# ROBUSTNESS OF NIR

# CALIBRATIONS FOR ASSESSING

# FRUIT QUALITY

by

**JOHN AUSTIN GUTHRIE**

**QDAH, D App Sc (Food Tech), B App Sc (Food Tech), M App Sc**

**A thesis submitted for the degree of Doctor of Philosophy**

**to**

**Faculty of Arts, Health and Sciences.**

**Central Queensland University, Rockhampton.**

**February 2005**

# TABLE OF CONTENTS

# DECLARATION

I hereby declare that the main text in this thesis is an original work and no part of it has been used in the award of another degree.

John Austin Guthrie

Predictive models based on near infra-red spectroscopy for the assessment of fruit internal quality attributes must exhibit a degree of robustness across the parameters of variety, district and time to be of practical use in fruit grading. At the time this thesis was initiated, while there were a number of published reports on the development of near infra-red based calibration models for the assessment of internal quality attributes of intact fruit, there were no reports of the reliability ("robustness") of such models across time, cultivars or growing regions. As existing published reports varied in instrumentation employed, a re-analysis of existing data was not possible.

An instrument platform, based on partial transmittance optics, a halogen light source and a (Zeiss MMS1) detector operating in the short wavelength near infra-red region was developed for use in the assessment of intact fruit. This platform was used to assess populations of macadamia kernels, melons and mandarin fruit for total soluble solids, dry matter and oil concentration. Calibration procedures were optimised and robustness assessed across growing areas, time of harvest, season and variety. In general, global modified partial least squares regression (MPLS) calibration models based on derivatised absorbance data were better than either multiple linear regression or 'local' MPLS models in the prediction of independent validation populations. Robustness was most affected by growing season, relative to the growing district or variety. Various calibration updating procedures were evaluated in terms of calibration robustness. Random selection of samples from the validation population for addition to the calibration population was equivalent to or better than other methods of sample addition (methods based on the Mahalanobis

distance of samples from either the centroid of the population or neighbourhood

samples). In these exercises the global Mahalanobis distance (GH) was calculated

using the scores and loadings from the calibration population on the independent

validation population. In practice, it is recommended that model predictive

performance be monitored in terms of predicted sample GH, with model updating

using as few as 10 samples from the new population undertaken when the average GH

value exceeds 1.0.

# ACKNOWLEDGEMENTS

# Non Invasive Assessment Group

# Central Queensland University



Back row (left to right):          Kerry Walsh, Christo Liebenberg, Rob Long

Middle row (left to right):        John Guthrie, Donna Hobbs

Front row (left to right):         Aaron Slater, Mirta Golic, Phul Subedi

# PUBLICATIONS / PRESENTATIONS

The following publications / presentations were derived from work undertaken during the candidature of this thesis.

## Refereed Journal Articles

Guthrie JA, Wedding BB, Walsh KB (1998) Robustness of NIR calibrations for soluble solids in intact melon and pineapple. *Journal of Near Infrared Spectroscopy* **6**, 259-265.

Guthrie JA, Walsh KB (1999) Influence of environmental and instrumental variables on the non-invasive prediction of Brix in pineapple using near infra-red spectroscopy. *Australian Journal of Experimental Agriculture* **39**, 73-80.

Walsh KB, Guthrie JA, Burney JB (2000) Application of commercially available, low cost, miniaturised NIR spectrometers to the assessment of the sugar content of intact fruit. *Australian Journal of Plant Physiology* **27**, 1175-1186.

Guthrie J, Greensill C, Bowden R, Walsh. K (2004) Assessment of quality defects in macadamia kernels using NIR spectroscopy. *Australian Journal of Agricultural Research* **55**, 471-476.

Guthrie JA, Walsh KB, Reid DJ Liebenberg CJ (2005) Assessment of internal quality attributes of mandarin fruit. 1. NIR calibration model development. *Australian Journal of Agricultural Research* **56**, 405-416.

Guthrie JA, Reid DJ, Walsh KB (2005) Assessment of internal quality attributes of mandarin fruit. 2. NIR calibration model robustness. *Australian Journal of Agricultural Research* **56**, 417-426.

## Consultancies and Reports

Guthrie JA, Walsh KB (1999) Development of non-invasive objective test to measure eating quality of fruit. Final Report to Horticulture Australia Limited (FR548).

Walsh KB, Guthrie JA (2000) Further development of NIRS to measure eating quality of melons. Final Report to Horticulture Australia Limited (VG97109).

Guthrie JA (2002) Detection of insect damage in intact sweet corn cobs using NIR reflectance spectroscopy. Final Report for the project titled 'Insect Management in Sweet Corn' authored by Peter Deuter (2002) and funded by Horticulture Australia Limited (VG97036).

## Conference Publications

Guthrie JA, Walsh KB (1997) Non-invasive assessment of pineapple fruit sweetness by near infra-red spectroscopy. In: Proceedings of the Sensors for Non-destructive Testing Conference, Orlando, Florida, 18-21 February.

Walsh K, Greensill C, Guthrie J, Newman S, Esson C (1997) Near infra-red spectroscopic assessment of fresh fruit - an Australian perspective. Australian Near Infrared Spectroscopy Group Conference, Adelaide, 23-24 April.

Guthrie J, Walsh K, Deeth H (1997) Getting it right before they take a bite. Australian Nurseryman's Fruit Improvement Company Conference, Sydney, 29-30 April.

Guthrie JA, Walsh KB (1997) Objective quality testing. Invited workshop presentation to 3rd National Stonefruits Conference, 'Growing stonefruits for future markets', Swan Hill 22-25 July.

Guthrie JA, Wedding BB, Walsh KB (1997) Robustness of NIR calibrations for soluble solids in intact melon and pineapple. 8th International Conference on Near Infrared Spectroscopy, Essen, Germany, 15-19 September.

Guthrie JA, Walsh KB (1997) Near infra-red technology for melon Brix measurement. Invited presentation to 4th Australian Melon Conference, Bundaberg, 16-18 October.

Walsh KB, Greensill CV, Guthrie JA (1999) Use of an at-line NIR instrument to evaluate robustness of fruit Brix calibrations. 9th International Conference on Near-Infrared Spectroscopy, Verona, Italy, 13-18 June.

Walsh K, Guthrie J (1999) Non-invasive spectroscopy as a tool to measure sink 'status'. International Conference on Assimilate Transport and Partitioning, Newcastle, 15-20 August.

Walsh K, Guthrie J (1998) Rapid near infra-red spectroscopic assessment of fresh fruit.  Australian Near Infrared Spectroscopy Group Conference, Cairns, 20-22 August.

Walsh KB, Guthrie JA (1999) Getting it right before they take the first bite – infrared spectroscopy and other developments. Fresh 99, National Conference and Exhibition for the fresh fruit and vegetable industry, Canberra, 2-5 September.

Guthrie J, Walsh K (2000) Calibration robustness with respect to variety and district in the assessment of soluble solids content of melons. Australian Near Infrared Spectroscopy Group Conference, Hamilton, Victoria, 4-6 April.

Guthrie JA, Walsh KB, (2001) Assessing calibration robustness for intact fruit. Xth International Conference on Near-Infrared Spectroscopy, Kyonjgu, Korea, 11-14 June.

Guthrie JA, Walsh KB (2002) Citrus sorting: a story of sugar and water. 10th Australian Near Infrared Spectroscopy Group Conference and Short Course, Coffs Harbour, NSW, 7-8 May.

Guthrie JA. (2002) Near Infra-red Spectroscopy in Horticulture. Australian Society for Horticultural Science Regional Conference, Brisbane, 12 June.

Walsh K, Golic M, Guthrie J (2002) Guaranteed sweet: NIS and the sorting of fresh fruit. 10th Australian Near Infrared Spectroscopy Group Conference and Short Course, Coffs Harbour, NSW, 7-8 May.

Guthrie JA, Greensill CV, Bowden RP, Walsh KB (2003) Assessment of quality defects in macadamia kernels using NIR spectroscopy. 11th International Conference on Near Infrared Spectroscopy, Cordoba, Spain, 6-11 April.

Guthrie J, Reid D, Walsh K (2004) Optimising math treatments using significance testing. 11th Australian Near Infrared Spectroscopy Group Conference and Short Course. Fremantle, WA, 18-21 April.

# ABBREVIATIONS

| | |
|---|---|
| 1-VR | variance ratio |
| A/D | analogue to digital |
| ANOVA | analysis of variance |
| AOTF | acousto optical tuneable filter |
| Brix | total soluble solids by refractometry |
| CCD | charged coupled device |
| CH | methine group |
| CH2 | methylene group |
| CP | crude protein |
| CV | coefficient of variation |
| D | detected area |
| DM | dry matter |
| DT | detrending |
| E | longitude, east |
| ED | euclidean distance |
| EI | evaluation index |
| eqn. | equation |
| FWHM | full width half maximum height |
| GH | global Mahalanobis distance |
| H | hydrogen |
| HgAr | helium argon |
| InGaAs | indium gallium arsenic |
| log 1/R | absorbance |

| | |
|---|---|
| M | mole per litre |
| MD | Mahalanobis distance |
| MLR | multiple linear regression |
| MMS1 | Zeiss Si PDA spectrometer |
| ms | milliseconds |
| MSC | multiplicative scatter correction |
| N | nitrogen |
| n | number of samples |
| NaOH | sodium hydroxide |
| NH | neighbourhood Mahalanobis distance |
| NIR | near infra-red |
| NIRS | near infra-red spectroscopy |
| nm | nanometres |
| NSAS | near infra-red spectral analysis software |
| OH | hydroxy functional group |
| P | probability |
| PbS | lead sulphide |
| PC | principal components |
| PCA | principal components analysis |
| PCR | principal component regression |
| PDA | photodiode array |
| PLS | partial least squares regression |
| $R^2$ | coefficient of determination |
| REP file | repeatability file |

| | |
|---|---|
| RMSEC | root mean square error of calibration |
| RMSECV | root mean square error of cross validation |
| RMSEP | root mean square error of prediction |
| RMSEP (C) | root mean square error of prediction corrected for bias |
| RPD | ratio performance deviation |
| S | latitude, south |
| S/N | signal to noise ratio |
| SC | scatter correction |
| SD | standard deviation |
| SDR | standard deviation ratio |
| SG | specific gravity |
| Si | silicon |
| SIMCA | soft independent modelling of class analogy |
| SMA | Sub multi assembly |
| SNR | signal to noise ratio |
| SNV | standard normal variance |
| SS | soluble solids |
| SW-NIR | short wave near infra-red |
| TA | titratable acidity |
| TSS | total soluble solids |
| WINISI | Windows infrasoft international chemometric software |

As 3 out of the 4 chapters have been published in the *Australian Journal of Agricultural Research* (CSIRO Publishing), this thesis has used the format of that journal.  Please note that the policy with regard to the use of numerals throughout the text is "Use arabic numerals in the text, except at the start of a sentence."

# CHAPTER CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# APPENDICES

# 1

## THEORY AND LITERATURE REVIEW

### INTRODUCTION

Prediction models for the assessment of fruit internal quality attributes based on near infra-red (Yermiyahu *et al.* 1997) spectra must exhibit a degree of robustness to be of practical use in fruit grading. Robustness may be defined as the ability of the calibration equation to endure and predict with acceptable accuracy across the parameters of variety, district and time, for the attribute of interest. The calibration model, built at past or current time, should allow valid predictions for future spectroscopic measurements (Thomas and Ge 2000). To achieve this aim, the population used to derive the calibration equation must have sufficient variability within the population to best represent all subsequent samples presented to the instrument. However, inclusion of ever increasing sample variability within the calibration population will, at some point, lead to a loss of accuracy. At this point, the development of a second model is warranted.

### The Problem

de Noord (1994) states that there are a number of situations in which a multivariate calibration model may become invalid. Such instances include replacement of an instrument, drift in the instrument response, measurement at a

different temperature, or change in the physical or chemical composition of the samples. These situations involve change in the sample (e.g. band assignment, light scattering and light absorbing characteristics) or changes in the instrument (e.g. wavelength accuracy, photometric accuracy or optical configuration). These changes will set the new samples being scanned spectrally apart from the 'normal' calibration population. At some point these changes will be sufficient to adversely affect the prediction accuracy of the attribute being assessed by near infra-red spectroscopy (NIRS).

Photometric accuracy of a given spectrometer will be influenced by: (a) the signal to noise (S/N) of the detector and associated electronics, (b) the efficiency of the detectors, and (c) the efficiency of the dispersion grating in its ability to disperse the reflected or transmitted light with minimal loss of signal to the detectors. If these changes are not wavelength dependent, the derivatisation of the spectral data should minimise their effect on the calibration model's prediction accuracy.

Where these changes are wavelength dependent, a prediction *bias* will result. The wavelength accuracy of a given spectrometer is commonly affected by mechanical malalignment. In scanning and filter instruments (not linear array instruments), wear in moving parts such as scanning dispersion gratings and filter wheels will affect the instrument wavelength repeatability and efficiency. In diode array instruments, any shift in the optical bench (e.g. thermal expansion of components) will affect wavelength accuracy.

