by NIR spectroscopy. The type of puree is largely irrelevant, although samples which are not included in the calibration set are not predicted as successfully. Figure 3 shows that it is comparatively easy to develop a model which can separate pure strawberry purees from other fruit purees and mixtures. The squared Mahalanobis distance method predicts the authenticity of the validation set in the majority of cases and there is no indication that the models are overfitting.

Future work will include using more varieties of strawberry across different growing seasons. We will also develop models for other pure fruits such as apple and raspberry.

Acknowledgements

The authors would like to thank the University of Plymouth for funding this project and Mr Nigel Askew for providing samples. I.J.W. would like to thank the International Committee for Near Infrared Spectroscopy (ICNIRS) for financial contributions towards attendance at this conference.

References

1. M. Defernez, E.K. Kemsley and R.H. Wilson, *J. Agric. Food Chem.* **43,** 109 (1995).