The complex matrix making up the high moisture biological sample (i.e. intact respiring fruit containing simple and complex carbohydrates, lipids, protein and greater than 80% water), can change in both physical and chemical composition, resulting in changes in the spectral characteristics in both the visible and near infra-

2

red (NIR) parts of the spectrum. Such changes can be used to advantage. For example, Kim *et al*. (2000) used pattern recognition models to classify kiwifruit according to various pre-harvest fruit management treatments, based on short wavelength near infra-red (SW-NIR) spectra. However, if the calibration model is developed using a linear multiple variate regression (e.g. partial least squares (PLS) or MLR, (with coefficients for each wavelength), any change in the matrix causing a wavelength shift in band assignments will degrade the prediction capacity of the model.

## Data pre-treatment

Absorbance of reflectance spectra can be influenced by changes in the sample that are not related to the component of interest. For example, scattering of light from the sample surface (specular radiation) can cause a baseline shift, with more effect at longer wavelengths than shorter wavelengths (for this reason the sky looks blue). Pre-treatment of spectral data prior to analysis is undertaken to remove or reduce undesirable variations such as baseline tilt or light scattering within the sample. In diffuse reflectance NIR measurements, significant difference in the spectra may occur due to the non-homogenous distribution of the particles making up the sample. The degree of scattering is not uniform throughout the spectrum being greater at the longer wavelengths and the effect of scattering on a spectrum is apparent as a baseline shift, tilt and sometimes curvature (more so at wavelengths greater than 1,100 nm). Pre-processing techniques to address this problem includes multiplicative scatter correction (MSC), standard normal variance (SNV), detrending (DT) and derivatives.

### *Multiplicative scatter correction*

This technique assumes the wavelength dependency of the light scattering from and within the sample is different from the absorption by the sample, due to the constituent of interest. Multiplicative scatter correction (MSC) involves calculation of a linear regression of the spectral responses in each spectrum against the average spectrum of the population. The mean offset value is then subtracted from the mean spectra, and the result divided by the slope to give the MSC corrected spectrum. Barnes *et al.* (1989) recommends that MSC is only applicable to spectra with a fairly linear response to concentration of the analyte of interest. If the spectra in the calibration population are substantially different from one another due to a wide range of variability in the sample composition, MSC correction will not give the desired result.

Mean Spectrum: $\qquad\qquad\qquad\qquad \overline{A} = \sum_{i=1}^{n} A_{i,j}$

Linear Regression: $\qquad\qquad\qquad\qquad A_i = m_i \overline{A} + b_i$

MSC Correction: $\qquad\qquad\qquad\qquad A_{i(MSC)} = \dfrac{(A_i - b_i)}{m_i}$

Where A is the *n* by *p* matrix of the calibration population spectral responses for all wavelengths, $\overline{A}$ is a *1* by *p* vector of the average responses of all the training population spectra at each wavelength, $A_i$ is a *1* by *p* vector of the responses for a single spectrum in the training population, *n* is the number of training spectra, and *p* is the number of wavelengths in the spectra. The $m_I$ and $b_I$ values are the slope and offset coefficients of the linear regression of the mean spectrum vector $\overline{A}$ versus the $A_j$ spectrum vector.

***Standard normal variance and Detrending***

Standard normal variance and DT are 2 other techniques used to reduce the effect of light scattering on the sample spectrum. Standard normal variance removes scattering by normalising each spectrum in the calibration population by the standard deviation of the absorbances across the entire spectral range of that spectrum.

Mean Response $\qquad \bar{a} = \sum_{j=1}^{p} A_{i,j}$

SNV Correction $\qquad A_{i(SNV)} = \dfrac{\left(A_i - \bar{a_i}\right)}{\sqrt{\dfrac{\sum_{j=1}^{p}\left(A_{i,j} - \bar{a_i}\right)^2}{(p-1)}}}$

Where A is the *n* by *p* matrix of training population spectral responses for all the wavelengths, $A_i$ is a *1* by *p* vector of the responses for a single spectrum in the training population, $\bar{a_i}$ is the average of all the spectral responses in the vector, *n* is the number of training spectra, and *p* is the number of wavelengths in the spectra.

With DT, each spectrum is treated independently of the other spectra in the calibration population. In the calculation, a linear least squares regression is used to fit a quadratic polynomial to the responses in the spectrum (Barnes *et al.* 1989). Then this curve is subtracted from a given spectrum. This quadratic curvature component of the calculation is an attempt to correct the effects of particle size and packing of the sample.

Barnes *et al.* (1989) found that SNV and DT correction of spectra of crystalline sucrose samples achieved superior calibration statistics to those obtained with straight absorbance (log 1/R) and results equal to or better than derivatised (first or second) transformations. Further, derivatisation of SNV and / or DT spectra produced better calibration results than derivatised only spectra, but not significantly better than straight SNV and /or DT transformations.

Generally, SNV is applied first to correct for the effects of interference of scatter and particle size, similar to MSC.  Then DT is used to attempt to remove the additional variations in baseline shift and curve-linearity.

*Derivatives*

The first derivative is simply a measure of the slope of the spectrum for each point in the spectrum, and is 1 of the best methods for removing the baseline shift (Barnes *et al.* 1989).  The second derivative (rate of change of the slope of the spectrum), as well as removing any offset is not affected by any linear tilt in the data.  Most applications of NIR spectroscopy, utilising either diffuse reflectance or transmission modes on intact fruit, use second derivative absorbance data in calibration development (e.g. Guthrie and Walsh (1997b), Kawano *et al.* (1993), Kawano *et al.* (1989), Peiris *et al.* (1998b), Slaughter *et al.* (1996), Lammertyn *et al.* (1998)).  Different chemometric software packages offer different methods of derivative calculation.  The software package WINISI offers a 'Norris' method of derivative calculation whereas The Unscrambler offers both 'Norris' and 'Savitzky-Golay' methods of derivative calculation.

In the 'Savitzky-Golay' method, data points around a central point are fitted to a polynomial function.  A least squares curve fitting routine is combined with differentiation to assess the derivative (or second derivative) at the wavelength ($\lambda$).

For slowly changing functions, derivatives can be approximated by taking the difference in y (absorbance) values between non-adjacent data points.  To reduce noise, the difference of the 2 averages can be taken, formed from points surrounding the selected y values.  To further reduce computations, division by $\Delta x$ can be omitted.  Norris (1982) defined the term "segment" to indicate the length of the x interval over which the y values are averaged, and the term "gap" for the length of the x interval that

separates the 2 segments to be averaged.  This protocol, known as a 'Norris' derivative, is adopted in the WINISI software.

Using the 'Norris' derivative protocol, second derivative of 3 data points at x values (a; b; and c) can be calculated as a – 2b – c.  Using WINISI terminology, a 2:10:6:1 data treatment is computed as (1 + 2 + 3 + 4 + 5 + 6) – 2 * (11 + 12 + 13 + 14 + 15 + 16) + (21 + 22 + 23 + 24 + 25 + 26), where the number in parentheses refers to the absorbance data at that data point number.  The range of data points involved in each treatment is n data = 2 * 10 + 6 + 1 – 1 = 26.  The data point associated with this treated value is (n data + 1)/2 = (26 +1)/2 = 13 (any fractional data point is truncated) (pers. com. Mark Westerhaus 1999).  The gap size (number of data points either side of the data point of interest) used in the derivative calculation, will affect the extent of smoothing of data, and will vary with the instrument used. The setting of the gap is generally optimised by trial and error.  Inherently, it is expected that the gap should be ≥ to the λ resolution of the instrument.  For example, the Foss NIRSystems 6500 records data at 2 nm steps, but has a wavelength resolution of (full width at half maximum height, FWHM) of 10 nm.  Shenk *et al*. (1992) typically recommends a gap of 4 data points for calculation of derivative shifts for this data, entailing averaging over 8 and 16 nm for first and second derivative, respectively.

**Measurement of Calibration Model Performance**

The success of a calibration model in terms of calibration and prediction indices can be judged through the use of the following: the coefficient of determination ($R^2$), root mean square standard error of calibration (RMSEC), the root mean square standard error of prediction (RMSEP), root mean square standard error of cross

validation (RMSECV), variance ratio (1-VR), *bias* (the difference between the mean of actual and predicted values) and the ratio of the standard deviation of the population divided by the root mean square standard error of prediction or the root mean square standard error of cross validation (SDR) (Shenk *et al.* 1992) and (McGlone and Kawano 1998) (Table 1). The term ratio performance deviation (RPD) is similar to SDR (except RPD uses a *bias* corrected RMSEP or RMSECV values).

The key terms are SD and RMSECV for calibration results and SD, RMSEP and *bias* for prediction results. The $R_c^2$ (calibration model coefficient of determination) is a function of SD and RMSEC, and $R_v^2$ (validation coefficient of determination) is a function of SD and RMSEP (when *bias* = 0). Obviously the $R^2$ can be improved by increasing the SD of the population. Evaluation of a model using the $R^2$ statistic should therefore only be considered with knowledge of the SD (which should be equivalent to that of the populations to be predicted).

# Table 1. Definition of chemometric terms.

**The following notation is used:**

The following notation is used:

$n$ = number of samples in the calibration population

$x_i$ = actual reference values of samples in the calibration population $i$; $i = 1, 2 ... n$

$\bar{x}$ = mean of reference values for the calibration population $= \dfrac{1}{n} \sum_{i=1}^{n} x_i$

$y_i$ = predicted sample values $i$; $i = 1, 2 ... n$

$\bar{y}$ = mean predicted value of samples in prediction/validation population $i$, $i = 1, 2 ... n$,

$f$ = number of latent variables (principal components) in the calibration regression model

$m$ = number of samples in the prediction / validation population

$c_i$ = actual reference value of samples in the prediction / validation population $i$; $i = 1, 2 ... m$,

$\hat{c}_i$ = predicted value of samples in the prediction / validation population $i$; $i = 1, 2 ... m$,

| Term | Abbrevn. | Equation | Comments |
|---|---|---|---|
| Standard deviation | SD | $SD = \sqrt{\dfrac{\sum_{i=1}^{n} \left( x_i - \bar{x} \right)^2}{n-1}}$ | Measure of the dispersion of the reference values for the population |
| Bias | *bias* | $bias = \sum_{i=1}^{n} \dfrac{\left( x_i - y_i \right)}{n}$ | Measure of the average difference between reference and predicted values |
| Coefficient of determination | $R^2$ | $R^2 = \dfrac{\text{Regression SS}}{\text{Total SS}}$ <br><br> $= \dfrac{\sum_{i=1}^{n} \left( x_i - \bar{x} \right)\left( y_i - \bar{y} \right)}{\sum_{i=1}^{n} \left( x_i - \bar{x} \right)^2 \sum_{i=1}^{n} \left( y_i - \bar{y} \right)^2}$ <br><br> $= 1 - \left( \dfrac{RMSEC}{SD} \right)^2$ | The fraction of the variance within the predicted values explained by the calibration model |
| Standard error of calibration | RMSEC | $RMSEC = \sqrt{\dfrac{\sum_{i=1}^{n} \left( x_i - y_i \right)^2}{n - f - 1}}$ | Estimate of the variation of the reference and predicted values of the calibration population |
| Standard error of cross-validation | RMSECV | $RMSECV = \sqrt{\dfrac{\sum_{i=1}^{n} \left( x_i - y_i \right)^2}{n-1}}$ | Where $x_i$ and $y_i$ are samples left out of the calibration population |
| Standard error of prediction | RMSEP | $RMSEP = \sqrt{\dfrac{\sum_{i=1}^{m} \left( c_i - \hat{c} \right)^2}{m}}$ | Estimate of the variation of the reference and predicted values of the validation population |
| Standard error of prediction corrected for *bias* also known as SDVD | RMSEP(C) | $RMSEP(C) = \sqrt{\sum_{i-1}^{m} \dfrac{\left( c_i - \hat{c}_i - bias \right)^2}{m-1}}$ | Estimate of the variation of the reference and predicted values of the validation population, corrected for *bias* |
| Variance ratio | 1-VR | $1 - VR = \left( \dfrac{SECV}{SD} \right)^2$ | WINISI software term, same as $R^2$ but RMSECV used, not RMSEC |
| Standard deviation ratio | SDR | $SDR = \dfrac{SD}{RMSECV \text{ or } RMSEP \text{ or } RMSEC}$ | Standardising with either the RMSEP, RMSEC or RMSECV between populations |
| Ratio performance deviation | RPD | $RPD = \dfrac{SD}{RMSEP(C)}$ | As above but *bias* corrected |

McGlone and Kawano (1998) 'rule of thumb' for evaluation of a model considered that a SDR of at least 3.0 (defined with RMSEP) was necessary to grade fruit into 2 classes with an acceptable degree of accuracy. However, this requirement is dependent on the required accuracy of the grading exercise. Certainly, the SDR must exceed 1, otherwise the grading is meaningless.

In the development of a calibration model, the data treatments chosen should minimise the standard error of calibration (RMSEC). An "acceptable" RMSEC is defined as usually from 1 to 2 times the laboratory error (Hruschka 1987). Where cross validation is used, the standard error of cross validation (RMSECV) should be no more than 20% greater than the RMSEC (Shenk *et al.* 1992). Overfitting occurs when a calibration regression approximates non-representative features of the particular samples used for calibration. Evidence of overfitting can be found where (Hruschka 1987):

(i)     The RMSEC is much lower than the laboratory error. This case usually resulting from too many terms in the regression or too few samples.

(ii)    The *bias* is significant compared with the standard deviation of validation differences, also known as RMSEP(C). The value refers to the SD of the variance between the NIR measured and reference values of the samples. The contribution by *bias* to variance can be explained by the fraction of $bias^2$ / $RMSEP(C)^2$. For example, if the RMSEP(C) = 1.0, then a *bias* of 0.40 contributes only 16% of the variance, but a *bias* of 0.80 will contribute 64% of the variance and is significant compared to the RMSEP(C).

(iii)   The *bias* is negligible, but the RMSEP(C) is more than twice the RMSEC. If this case is due to instrumentation issues, rather than the sample matrix, it may be possible to obtain a good NIR measurement from a bad calibration by averaging

the results of several NIR measurements on sub-samples (averaging out sampling error).

Over-fitting of the calibration model to the spectral data and the attribute of interest, may occur within a particular population by the fitting of 'noise' (i.e. to spectral information that does not either directly or indirectly relate spectroscopically to the attribute of interest). The 'rule of thumb' recommended by Shenk *et al*. (1992) was to use approximately 1 wavelength in a multiple linear regression (MLR) model and 1 principal component (PC) in a partial least squares (PLS) model per 10 samples used in the calibration population. Hruschka (1987) suggested the use of a minimum of 5 - 15 samples for each regression variable, data treatment constant, or any parameter of the data treatment (such as wavelength) that is allowed to vary. The number of PC's in a PLS model is generally determined by the minimum root mean square error of cross validation ((Shenk and Westerhaus 1993), (Slaughter *et al.* 1996)), but for large populations (e.g. 2000 samples), the chemometric software package normally sets a default of 16 to 25 PC's (e.g. WINISI ver. 1.04a in MPLS has a default of 16 PC's).

## Population structuring

### *Building in variation*

In order for calibration populations to best predict on unknown samples, the population must be made up of samples covering all aspects of both spectral and wet chemistry data. Therefore procedures are used to structure the calibration population and assess whether this calibration population is capable of predicting on any subsequent unknown population.

Structuring the reference values of the calibration population in a 'boxcar' distribution can be achieved by reducing the number of samples with reference values close to the mean. For example, Guthrie and Walsh (1999) (Appendix A) report that a pineapple fruit total soluble solids model based on a boxcar distribution gave a higher $R_c^2$ than that based on normally distributed data, although prediction results were poorer. Fearn (1992) considers that it is a mistake to discard samples if calibration sample numbers are limited. If there is an ample supply of calibration samples and the constraint is reference analysis (e.g. cost of analysis), then flattening the distribution by selection is a "good idea" (Fearn 1992). This procedure increases the standard deviation of the population but increases the emphasis of the regression on samples at the population's extremities (improving prediction accuracy far from the mean) and removes spectral variance from the population. The calibration population should also contain samples spectrally representative of the population being predicted (as discussed in the next section).

### *Temperature Compensation*

Near infra-red calibration models developed on samples scanned at constant temperature are not reliable in predicting samples at different temperatures (Kawano *et al.* 1995 see also Guthrie and Walsh 1999, Appendix A). For example, intact peaches (100 fruit each in calibration and validation populations, drawn from an original population of 200 fruit) were scanned at 3 different temperatures (21, 26 and 31ºC) and spectra from all temperatures were included in the calibration population. Using this model to predict fruit at each of the 3 temperatures, RMSEP values ranged from 0.39 to 0.465 TSS with *bias* from 0.01 to 0.03% TSS, using a model developed on fruit at 1 temperature only, but without this temperature compensation, RMSEP

values ranged from 0.44 to 0.50% TSS, with *bias* from 0.05 to –0.33% TSS.  The 'repeatability' file (REP file) of WINISI has been used to accommodate sample temperature change (Shenk and Westerhaus 1993).  The repeatability file contains spectral information about a source of variation that is to be minimised in the development of a calibration (e.g. variations due to sample temperatures or the use of different instruments).  In this method the spectra of 1 sample is collected at a range of sample temperatures.  The mean of these samples is then subtracted from each sample in the REP file and assigned a zero value to each constituent value before proceeding with the normal PLS procedure.  This structuring of the calibration population can be based on reference values ('wet' chemistry values), spectral data or a combination of both reference and spectral data.

### *Outlier removal*

Multivariate distance measuring computations can be used to assess the presence of outliers in both the calibration and the prediction populations.  According to de Noord (1994) and Mark (1986), the Mahalanobis distance (D) is preferable to the Euclidean distance (ED) to measure the suitability of samples for inclusion in the calibration population because the D takes sample variability into account.  That is, D weights the differences by the range of variability in the direction of the sample point. The Euclidean distance treats all values equally when calculating the distance from the mean point.  Mahalanobis distances look not only at variations (variance) between the responses at the same wavelengths, but also at the inter-wavelength variations (co-variance).

In practice, using raw absorbance or derivative data for more than 10-15 wavelengths can result in mis-classification of known samples (Mark and Tunnell

1985).  The above problem is mostly overcome by calculating D in principal

component algorithm (PCA) or partial least squares (PLS) space (i.e. calculation of D

from sample scores).  Measuring D in principal component or partial least squares

space usually explains most of the sample variation within a population in less than

fifteen PCA scores.  Scores are the values of factors (also known as principal

components) calculated using either the PCA or PLS algorithms relating to each

sample in the population (usually created from mean-centred absorbance data), and

loading are either PCA or PLS factors relating to the wavelengths used.

The Mahalanobis matrix equation is given as:-

$$M = \left[\frac{S'S}{n-1}\right] \qquad \text{eqn.1}$$

where M is a $f$ by $f$ Mahalanobis matrix, S is the n by $f$ matrix of training samples

PCA/PLS scores, n is the number of samples and $f$ is the PCA/PLS scores (Mark and

Tunnell 1985).  For mean centred score data (used by the chemometric software

package WINISI ver. 1.04a) the above equation becomes:

$$M = \left(\frac{(S_i - \overline{S})'(S_i - \overline{S})}{n-1}\right) \qquad \text{eqn.2}$$

Although it is better to use scores created from mean centred absorbance data,

the equation 2 approach is a calculation simplification.  If mean centred absorbance

data are used, then $M$ from equation 2 should approximate $M$ from equation 1

(different by the residual description of the mean spectrum given by $\overline{S}$ ).

Prediction of Mahalanobis distance for an unknown sample given in the $1*f$

matrix, $S_{unk}$ is given by

$$\mathbf{D}^2 = \left(S_{unk} - \overline{S}\right)M^{-1}\left(S_{unk} - \overline{S}\right)' \qquad \text{eqn.3}$$

Thus, the D value will increase as more factors are used in the calibration

model.  Further the average D value will vary between populations (Mark 1986).  For

example, adding outliers to a combined population will increase the D value. Shenk *et al.* (1992) reported the use of a 'standardized' D value (this was not defined, but was presumably GH of eqn.5).

Shenk *et al.* (1992) therefore recommend the use of a global Mahalanobis distance, where $GH = D^2/f$ referred to as Global H (GH) in WINISI ver. 1.04a for outlier determination.

$$GH = \frac{D^2}{f}$$

$$GH = \left(S_i - \bar{S}\right)\left(\frac{\left(S_i - \bar{S}\right)'\left(S_i - \bar{S}\right)}{n-1}\right)^{-1}\left(S_i - \bar{S}\right)' \qquad \text{eqn.4}$$

A GH unit value is considered to represent a standard deviation from the group mean spectrum (centroid) and therefore samples with a GH of $> 3.0$ have a 0.01 or less probability of belonging to that group.

However, additional extra factors to a model are not expected to linearly increase GH. Thus standardization by the division of the number of factors, while computationally simple (Shenk *et al.* (1992) were working within the constraints of a 640 Kb memory buffer) is not intuitively the best approach. For example, further work might also consider division by the average score of the calibration population (eqn.5).

The extent to which samples from the population are spread through multidimensional space can be calculated as:

$$rmsgroupsize = \sqrt{\frac{\sum^n D^2}{n-1}} \qquad \text{eqn.5}$$

Shenk and Westerhaus (1991) also recommended the use of a neighbourhood Mahalanobis distance value (NH). This value was used to estimate the proximity of each sample to every other sample in the population using Mahalanobis distances calculated as in equations. 2, 3 and 5.

$$NH = (S_i - S_j)(S'S) - 1(S_i - S_j)'$$     eqn.6

This value can be used, together with spectral similarity ($R^2$ of sample spectrum against mean spectrum of the calibration population) and GH (spectral distance from the mean spectrum of the population) to ascertain the ability of the calibration model to predict accurately on an unknown sample and to allow the removal of 'superfluous' spectra from the calibration population. Shenk and Westerhaus (1991) recommended the use of a GH value of 3.0 and a NH value of 0.6 (for the Foss NIRSystems 6500 scanning instrument) for elimination of outliers from a given population. For instruments utilising less data points per spectrum (e.g. PDA or filter instruments), Shenk and Westerhaus (1993) recommended a lower NH value of 0.2.

Another approach to sample outlier detection relies on the calculation of the difference between a spectra reconstructed from the scores and loadings, then the difference between the reconstructed sample and the original sample is the spectral residual. By calculating the sums of squares of the spectral residuals across all wavelengths, an additional value can be generated for each spectrum (Galactic Corporation 1999). Two methods can be used to include these spectral residuals in the matching of sample suitability for inclusion in the calibration population. One combines the D and soft independent modelling of class analogy (SIMCA) test on the spectral residuals to test the validity of the samples matching to the calibration population. The other combines the PCA/PLS scores and spectral residuals for each

spectrum and uses them all for the Mahalanobis group matrix calculations (Galactic Corporation 1999).

The SIMCA method employs principal component analysis of full spectra for the construction of mathematical models for each material to be analysed, retaining an optimal number of PC's (determined by a cross validation technique of optimal recognition and rejection rates in independently measured samples) for each (Gemperline *et al.* 1989). The sum of squares of the residual spectrum (difference between the original and model constructed spectrum) is used to calculate a '*F*' value. The probability level for the corresponding *F* value is used to classify the sample.

$$ F = \frac{S_p^{(q)^2}}{S_o^{(q)^2}} \frac{n_q}{\left(n_q - f - 1\right)} \qquad \text{eqn.7} $$

where $S_p^{(q)^2}$ is the residual variance of spectrum *p* fit to population $q$, $S_o^{(q)^2}$ (where *q* is the calibration population) the variance within class $q$, $n_q$ is the number of spectra used in the training population for population *q, M* is the number of points per spectrum, and *f* is the number of PC's used to model population *q* (Gemperline *et al.* 1989).

The local calibration technique, as patented by (Shenk and Westerhaus 1997), is based on a technique where the unknown spectrum is used to select a default number of similar spectra (e.g. 120 spectra) from a large database (up to 4,000 spectra). A PLS regression equation is then developed on these selected spectra. The developed calibration equation is then used to predict the unknown sample. An increase in prediction accuracy for analysis of dried ground corn and haylage for dry matter, crude protein and acid detergent fibre of approximately 28% (in RMSEP values) is claimed, relative to specific global calibrations developed utilising PLS regression.

# CASE STUDIES OF CALIBRATION ROBUSTNESS

## Robustness across growing districts

Robust models for prediction of the attributes of intact fruit will require the use of calibration populations that exhibit large variations in fruit origin, storage age and size, as well as large variation in the constituent being modelled (McGlone and Kawano 1998). For example, McGlone and Kawano (1998) used 306 fruit of 1 cultivar of kiwifruit sampled from 5 different orchards (both from New Zealand and Japan). The calibration and validation populations were equal in size, and derived from randomised selection of the various data populations. They found that a model (PLS regression on second derivative spectral data in the 800 to 1,100 nm region with 20 nm segment smoothing) based on a combined calibration population gave "very good results" in predicting both Brix (% TSS) and DM, ($R_c^2$ of > 0.90, RMSEP values of < 0.5 units, *bias* of 0.04 and -0.01% TSS, respectively and SDR of 3.1 for both attributes). However, this model was not tested on a subsequent separate population (e.g. next season's fruit from the same districts).

Peirs *et al*. (2003) investigated the robustness of PLS models for prediction of soluble solids in intact apples across orchards, seasons and cultivars. The major variation in spectral data occurred between season (31%) and cultivar (17%) with orchard explaining only a small amount of the variation. However, approximately 39% of the spectral variability was due to "other kinds of compositional variability and noise", not related to the season, cultivar and orchard. Including more variability into the calibration populations improved validation accuracy as long as atypical data was not included, and the number of latent variables in the model development did not

exceed 9.  A number greater than 9 resulted in over-fitting and less subsequent

robustness in terms of validation with external validation populations.  They

recommend the inclusion of as much variability in terms of orchards and seasons as

possible but the use of cultivar specific calibration models when possible.


## Robustness across different varieties / commodities

While few studies have been published involving NIRS-PLS model robustness

in prediction of fruit attributes across varieties, similar studies in relation to other

commodities/attributes can offer insight into this issue.  Korcak *et al*. (1990) predicted

leaf nitrogen (% N) using NIR spectroscopy of dried, ground fruit leaf tissue across 4

different fruit tree species (apple, peach, plum and pear) sampled over time (May to

August).  They found that models based on calibration populations (samples from

year 1) composed of 1 species gave the best predictions for that species (prediction

population samples taken the next year), with RMSEP values of 0.15% N or less.  A

model based on a combined calibration population of all species (600 leaf samples)

(SD not stated) predicted peach (RMSEP = 0.18) apple (RMSEP = 0.15), pear

(RMSEP = 0.36) and plum (RMSEP = 0.46) leaf total nitrogen (% N) from samples

obtained the following year.  The combined species model was recommended as

suitable for peach and apple but less accurate for pear and plum estimation of leaf

nitrogen.  Increasing the size of the calibration population over time may increase the

usefulness of the method for plum and pear tissue N analysis.

Kojima *et al.* (1994) used NIR spectroscopy to predict TSS of intact fruit for

various stages of development of 2 cultivars ('Kousui' and 'Housui') of Japanese

pear.  They used MLR (6 wavelengths in the spectral region between 1,500 and

2,400 nm) and found the calibration statistics were improved by adapting a *bias*

correction or combining populations of 2 consecutive years.  The populations of each cultivar were split into calibration and validation populations in the proportion of approximately 60:40.  The combined populations were simply the addition of the 2 years data.  The *bias* correction ranged from -0.01% TSS (RMSEP = 0.49% TSS) when 1 cultivar was predicted on its own prediction population, -0.53% TSS (RMSEP = 0.76% TSS) when used on the other cultivar's prediction population, and −0.03 and −0.12% TSS (RMSEP = 0.58 and 0.83% TSS) when the combined calibration was predicted on the 'Kousui' and 'Housui' validation populations, respectively.  It is worth noting the high *bias* compared to the RMSEP when the calibration model developed with the cultivar 'Housui' was used to predict on the cultivar 'Kousui'.

Slaughter *et al*. (1996) reported soluble solids (SS) on intact tomatoes across 30 different cultivars and 5 different locations on the fruit  - 4 around the equator and 1 at the blossom end.  Spectra were obtained from 400 fruit (100 fruit for calibration model development and 300 for validation) made up of 30 cultivars collected over 7 weeks, at various stages of growth (maturity and ripeness) (proportion of each cultivar in the calibration and validation populations not stated).  Although the purpose of this study was to optimise the position on the fruit for taking optical data, they did show that an acceptable single calibration model (using the spectral region 800 - 1,000 nm) across many cultivars could be developed.  Validation results on the remaining 300 fruit (proportion of each cultivar not stated) gave a $R_v^2$ of 0.79, RMSEP of 0.33% TSS with a *bias* of -0.05% TSS (SD of both calibration and validation populations not stated).

Ventura *et al*. (1998) used NIRS to non-destructively measure Brix (% TSS) in 2 cultivars of apple (*Malus domestica* Borkh., cv. 'Golden Delicious' and 'Jonagold'). A spectrum per fruit was taken on the "sunny side", and the population (total of 340

fruit) was split equally into calibration and validation populations. They used MLR

regression with up to 12 wavelengths covering the area of the spectrum from 811 -

999 nm. First derivative absorbance data gave better results than second derivative

absorbance data (data not shown). The best result on the combined calibration

population (both cultivars) was achieved using 12 wavelengths ($R_c^2$ = 0.56, RMSEC =

1.01% TSS (SD of 1.48% TSS), $R_v^2$ = 0.49, RMSEP = 1.14% TSS (SD of 1.45%

TSS) with *bias* = -0.13% TSS. When MLR was carried out separately on each

cultivar, results were better for 'Golden Delicious' ($R_c^2$ = 0.65, RMSEC = 0.91%

TSS, $R_v^2$ = 0.54, RMSEP = 1.10% TSS with *bias* = -0.25% TSS) than for 'Jonagold'

data ($R_c^2$ = 0.52, RMSEC = 0.91% TSS, $R_v^2$ = 0.46, RMSEP = 1.18% TSS with

*bias* = -0.08% TSS). The individual cultivar calibrations were not tested on the other

cultivar. However, these results are poor compared to other published studies

involving apples, a problem possibly attributable to the instrumentation (reflectance

optics) used in this study.


## Robustness – time / season

Change in growing conditions (within a season or between seasons) can effect

both the chemistry (e.g. acid or sugar levels) and the physical structure (e.g. skin

thickness, surface wax thickness). Both issues can impact NIRS-PLS predictions,

impacting the extent of overlapping bands, and the degree of incident light

reflection/penetration, respectively. In general then, models are not perfectly robust

in the prediction of attributes for agricultural produce and industries employing this

technique (e.g. protein content of wheat, oil content of rape seed) employ a procedure

of validation on new samples, with model updating or model replacement as

necessary.

Flinn and Murray (1987) used NIR reflectance spectroscopy to evaluate herbage quality in southern Australia. The study involved 195 dried and ground hay pasture samples gathered over 3 years from various western Victorian farms and samples consisted of many different grasses with a mean legume content of 16%. Calibrations were derived for estimation of crude protein (CP), neutral detergent fibre and *in vitro* dry matter, using MLR. Ninety-five samples were used in calibration and 97 in cross-validation. In the case of CP (range of values from 5 to 20%), the MLR equation developed in year 1 predicted for year 2 and 3, (RMSEP = 0.71% CP, *bias* = 0.11% CP, RMSEP = 1.04% CP, *bias* = 0.37% CP, respectively). No SD results were stated and all results were expressed as a percentage of dry matter. According to Flinn and Murray (1987), most calibration equations derived for individual years were "sufficiently robust" to extrapolate to other years.

Garcia-Ciudad *et al*. (1999) tested the robustness of a near infra-red reflectance spectroscopy (1,100 – 2,500 nm) calibration model developed with dried grass samples from 1 sampling year to predict the nitrogen content of samples from the same area but subsequent years. The population (selected to represent all variables in the sample population) of grass samples was composed of heterogeneous and botanically complex samples. These samples were gathered from different places, date of collection, soil type, growth stage and botanical composition in the 1 year (calibration population, SD of 4.0 g/Kg$^{-1}$ N) and then used to predict nitrogen (N) content on samples gathered in that year and the next 3 years (validation populations, SD values of 6.1, 3.7 and 5.6 g/Kg$^{-1}$ N, respectively). These workers used a MLR regression program on derivatised absorbance values, utilising 6 wavelengths over the 1,100 – 2,500 nm region of the spectrum. The model (developed in year 1) was robust enough to accurately predict nitrogen concentration in samples of grass

collected in different years, with robustness defined as a RPD ( $RPD = \dfrac{SD}{RMSEP}$ )

exceeding 3.0.  The RPD ranged from 3.46 to 6.04 in the prediction of nitrogen

content in the validation populations.  *Bias* was lowest in the validation population

taken from the same population as the original calibration population (0.13% N) and

highest in the year 2 validation population (2.01% N) in an analyte range of 6.0 to

38.7% N.

Peiris *et al*. (1998b) used a NIR transmittance technique for the assessment of

soluble solids in intact peaches (*Prunus persica* (L.) Batsch).  Calibrations to predict

SS were developed using MLR regression on second derivative absorbance data with

2 wavelengths (falling between 870 and 910 nm).  Selected samples from each of the

individual cultivar (4 cultivars) calibration populations were combined to create both

season (3 seasons) and cultivar calibrations populations to cover the entire range of SS

contents within the season or cultivar.  They found the best predictions (higher

correlation coefficients and lower RMSEP values) in 3 out of the 4 cultivars were

obtained after using calibration populations to predict on validation populations

(RMSEP values from 0.49% to 1.63% SS with *bias* values of 0.01% to –2.62% SS,

respectively) tested for each cultivar, in each year but not between years.  The

RMSEP values ranged from 0.90% to 1.41% SS with *bias* values from –0.01% to –

2.08% SS for a model developed using data of 3 seasons to predict on subsequent

seasons.  They concluded that the cultivar or season calibration can be "successfully

employed" to predict the SS content of fruit from different cultivars in different

seasons.  In the individual cultivar and year predictions, calibration and validation

populations were derived from the original population.

Obviously the extent of model robustness across seasons will impact greatly on

the practical application of this technology.  While the results of Peiris *et al*. (1998b)

are encouraging, further work is required to extend these observations across other commodities.

## Robustness – sample and equipment temperature

The work of Sanchez *et al.* (2003) focused on the external factors of fruit temperature, spectrometer temperature and ambient light on their effect on lack of robustness of the NIR calibration model for sugar in intact apples.  They concluded that the sugar prediction was affected by both changes in the fruit and spectrometer temperature, but not by ambient light in their optical configuration (interactance probe).  The influence of spectrometer temperature was more than twice that of fruit temperature in terms of *bias*, while there was little effect on RMSEP(C) from change in either fruit or equipment temperature.  Several procedures were recommended by which spectra collected at different sample or equipment populations could be added to the calibration population to yield a model insensitive to these parameters.

## CALIBRATION TRANSFER BETWEEN INSTRUMENTS

Multivariate calibration models are often intended for use over an extended period.  Models may be based on tens to even thousands of samples collected over time and covering the full range of expected sample spectral and attribute variability.  Thus these calibration models require considerable effort and time in development.  Obviously it is desirable to develop one model for use on multiple instruments.  However, spectroscopic data will contain variation specific to a given instrument and measurement conditions (de Noord 1994).  Further, a predictive model developed and

used on one instrument may change or lose robustness because of instrument drift or shift (change in response by the instrument due to wear, replacement of vital parts etc.), change in the sampling or measuring environment (e.g. temperature), or a change in the sample physical or chemical constitution.

Spectroscopic data will vary from instrument to instrument (scanning, filter and PDA instruments); over time on the 1 instrument (e.g. ageing of detectors energy quantum, replacement of vital parts); due to changes in measurement conditions (e.g. temperature) and changes in the chemical and physical composition of the sample (particle size or surface texture). Standardisation procedures have been developed in an attempt to eliminate the need for full recalibration when changing instrumentation.

## Standardisation Strategies

Standardisation between 2 different instruments from the same manufacturer, where the differences in spectral data are small, may be achieved by hardware matching (detector linearity, frequency accuracy, angle of incidence and purging) coupled with some data pre-processing and selection of 'robust' parts of the spectrum. However, this approach can not overcome the problem of calibration transfer between instruments of different types (scanning, PDA and filter). Further, in practice, no 2 instruments can be made to have exactly the same specifications.

Another approach is to eliminate the sources of data variation that are not intrinsically related to the attribute being predicted, but are a result of features which are specific to the instrument, the experimental conditions or the sample. Such variation may be removed using data pre-processing. de Noord (1994) and Swierenga *et al*. (1998). used 2 methods to improve the robustness of multivariate calibration

models, namely data pre-processing and adaptation of the calibration model by piecewise direct standardisation. Data pre-processing techniques can eliminate effects such as baseline shifts (first and second derivatives); multiplicative and additive effects caused by different particle sizes (multiplicative scatter correction); information not relevant spectrally to the attribute being measured (wavelength selection); and wavelength shifts and slope variation in a spectrum (standard normal variance transformation) (Shenk and Westerhaus (1991), Barnes *et al.* (1989), Bouveresse and Massart 1996)).

The development of 'robust' calibration models is another strategy to minimise the need to standardize instruments (de Noord 1994). One approach involves simply incorporating all relevant sources of variation within the model (e.g. undertaking spectral measurements over time, different instruments, instrument temperatures etc.). According to de Noord (1994) this technique is generally not applicable unless the differences between populations are small to moderate. This is because of the clustering of spectral data due to the discrete performance of each instrument.

Shenk and Westerhaus in their chemometric software package WINISI standardize within instrument types (monochromators) and between instrument types (monochromator to filter) using single sample or multiple sample standardisations to make spectra alike between master and host instruments. The technique involves scanning known samples ('check samples') on the master and then the host instrument. An offset correction (photometric correction) is then made in the simple single sample standardisation method, while more complex mathematics (e.g. quadratic correction for wavelength differences between instruments, linear slope and *bias* corrections for each wavelength) are used in the multiple sample standardization.

Another technique involving collection of spectra of standard samples on both slave and master instruments is that of direct standardisation. Direct standardisation involves transforming spectra gathered on 1 instrument (master) to appear as if they were measured on another instrument (slave). Swierenga *et al*. (1998) concluded that direct standardisation gives slightly better or comparable results with regard to data pre-processing when the model is transferred between instruments. However, direct standardisation requires more information than data pre-processing. The same samples must be run on all instruments to be standardised, and 'standard' samples as close as possible to the samples being sorted must be used. For stable products such as wheat this is possible (although with the danger of sample repacking during transport between instrument locations). However, for perishable items such as intact fruit, this issue precludes standardisation of instruments at different physical locations. The development of a stable 'standard', spectrally similar to a fruit, would be very useful!

## CONCLUSION

There are many reports of the use of NIRS to predict internal quality attributes of fruit. However, interpretation of literature reports of NIRS model development requires a level of detail to be provided that is, unfortunately, frequently lacking. All reports of uses of NIRS should be clear with respect to population structure (e.g. selection method, source and relationship between the calibration and validation populations) and should report the populations' SD values, spectral range used and data pre-processing techniques. Calibration and prediction results should be reported as $R_c^2$, RMSECV, SD and $R_v^2$, RMSEP and SD, respectively.

Near infra-red spectroscopy has been used to ascertain quality attributes non-destructively of both wet and dry agricultural commodities. Some work has been undertaken on intact fruit with regard to robustness of NIR calibration models (time and different cultivars/varieties). However, many previous studies have been limited because of the lack of true validation that is, validation of the calibration model on a population truly independent of the calibration population. For NIR technology to have practical and commercial applications in the food/horticultural industry, the question of how robust NIR calibration models are over time, cultivar/variety and growing district must be answered. This involves optimising and establishing protocols for data pre-treatments, population structuring and subsequent chemometric analysis utilising various regression techniques.

### *Thesis direction*

Many NIRS based calibration models for intact fruit assessment have been developed and reported by various groups. However, at the time of conception of this thesis (1997), there were no published reports on the robustness of calibration models on new validation populations. Subsequently literature reports (see above) have identified problems with robustness of NIR calibrations for intact fruit, particularly across seasons. Existing fruit spectra and related attribute databases were available (the Journal of Near Infrared Spectroscopy maintains a collection of spectral data associated with published manuscripts), which could be used to consider model robustness. However different spectrometers and optical arrangements used in the collection of these databases make integration of data from different population within such databases meaningless.

Therefore, in this thesis an instrumentation platform was developed with regard to detector, light source and optical mode (Chapter 2).  This platform was then used to collect spectra of various horticultural produce (intact mandarins, melons and macadamia kernels) from different growing districts, time of harvest, variety and seasons.  Calibration model development was optimised with regard to attribute distribution within the sample, spectral window, mathematical treatment of the absorbance data (e.g. derivatisation and gap size), data smoothing and scatter correction techniques (e.g. in mandarin, Chapter 3).  The improvement in performance of calibration models for prediction of TSS and DM in intact mandarin fruit (a high moisture product) was assessed with regard to various techniques of sample selection and subsequent inclusion in the existing calibration population (model updating) (Chapter 4).  This work was repeated using intact melon fruit of different varieties (Chapter 5).  A low moisture product (intact macadamia kernels) was then assessed for oil and moisture content with various instruments and optical arrangements. Model updating was also assessed for this product using oil content (Chapter 6).  The final chapter draws together these results with final conclusions and recommendations.

# 2

# APPLICATION OF COMMERCIALLY AVAILABLE, LOW COST, MINIATURIZED NIR SPECTROMETERS TO THE ASSESSMENT OF THE SUGAR CONTENT OF INTACT FRUIT[1]

## ABSTRACT

Recent decreases in costs, and improvements in performance, of silicon (Si) array detectors open a range of potential applications of relevance to plant physiologists, associated with spectral analysis in the visible and short-wave near infra-red spectrum. The performance characteristics of 3 commercially available 'miniature' spectrometers based on Si array detectors operating in the 650 – 1,050 nm spectral region (MMS1 from Zeiss, S2000 from Ocean Optics, and FICS from Oriel, operated with a Larry detector) were compared with respect to the application of non-invasive prediction of sugar content of fruit using near infra-red spectroscopy. The FICS-Larry gave the best wavelength resolution, however, the narrow slit and small pixel size of the CCD detector resulted in a very low sensitivity and this instrumentation was not considered further.

---

Wavelength resolution was poor with the MMS1 relative to the S2000 (e.g. full width half maximum height (FWHM) of the 912 nm Hg peak, 13 and 2 nm for the MMS1 and S2000 respectively), but the large pixel height of the array used in the MMS1 gave it a sensitivity comparable to the S2000. The signal to noise (S/N) standard error ratio of spectra was greater by an order of magnitude with the MMS1, relative to the S2000, at both near saturation and low light levels. Calibrations were developed using reflectance spectra of filter paper soaked in a range of concentrations (0 – 20% w/v) of sucrose, using a modified partial least squares procedure. Calibrations developed with the MMS1 were superior to those developed using the S2000 (e.g. coefficient of correlation of 0.90 and 0.62, and standard error of cross-validation of 1.9 and 5.4%, respectively), indicating the importance of high signal to noise ratio over wavelength resolution to calibration accuracy. The design of a bench top assembly using the MMS1 for the non-invasive assessment of mesocarp sugar content of (intact) melon fruit is reported in terms of light source and angle between detector and light source, and optimisation of math treatment (derivative condition and smoothing function).

## INTRODUCTION

Throughout the past 50 years, breeding and post-harvest physiology programs associated with fruit have focussed on the production issues of quantity, and quality with respect to storage life and visual appearance. The consumer perceives that eating quality of fruit has decreased over this time frame. Agronomic and breeding programs can deliver fruit with improved eating quality, however this goal has not received emphasis because of the difficulty of assessing internal attributes of every item of fruit. Various non-invasive technologies such as nuclear magnetic resonance, chlorophyll fluorescence, acoustics, and near infra-red spectroscopy (NIRS) can be

applied to the task of non-invasive assessment of fruit eating quality attributes.  At the present time, NIRS is the most appropriate technique in terms of speed of assessment and cost.

Near infra-red spectroscopy has been applied to the non-invasive estimation of fruit eating quality, and is in commercial use in Japan (Kawano 1994b).  Published reports of such applications have largely involved either the use of research grade near infra-red instrumentation (Yermiyahu *et al.* 1997), unsuited to packing shed or field use (e.g. Guthrie and Walsh (1997a), instrument value ca. $AUD 100,000) or the use of purpose built spectrometers, unavailable for general application (e.g. Jaenisch *et al.* (1990), Bellon *et al.* (1993), Peiris *et al.* (1998b)).  During the mid to late 1990s, however, several low cost (< $AUD 10,000), miniature (spectrometer size < 500 $cm^3$), array spectrometers capable of operation up to 1,050 nm became commercially available.  These spectrometer modules are finding use in a range of instrumentation of interest to plant physiologists (e.g. portable spectroradiometers).

Osborne *et al.* (1996) reported the use of the Zeiss MMS1 spectrometer for the prediction of sugar content of kiwifruit, and Bellon and Vigneau (1995) reported the use of an Oriel Instaspec 2 to predict sugar content of apples, under laboratory conditions.  Mowat and Poole (1997) employed an Ocean Optics S1000 and a laptop computer to discriminate between field populations of kiwifruit.  However, the choice of instrumentation for the task of non-invasive assessment of fruit quality is made difficult by a lack of published specification requirements for this task.  We have contributed to this field with a consideration of the wavelength resolution and signal to noise (S/N) requirements of the task (Greensill and Walsh 1999).  In the current manuscript we review the design requirements for this application, compare commercially available spectrometer modules that have been used by different

researchers with respect to these criteria, and report on the design of an optical system suited to the assessment of melons.

Infra-red (IR) radiation is strongly absorbed by organic molecules, with the wavelength of absorption characteristic of the molecular bond. Overtones of the fundamental band (IR) frequencies, particularly those arising from R-H stretching modes (O-H, C-H, S-H, N-H, etc.), cause absorbance in the NIR region of the spectrum, although this absorbance is typically 10 – 1,000 times weaker than that of the fundamental band. Infra-red peaks are narrow and diagnostic, and thus instrumentation capable of high wavelength resolution is desirable. In contrast, peaks in the NIR spectra are broad, up to 100 -150 nm wide. However, as radiation sources are readily available to deliver high intensities in the NIR region and as detectors sensitive to this region have a relatively low S/N, NIRS lends itself to the quantification of organic constituents.

Near infra-red spectroscopy has been used in many fields, with most work carried out in the region of 1,100 – 2,500 nm (lead sulphide (PbS) detector). However, strong absorbance by water at around 1,600 nm has restricted use of the technique to dry materials and to reflectance optics. Hydrated objects are characterised by complicated hydrogen-bonding interactions between water, sugar, protein, etc., which complicate the spectra obtained. The application of short wave NIR (700-1,100 nm) is promising because: (1) the bands are ascribed to the third and fourth overtones of O-H and C-H stretching modes and are expected to be separated due to anharmonicity, (2) lower absorbance at these wavelengths allows for transmission optics, and (3) the corresponding instrumentation is low-cost and suited to process control, and portable enough for *in situ* field measurements. The ability to collect and interpret spectra of hydrated objects using short-wave NIR has blossomed

in the past decade, with advances in detector arrays, fibre optics and personal computing power.

Unfortunately, due to the complexity of NIR spectra (band overlaps), relatively sophisticated chemometric procedures (data processing such as derivatives, and data analysis using multiple linear, partial least squares or neural network regression techniques) are required for spectral analysis.  Spectral data can easily be over-fitted in the regression analysis.  The resultant calibration is useful for the predictions within the populations from which it was developed, but can fail in use on new populations (i.e. the calibration is not robust).  Instrumentation drift over time can also result in prediction failure, and differences between instrument units can preclude calibration transfer between instruments.

The design of the spectrometer can be rationalised with respect to the application.  For example, given the broad character of the absorption peaks in the NIR region, it is possible that spectral resolution may be traded off to increase detector sensitivity (i.e. a wider slit, or wider pixels).  Spectral resolution may be determined by pixel dispersion (the range of wavelengths divided by the number of pixels), but is otherwise a function of slit width as well as the quality of the dispersive element (e.g. density of lines on grating).  Further, the dispersive element may be chosen with transmissivity characteristic, rather than wavelength resolution, as the primary feature.  The type of detector should also be considered with respect to the application requirements.  Silicon (Si) detectors are sensitive into the NIR up to about 1,100 nm, while indium gallium arsenide (InGaAs) detectors are useful over the 900 to 1,700 nm spectral region.  However, Si detectors are preferred for reasons of cost and S/N.  Photodiode silicon detectors are approximately 100 times less sensitive to light than charge coupled device (CCD) Si detectors, but the higher saturation level of

the photodiode support a 10 fold higher maximum S/N ratio for this detector, relative to CCD detectors (i.e. 10,000 cf. 1,000). Overall, CCD's are preferred for very low light applications, while photodiodes are the better choice for accurate absorbance measurements when higher light levels are available (Oriel 1994). However, signal per detector pixel can be increased by increasing the height of the pixel and slit, by focussing light from a high slit onto the array or by summing columns in a 2 dimensional array.

The specification requirement for a spectrometer to support NIRS assessment of fruit in an in-line or field setting includes high S/N ratio, relatively high sensitivity (particularly if complete transmission spectroscopy is intended), and tolerance to vibration and dust. In-line application also requires the capacity for rapid spectral acquisition, with assessment of up to 10 pieces of fruit per second. Scanning grating instruments, with light detected by a single detector, are too slow in this respect, and are also vibration sensitive in terms of wavelength calibration. A stationary dispersive element and a fixed detector array can be very robust in terms of wavelength reproducibility, and very rapid in terms of spectra acquisition. Therefore, the typical spectrometer for the in-line sorting of fruit will consist of an entrance slit (with an inverse relationship between spectrometer sensitivity and wavelength resolution), a dispersive element (prism, grating or acousto-optical tuneable filter (AOTF)), a fixed array detector (linear Si or InGaAs photodiode array (PDA); or linear or 2 dimensional charge coupled device array (CCD), and an analogue to digital conversion device (usually 8 to 16 bit, i.e. a grey scale of 256 to 65,536 levels, typically up to the dynamic range – maximum signal/detection limit – of the instrument). A grating is the usual choice for the dispersive element, blazed at a wavelength in the NIR to maximise the efficiency of light transmission in this range,

although the spectrometer used by Dull *et al*. (1989) and Dull *et al*. (1992); to assess melon soluble solids content and Peiris *et al*. (1998b); to assess peach soluble solids content, utilised an AOTF as the dispersive element. The detector is usually either a Si PDA (as in the Zeiss MMS1, as used by Osborne *et al*. (1996); and the Oriel Instaspec 2, as used by Bellon and Vigneau (1995), or a linear CCD array (as in the Ocean Optics S1000, as used by Mowat *et al*. (1997)). Bellon *et al*. (1993) described the application of a 2 dimensional CCD array, comprised of 500 by 582 pixels (pixels 17 by 11 μm). A grating was used to disperse the light such that rows represent spectra, and columns were averaged to increase the S/N ratio. The Oriel FICS unit is capable of accepting various detectors, but is optimally used with a 2,500 μm high PDA detector. As costs decrease, InGaAs arrays will offer potential, operating over the wavelength range 900 – 1,700 nm.

The spectral response, and the stability of this response, of a spectrometer will be effected by the spectral output of the light source, transmission and reflection characteristics of the optical path within the spectrometer (e.g. entrance slit width, grating groove density), the stability of the mounting of the optical components (with respect to vibration and thermal expansion coefficients), the spectral response of the detector, and the stability of the electronics. The effect of trade-offs between wavelength resolution and decreased light levels at the detector (e.g. narrower entrance slit), and between S/N ratio and detector sensitivity (i.e. photodiode cf. CCD) encountered in the choice of NIR instrumentation, deserve attention with respect to the task of assessing the sugar content of intact fruit.

In this manuscript we evaluate 3 commercially available, low cost NIR spectrometers which differ in terms of the aforementioned parameters, with respect to the non-invasive measurement of Brix (sugar content) of melon fruit. We also report

on the optimisation of an optical configuration suited to the assessment of melons, and the optimisation of chemometric processing technique. A field portable unit has subsequently been based on this design, and melon spectra collected across growing district and time to explore calibration robustness issues.

## MATERIALS AND METHODS

### *Spectrometer description*

Three commercially available miniature spectrometers, with gratings chosen for operation in the NIR, were acquired – the Zeiss MMS1 (Zeiss, Germany), the Ocean Optics S2000 (Dunedin, Florida, USA, distributed through LasTek, Adelaide, Australia) and the Oriel Fixed Image Compact Spectrometer (FICS, model 77443), using a Larry linear CCD array (distributed through LasTek, Adelaide, Australia).

The Zeiss MMS1 (Monolithic Miniature Spectrometer), released in 1994, consists of a block of glass (UBK 7) with the imaging grating directly replicated onto 1 surface. The body thus acts as the dispersive element, and also images the entrance slit onto the diode array by varying groove density and using curved grooves to correct coma and flatten the focal curve to optimise use of the flat detector structure (6 mm long). The refractive index of the material (UBK 7) used in the construction of the body is higher than that of flint glass, giving greater angles of refraction and thus enabling the unit to be reduced in size. With the monolithic construction, the grating is immovable and thus vibration tolerant and protected against dust, and the spectrometer is relatively tolerant of temperature changes (wavelength drift of 0.012 nm/K specified). A fibre optic cross section converter is employed, with a linear arrangement of 30 quartz fibres (each 70 μm wide) acting as slit for the instrument. Thus slit width is not alterable. A Hammamatsu diode array (S3904-256Q, 256

elements, each 25 x 2,500 μm, 6 mm total length) was used as the detector. With detection of wavelengths between 300 and 1,150 nm, the MMS1 has a pixel dispersion of 3.3 nm/pixel. Order sorting filters are applied during manufacture to different regions of the array to eliminate detection of second order spectra over this wide wavelength range. A 12 bit analogue to digital conversion device was used, under the control of Zeiss software. The Zeiss software supplied with the MMS1 employs a smoothing function for its graphical display, but not on saved data, as used in the calculations of mean and standard error of signal in this study.

The Ocean Optics S2000, released in 1997, has increased sensitivity relative to the original 1992 release (S1000). A 2,048 element linear CCD array (each 12.5 x 200 μm, 4 mm total length) is employed, with only the mid-section used to minimise problems with field distortion. To optimise use in the NIR region, an order sorting filter (550 nm) was factory installed. With the grating (# 14, blaze 1,000 nm, 600 l/mm) and the slit width (50 μm) chosen (factory installed), a 3 nm resolution is specified. With a nominal wavelength range of 632 to 1,278 nm, and a number of blackened pixels, the S2000 has a pixel dispersion of 0.36 nm/pixel. A 12 bit analogue to digital conversion device was used, with data acquisition controlled by Spectra Array (LasTek, Adelaide, Australia).

With the Oriel, the slit is mounted on a slide, and so can be varied. A 25 μm slit was used in this study. A 2,048 element linear 'Larry' CCD array (12.5 x 200 μm) was employed as the detector. With the grating (blaze 1,000 nm, 600 l/mm) chosen, a wavelength range between 300 and 1,150 nm was detected, giving a pixel dispersion of 0.41 nm/pixel. A 12 bit analogue to digital conversion device was used, with data acquisition controlled by Spectra Array (LasTek, Adelaide, Australia).

*Spectrometer comparisons*

The performance of the spectrometers was compared in terms of wavelength resolution and stability, relative spectral sensitivity, relative detector sensitivity, signal to noise ratio, stability over time and variation in temperature and calibration performance.  To achieve these comparisons, wavelength calibration was undertaken using a mercury argon lamp (HG1, Ocean Optics), and spectra were acquired of a reference material (WS - 1, halon reference, Ocean Optics) and of samples (i.e. Whatman # 1104 filter papers immersed in a range of concentrations of sucrose, then allowed to drain prior to scanning.

Spectra of filter paper soaked with sucrose solutions were collected using a reflectance probe with 6 illumination fibres and 1 read fibre (all fibres 400 μm in diameter, R400 - 7, Ocean Optics).  The read fibre was directed to a spectrometer, while the illumination fibres were connected to a 6 W tungsten halogen light source (LS - 1, Ocean Optics).  Sub multi assembly (SMA) connectors (NA 0.22) were used to connect these items.  Spectrometer temperature was varied by placing the spectrometer within the oven of a Fisher Gas Partitioner (Model 1200) and within an ice - box, and monitored by a thermocouple placed in contact with the spectrometer body.  Temperature was ramped from ambient (22°) to 0°C, and then increased to 45°C, at approximately 0.2°C/min.

*NIRS calibration technique*

Two populations ($n$ = 40 (2 spectra per fruit); $n$ = 210 (1 spectra per fruit); combined Brix range 5.4 to 11.2°) of rockmelons (var. 'Doubloon') were sourced from the Bowen-Burdekin (North Queensland) region in November 1998 for use in the comparison of instrumentation.  A further 10 populations (total $n$ = 1,991, 2

spectra per fruit, Brix range 4.4 to 12.2°) of rockmelons (4 varieties, from various

growing regions) were obtained during 1998 and 1999 and spectra collected using the

purpose built instrument described below.  This larger data set was used in the

consideration of the optimal data treatment for calibration, with the optimal data

treatment then used for the calibrations involving the comparison of instrumentation

reported in this manuscript.  Spectra acquisition and wet chemistry occurred within 3

days of harvest at the 'slip' stage (fruit breaking away from peduncle).  Spectral data

was acquired of an area of the fruit equidistant between point of attachment of

peduncle and corolla, but not of an area which had been in contact with the ground

during fruit growth.  Where 2 spectra were acquired per fruit, spectra were acquired

from opposite sides of the fruit.  Juice was extracted of 40 mm diameter plugs of fruit

mesocarp tissue underlying the assessed areas, with soluble solids concentration

assessed using an Erma (Tokyo, Japan) digital refractometer.  While a range of ca. 6°

Brix was recorded between different fruits (population ranges reported above), a

range of ca. 1.5° Brix units was also recorded within the central region of a given fruit

(i.e. around the 'equator' of the fruit).  Brix of prepared sucrose solutions was also

measured using the Erma refractometer.

Zeiss and Spectra Array files were converted to JCAMP format, and imported

into the chemometric package WINISI (version 3.0, Infrasoft International, PA,

USA).  Spectral outliers were defined using the WINISI critical 'GH' statistic (a

measure of distance of spectral sample to population mean, based on an estimate of

the Mahalanobis distance, D, calculated on principle component scores, defined as

$GH = D^2/f$, where $f$ is the number of factors in the PLS regression) set to a value of 3.

Analysis involved a modified partial least squares (MPLS) procedure using raw, first

or second derivative absorbance data and 6 cross-validation groups.  Standard normal

variance (SNV) and detrend were used for scatter correction. The effect of the number of data points used in the derivative calculation ('gap') and the number of data points used in a smoothing routine offered in the WINISI software was considered. As suggested in the WINISI manual, calibrations were compared primarily on the root mean square error of cross validation (RMSECV) statistic, where RMSECV should not be more than 20% greater than RMSEC, and attention given to the 1-VR (variance ratio, i.e. regression coefficient of predicted on actual Brix for the validation population) statistic. The standard error of calibration (RMSEC) and regression coefficient of predicted on actual Brix for the calibration set ($R_c^2$) are reported. Note that the terms RMSECV and 1-VR, and RMSEC and $R_c^2$, respectively, are directly related for a given population (e.g. $R_c^2 = 1 - (RMSEC/SD)^2$, where SD is the standard deviation of the actual Brix for the population).

***Optimisation of optical configuration for the non-invasive assessment of melon fruit sugar content***

Philips halotone (12V, 50W, 50$^o$ light spread, aluminium reflector) lamps were used as light sources and the Zeiss MMS1 unit used as the detector. Reflectance (in which specularly reflected light is received by the detector) and partial transmission optical arrangements were trialled, as the optical density of the fruit prevented full transmission optics. The core configuration consisted of the detector fibre optic positioned to view the 'top' of the fruit (a position on the fruit equidistant from peduncle and blossom ends which was not an area of the fruit which had rested on the ground during fruit growth). Lamp(s) were positioned to illuminate the fruit at some (varied) distance from the area seen by the detector (i.e. angle between area of detection, centre of fruit and area of illumination varied).

The intensity of light received by the detector and the calibration performance (for prediction of mesocarp Brix) was considered with reference to the following variables: (i) the angle of incidence of light onto the fruit surface, (ii) the angle between detected area and illuminated area with reference to the centre of the fruit (i.e. distance between detected area of fruit and illuminated area), (iii) the number of lamps employed, (iv) the distance from detector fibre optic to fruit, (v) the presence of a shroud between detector fibre optic and fruit surface and (vi) the duration of illumination (with respect to temperature of fruit).

# RESULTS AND DISCUSSION

## *Wavelength accuracy and resolution*

The instruments were calibrated using a mercury argon (HgAr) lamp, and spectra of the mercury argon lamp acquired at near saturation count at the 842.5 nm emission line (Fig. 1). The spectrum acquired with the MMS1 unit demonstrated a poor wavelength resolution relative to either the S2000 (Fig. 1) or the Oriel unit (data not shown). Second order spectral peaks were recorded with the S2000, but not the other instruments (e.g. at 1,080 and 1,155 nm, data not shown). The peak at 912.3 nm was chosen for further characterisation as it was isolated from other peaks in the MMS1 spectrum. Spectra were acquired with the count of this peak at near saturation and normalised between instruments. The line width (full width at half maximum, FWHM) of the Oriel and S2000 was 1.2 and 2.1 nm, respectively, an order of magnitude superior to the MMS1 result of 13.1 nm. These results are consistent with the slit widths, pixel dispersion and geometries of the 3 units.
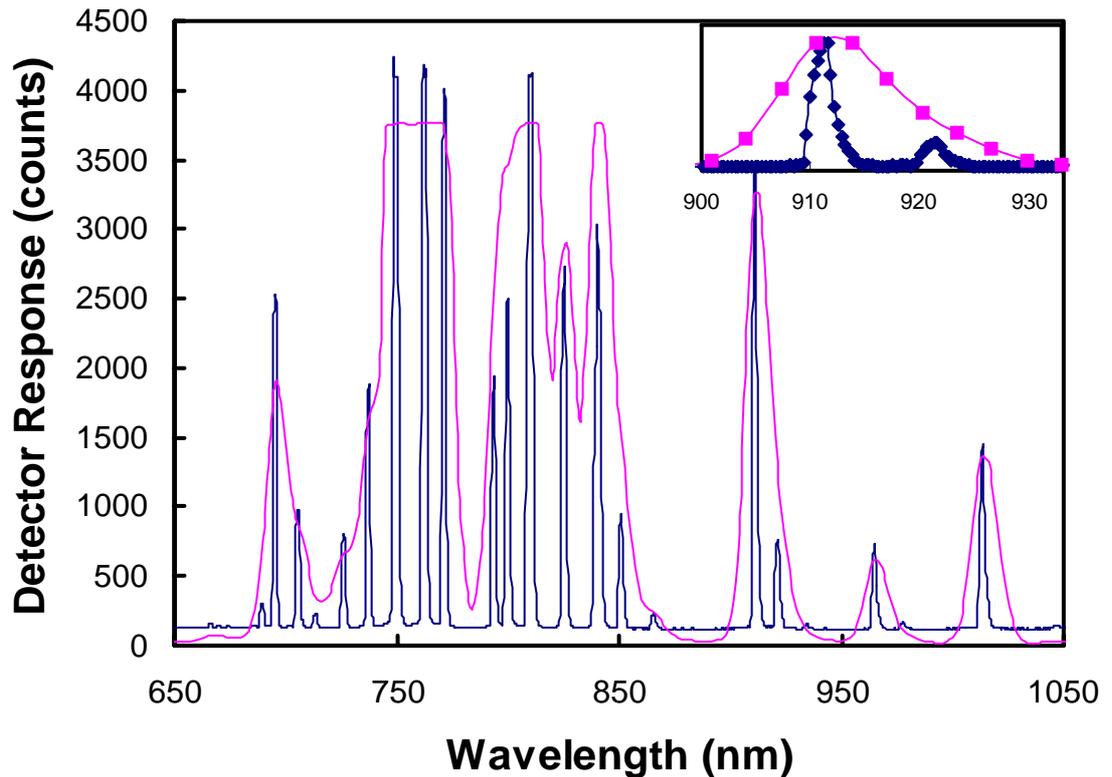
**Figure 1. Spectra of a mercury argon lamp acquired with the Zeiss MMS1 (red line) and the Ocean Optics S2000 (black line) spectrometers. Inset illustrates the resolution of the 912 nm peak by the 2 devices (with detector response normalised to output at this wavelength).**

Array spectrometers have a reputation for wavelength precision, relative to instruments in which the monochromator is a moving grating (and therefore sensitive to mechanical disturbance). The wavelength calibrations of the MMS1 and S2000 were checked periodically over a period of 6 months. During this period the instruments were used in air-conditioned laboratories, but were subject to mild shocks and temperature fluctuations during transport between laboratories. No recalibration was necessary over this period for either instrument (i.e. measured position of 912.3 nm spectral line of HgAr lamp did not vary by more than 0.3 nm). However, the FWHM of array spectrometers can be sensitive to temperature, as differential expansion of materials within the spectrometer changes the geometry of the light path. The monolithic construction of the MMS1 should be advantageous in this respect.

Wavelength accuracy and FWHM was stable for both the MMS1 and the S2000 over the temperature range expected in a packing shed environment. The FWHM of the MMS1 was estimated at between 13.04 and 13.13 nm, with no consistent change as the temperature of the spectrometer was varied between 4 and 45°C (data not shown). The FWHM of the S2000 varied between 2.06 and 2.12 nm as temperature was varied over this range, tending to increase with temperature (data not shown).

### *Relative spectral sensitivity*

The 3 spectrometers employed silicon based detectors, and so are expected to show decreasing sensitivity through the region 700 – 1,100 nm, with no response beyond 1,100 nm. However, the spectral sensitivity of the instrument can be altered by doping of the silicon in the detector, by use of coatings over the surface of the detector elements, and with respect to the spectral efficiency of the grating (primarily determined by the blaze wavelength). Spectra were acquired using the 3 instruments of the reference material in reflectance mode, using the interactance probe and a tungsten halogen light source. Spectra with a maximum count level near saturation were acquired for each spectrometer, and spectra compared after normalization to the count at 730 nm (Fig. 2). The MMS1 was more sensitive than the other instruments over the wavelength range 750 – 1,050 nm, and particularly over the region 800 – 900 nm. The Oriel-Larry unit was more sensitive at wavelengths between 650 and 700 nm than at 720 nm (data not shown). The spectrum acquired using the MMS1 was also smoother than equivalent spectra acquired with the FICS or S2000. An increase in count after 1,060 nm was recorded with the S2000, a result interpreted as a second order spectra (the S2000 unit employed a 550 nm primary cut-off filter).
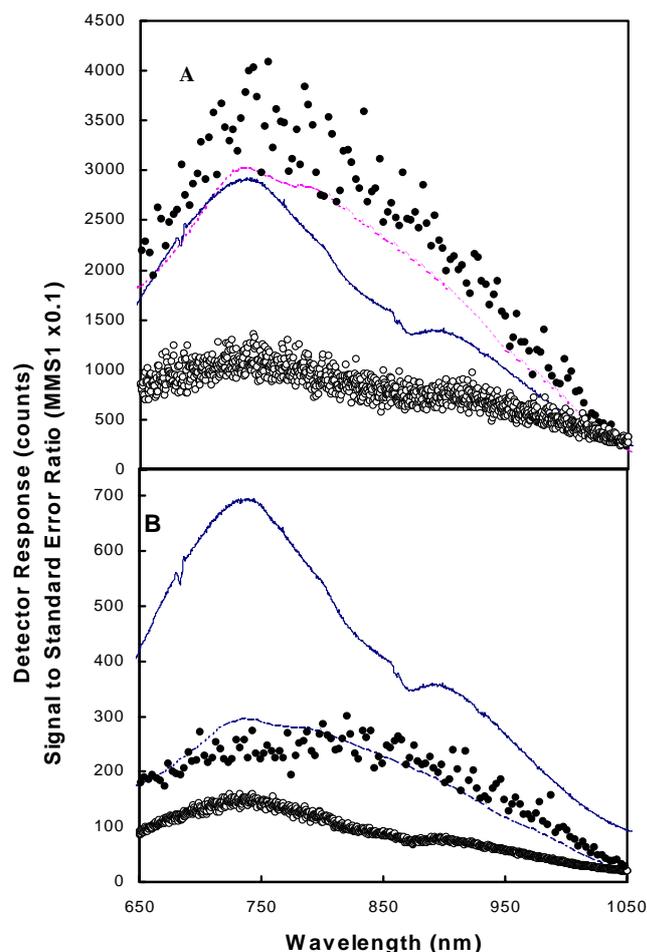
**Figure 2. Relative spectral sensitivity (lines) and signal to standard error ratio (circles) of spectra collected using the MMS1 (red dotted line, solid circle) and S2000 (solid line, open circle) spectrometers. Spectra were acquired using the same integration time (100 ms), light source, fibre optic guides and sample (reference material) for the 2 devices. Mean signal and mean signal divided by standard error of measurement at each wavelength (*n* = 50) are displayed. Note the scale change for the Zeiss MMS1 signal to noise ratio. (A) Light intensity was adjusted such that the output of each detector was near saturation, and normalised to output at 720 nm. (B) Spectra were acquired on both instruments at the same, relatively low, light intensity.**

Spectra were recorded of the HgAr lamp using the MMS1 spectrometer, while

altering the temperature of the spectrometer between 0 and 45°C (data not shown).

The measured count of a 'dark ' region of the HgAr lamp spectrum (870 nm)

increased with temperature by a count of 0.33 per °C (on a count of 29 at 0°C, linear

45

regression, $R^2 = 0.913$). This increase will reduce dynamic range with temperature increase. The measured count of the 912 nm emission line was more responsive to temperature, increasing by a count of 10.96 per °C (on a count of 2,757 at 0°C, linear regression, $R^2 = 0.913$). This result is consistent with Zeiss MMS Spectral Sensor product information (7 - 802e), which reports a sensitivity increase of ca. 0, 0.18, 0.47 and 0.69% per °C at 500, 735, 912 and 1,000 nm, respectively. Thus an increase of ca. 12 counts per °C on a count of 2,757 is expected.

Thus detector spectral sensitivity and dark current are changing with instrument temperature. These changes may be accommodated in a field application by minimising detector temperature change, and by collecting reference and dark spectra at the same instrument temperature as experienced while collecting sample spectra.

*Relative detector sensitivity*

Reflectance spectra of a reference material under halogen lamp illumination were acquired using the 3 spectrometers at a range of probe heights (i.e. different illumination levels) but the same acquisition time per spectrum (100 ms). Regression relationships were established between the readings of the 3 instruments. Detector response was recorded at 735 nm, as the wavelength at which highest counts were recorded in the MMS1 and S2000 units, and also a wavelength likely to be used in calibrations developed for the sugar content of fruit (e.g. Guthrie *et al.*(1997b)). The S2000 gave count readings 2.65 times higher than that of the MMS1 ($R^2 = 0.998$) (e.g. Fig. 2B), with a saturation count reached at only 25% of the range of the MMS1. In contrast, the slope of the Oriel – MMS1 regression was only 0.019 ($R^2 = 0.996$) (data not shown).

The sensitivity of a CCD array to light, in terms of electrons/count, is reported to be ca. 150 times greater than that of a PDA (Oriel 1994). The low sensitivity of the Oriel FICS assembly was primarily due to the design of the instrument optics to focus light onto a 2,500 μm height (PDA) array, not onto a (CCD) detector array only 17 μm tall, as used in this study. The relatively high sensitivity of the MMS1, as a photodiode array, relative to the S2000, as a CCD array, is explained by the degree of pixel dispersion in the 2 units (0.36 and 3.3 nm/pixel in the S2000 and MMS1, respectively), and also by the size of the pixels in the 2 arrays. The MMS1 pixel (25 x 2,500 μm) has an area 400 times greater than that of the S2000 CCD (12.5 by 12.5 μm). Also, the effective slit width of the MMS1, at 70 μm (diameter of fibre optic), was greater than that employed in the S2000 (50 μm). Thus each pixel of the MMS1 array received ca. 2,500 times greater illumination (number of photons) than in the S2000 array, for a given level of detected surface radiance.

*'Signal to Noise' Ratio*

Fifty reflectance spectra of the reference material were recorded at near saturation levels (3,000 counts with saturation recorded at 4,096 counts on a 12 bit analogue to digital (A to D) device for each instrument), and at a low light intensity (peak counts of > 300), for each instrument. The mean was divided by the standard error of the count at each wavelength as an estimate of the S/N ratio. Signal to noise ratio was estimated at the peak wavelength of 735 nm. With spectra recorded at near saturation levels, the maximum S/N ratio was approximately 40,000, 1,000 and 4,000 in the MMS1, S2000 (Fig. 2A) and Oriel units (data not shown), respectively. At a low light level of 10% of saturation for each instrument; the maximum S/N ratio was

approximately 3,000, 250 and 400 for the MMS1, S2000 (Fig. 2B) and Oriel instruments (data not shown), respectively.

The total pixel noise in the signal from either the photodiode or CCD array can be approximated as the square root of the sum of squares of the following 3 components, (a) read out noise, which is due to amplifier and electronics, (b) shot noise from the signal itself, equivalent to the square root of the signal, and (c) the shot noise of the dark current, which is dependent on exposure time and very dependent on temperature. The spectral shape of the noise (mean/standard error) values followed that of the signal, reflecting the importance of the signal shot noise to the total noise. The S/N ratio should be better for a CCD than a PDA for the operating range of the CCD, reflecting a lower read out noise, but the maximum S/N ratio of the PDA (achieved at higher signal levels) is expected to be an order of magnitude greater than that of the CCD (ca. 10,000 cf. 1,000, respectively; Oriel 1994). The results obtained with respect to maximum S/N of the PDA and CCD detectors was as expected. However, the S/N ratio was also higher for the PDA based spectrometer than with the CCD based instruments for light levels within the range of operation of the CCD. The low noise of the MMS1 at the lower light level is therefore attributed to a low read out noise, relative to that expected for a PDA.

Bellon *et al.* (1993) estimated the S/N ratio of their CCD based system by dividing the spectrum of a reference material by the standard error of 10 reflectance ratios (spectrum of reference material divided by a reference spectrum of the same material) at each assessed wavelength (rather than by the standard error of the repeated raw spectra). Assuming spectra were acquired at near saturation levels, and given the use of an 8 bit analogue (A) to digital (D) card (i.e. saturation at a count of 256), the report of a maximum S/N ratio of 90,000 is equivalent to a ratio of 360 (i.e.

90,000/256) in terms of the current study.  Thus the S/N ratio achieved by Bellon *et al.* (1993) was similar to that obtained with linear CCD arrays in the current study. This is surprising, in that as Bellon *et al.* (1993) averaged data over 512 rows, noise should have been decreased by a factor of the square root of 512 (22.6) over that of a single pixel.  The difference is attributed to noisier electronics in Bellon's equipment.

### *Stability of spectrometer and lamp output*

Using a light source which had been activated some 3 hours earlier, the output of the MMS1 and S2000 was recorded with respect to time from instrument activation (Fig. 3A, 4A).  For the MMS1, counts generally decreased (e.g. at 750 nm, by 25 counts on 3,500) (Fig. 3A, 4A).  The MMS1 was considered stable within 60 minutes of activation.  In contrast, the S2000 was relatively unstable, fluctuating by up to 3% of initial response, and not stable even after 90 minutes from activation.  The stability of instrument response is a critical parameter in consideration of the frequency of referencing required, or the preference for a dual beam over a single beam operation.

Using a MMS1 spectrometer which had been activated some 3 hours earlier, the spectral output of a tungsten halogen lamp was recorded with respect to time from activation (Fig. 3B, 4B).  Spectral output decreased by ca. 5%, across most wavelengths, but increased by ca. 2%, at 833 nm.  Most changes were complete within 30 minutes of lamp activation.  These spectral changes are attributed to the chemistry of the tungsten halogen lamp during a warm up period following ignition. Parallel data were collected with the S2000, with trends similar to that reported above (Fig. 4B).  The stability of the lamp output is also a critical factor in consideration of the frequency of referencing required in an application.
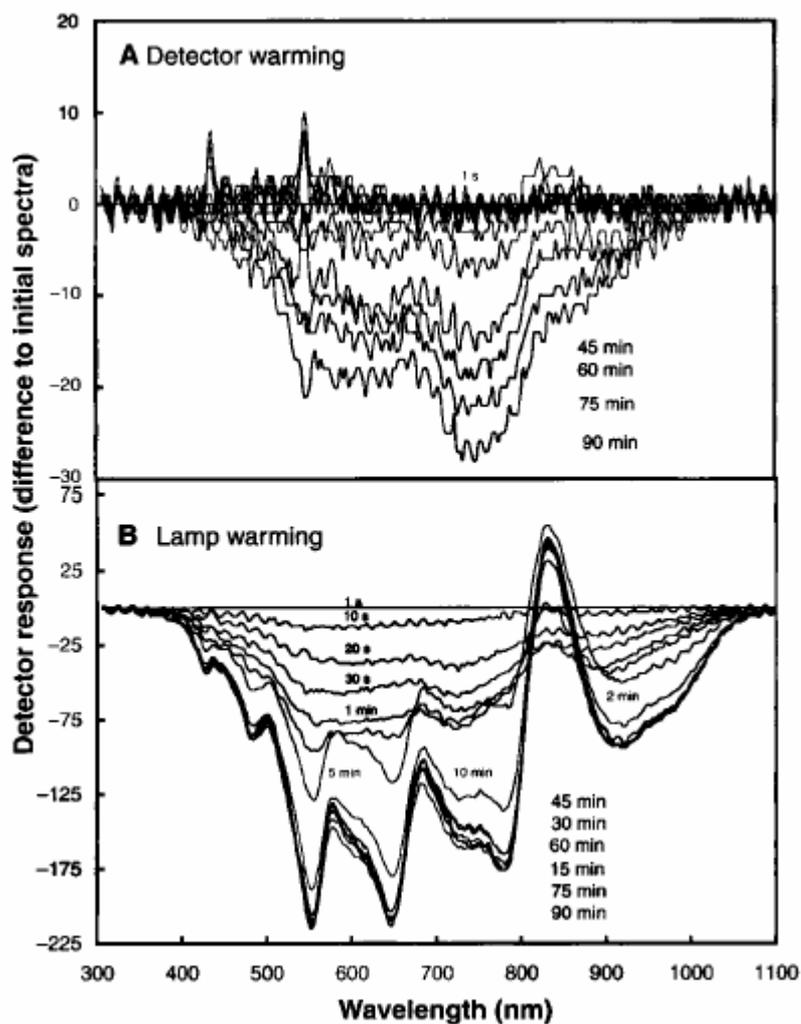
**Figure 3.** The stability of (A) detector response and (B) light source, as indicated by change in spectrometer response (interactance optics, reference sample) for the wavelength range 300 – 1,100 nm with time from instrument and lamp activation, respectively. Data expressed as the difference in the A/D card output to that of the first spectra acquired (at 1 second after detector and lamp activation, respectively). Spectra were obtained using a halogen light source and teflon as a reference sample.
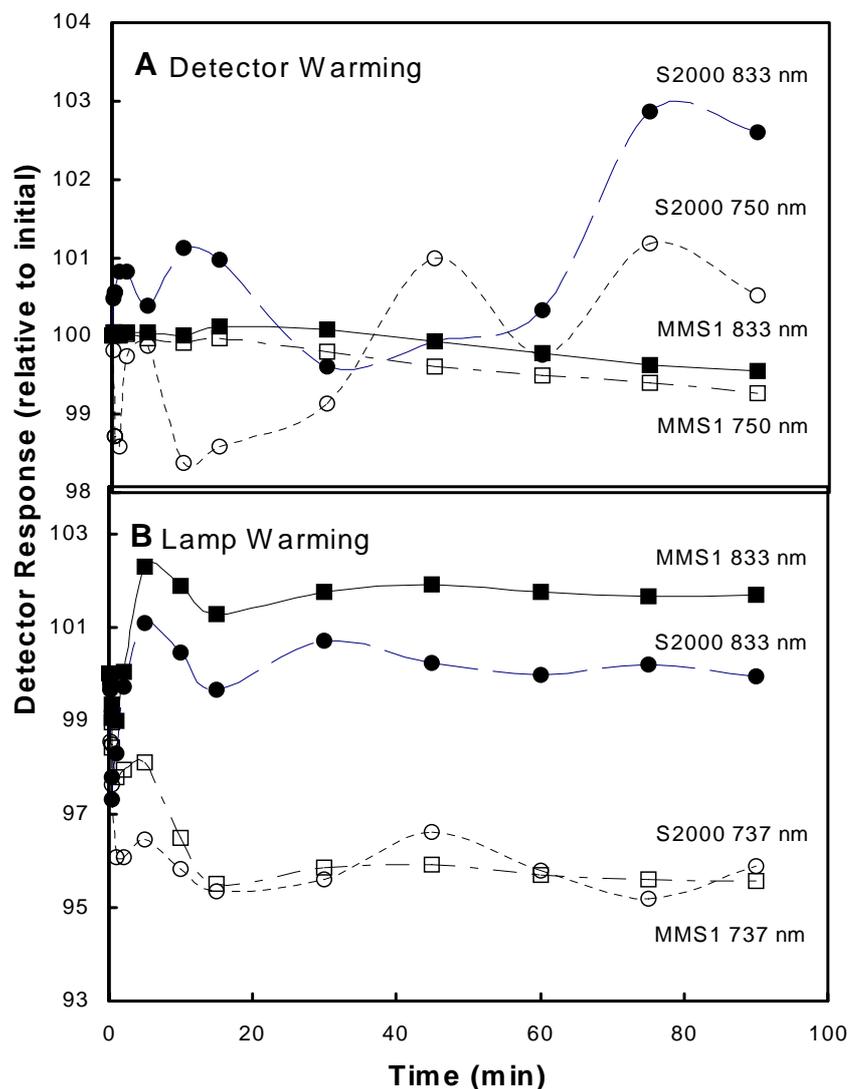
**Figure 4.** The stability of (A) detector (MMS1, squares; S2000, circles) response and (B) light source, as indicated by change in spectrometer response (interactance optics, reference sample) at 737 nm (open symbols) and 833 nm (closed symbols) with time from instrument and lamp activation, respectively. Data expressed as a percentage of the first recording (ca. 1 second after detector and lamp activation, respectively). Data is of the same experiment as presented in Fig. 3.

51

*Choice of spectrometer for the application of non-invasive sorting of fruit by NIRS*

The application of fruit sorting by NIRS requires an instrument which is relatively sensitive to light, in order to capture spectra of fruit in transmission or interactance modes without use of an unduly high incident radiation load (with the attendant sample heating problems). The instrument must be sensitive over the spectral region 700 – 1,050 nm (or higher), and the detector response must be relatively stable. As noted earlier, wavelength resolution below 10 nm is probably not necessary (e.g. Greensill (2000)).

The Oriel-Larry unit gave the best wavelength resolution of the 3 instruments. However, its sensitivity was poor (due to the detector used), and on this basis the instrument was eliminated from consideration. The S2000 gave better wavelength resolution and detector sensitivity than the MMS1. However, the relative response of the MMS1 in the near infrared (750 – 1,000 nm) region was higher than in the visible region than the S2000 (Fig. 2). Further, the S/N ratio of the MMS1 was an order of magnitude higher than the S2000, both at high light levels (i.e. near detector saturation) and at a low light levels (within the detection range of the CCD) (Fig. 2).

To compare the spectrometers for their application to the task of assessment of fruit by NIRS, we captured spectra of filter paper saturated with sugar solutions of varying concentration in one experiment. The MMS1 supported better calibrations than the S2000 (Table 1). We conclude that the attribute of wavelength resolution was not important to the calibration process, relative to the attribute of S/N ratio. Of the 3 instruments considered, we recommend the MMS1 for use in the application of fruit sorting by NIRS.

*Optimising optical configuration and instrument parameters for fruit sugar content calibration*

*Light angle relative to fruit and detector*

The intensity of light detected was not dependent on the angle of incidence of the light beam on the fruit surface (data not shown). This result is explained in terms of the diffuse transmission of light through the fruit, with incident radiation scattered within the fruit such that the angle of incident illumination has little effect.

As expected, the intensity of light detected at a given wavelength (800 nm) decreased as the light beam was moved away from the detected area. The decrease in detector response, $R$ (counts), was described with reference to the distance between the centre of the illuminated area and the detected area of the fruit, D (mm). This exercise was undertaken for an optical arrangement involving a single 50 W halogen lamp and the detector aligned to the centre of the fruit and positioned at 10 cm from the fruit surface (eqn. 1). The exercise was repeated incorporating a 45 mm diameter cylindrical shroud between the detector and the fruit surface, to eliminate specular radiation (eqn. 2).

$$R = 18525 \ e^{-0.0668 \ D} \qquad (R^2 = 0.943) \quad \text{(light without shroud) eqn. 1}$$

$$R = 64646 \ e^{-0.085 \ D} \qquad (R^2 = 0.994) \quad \text{(light shrouded)} \qquad \text{eqn. 2}$$

where $R$ is counts and $D$ is the distance between the centre of the illuminated area and the detected area of the fruit.

Thus, moving the detector from 40 to 50° with respect to the light source decreased the observed count from ca. 24 to 4% and 7 to 2% of maximum signal (i.e. detector saturation; 200 ms integration time) for the non-shrouded and shrouded arrangements, respectively.

**Table 1.** Calibration ($R_c^2$ and RMSEC) and validation (RMSECV) regression parameters (modified partial least squares procedure) of spectra collected with 2 spectrometers (MMS1 and S2000) using the same optical and sample presentation system (fibre optic interactance probe). Calibrations were performed using the second derivative (gap size of 4) of unsmoothed data from the wavelength range 700 – 1,050 nm. Scatter correction was not employed. Four spectra were acquired of each of 16 filter paper bundles, each saturated at different concentration of sucrose, at approximately 1.5° Brix intervals between 0 and 20.

| Instrument | Terms | $R^2$ | RMSEC | RMSECV |
|------------|-------|-------|-------|--------|
| MMS1 | 2 | 0.904 | 1.73 | 1.85 |
| S2000 | 2 | 0.619 | 3.44 | 5.40 |

We had expected that, at lesser angles of light source to detector, higher levels of radiation would be monitored, but that the 'path' of this light would be primarily through exocarp and outer mesocarp tissues. Therefore, increasing incident light to detector angle should allow for proportionally more spectral information on the tissue of interest, the edible mesocarp. However, at some incident light to detector angle, the disadvantage of decreased light transmission (i.e. decreased S/N ratio) must outweigh this advantage. Also, as the angle between incident light and detector is increased (particularly beyond 90°); it is expected that proportionally more of the detected light will have penetrated the seed cavity and carry spectral information about seeds, as well as about mesocarp tissue.

Spectra were acquired at 4 lamp-detector angles for 40 fruit, using a shroud between lamp and fruit. Calibration statistics (RMSEC and related $R_c^2$, and RMSECV, after outlier removal) were optimal at a detector-lamp-fruit angle of 60° (Table 2). For ease of fabrication, an angle of 45° between incident light and detector was adopted in further characterisation of the optical system.

*Number of lamps*

Increasing the number of lamps was expected to increase the available signal to the detector, and decrease the S/N ratio. Aoki *et al.* (1996) reported an optical arrangement employing 16 lamp positions around the equator of the melon fruit, with the detector viewing an area of the fruit at 90° to this plane. However, the use of more lamps will also increase the volume of the fruit 'sampled' by the light, which may degrade a calibration based on the analysis of a relatively small tissue sample in the primary analytical method.

Increasing the number of lamps from 1 to 4 had little effect on calibration performance (Table 2). Four lamps were employed in further characterisation of the optical system. These lamps were positioned at 45° in the vertical plane, with respect to the detector, and at 90° intervals in the horizontal plane, with respect to other lamps.

*Number of scans per spectra*

The S/N of spectra will improve proportionally to the square root of the number of scans averaged per spectrum. Calibration RMSECV decreased with increased number of scans, although this improvement was marginal and the $R_c^2$ and related RMSEC terms were degraded between 4 and 16 scans. Increased scan time will lead to sample change through heating, which could alter spectral characteristics and thus calibration performance (Guthrie *et al.* 1998). However, when fruit were held under the lamps for a period of 3 minutes, (fruit internal flesh increased in temperature by less than 1°C, while skin temperature rose by greater than 15°C), calibration performance was not significantly impacted (Table 2). Averaging of 4 scans per spectra was adopted in further characterisation of the optical system.

*Detector 'shrouding' and distance of detector to fruit*

Calibration performance was degraded by removal of the shroud between the detector and the fruit surface, in terms of $R_c^2$, RMSEC and RMSECV (Table 2). This result is consistent with the interpretation that the detection of specular and emergent light that has interacted only with the very surface layers (top few millimetres) of the fruit surface degrades calibration performance. The placement of a 40 mm high, 45 mm diameter collar on the fruit under the detector supported a calibration which was apparently superior to the arrangement employing a shroud between detector and fruit surface (Table 2). As both arrangements prevent specular reflections from the fruit surface from reaching the detector, equivalent calibrations were expected. Calibration performance was apparently slightly improved in terms of both calibration $R_c^2$ and RMSECV when the distance between fruit surface and detector/light source was allowed to vary in response to fruit size (i.e. by ca. 50 mm). Such change in fruit diameter altered the effective distance between illuminated and detected regions of the fruit surface. Further, as the detector fibre optic has a numerical aperture of 0.22 mm, an increasing area of the fruit surface is imaged as distance between the probe and the fruit surface is increased. This would result in an increased detector count, offset by a decrease in light intensity. Also, if the field of view of the detector overlaps the areas of direct lamp illumination of the fruit surface, specular reflection will also reach the detector. It was expected that calibration performance should therefore decrease when distance from detector to fruit varied.

**Table 2.** The effect of lamp-fruit-detector angle, a signal filtering routine, number of lights, number of scans averaged per spectrum acquired, and the presence of a shroud between lamp and fruit surface on the calibration of Zeiss MMS1 spectral data (700 – 1,050 nm) to melon flesh Brix, in comparison with 2 reflectance mode bench top NIR spectrometers. Calibrations were developed for spectra ($n = 40$) acquired of a population of fruit (mean 9.5, standard deviation 1.7° Brix) for various lamp – detector angles and for 3 spectrometers, and for spectra acquired of a population of 208 fruit (mean 8.3, standard deviation 1.2° Brix) for conditions varying with respect to the number of lights, presence of a shroud, and number of scans averaged. The default configuration consisted of shroud between detector and fruit, 4 lamps mounted to illuminate the fruit at 40°with respect to the detector, and averaging of 4 scans per spectrum, with a 200 ms integration time per spectrum. Data of treatments marked with an asterisk has been repeated for ease of data comparison. For condition 16b, fruit were held under the lamps for 3 minutes before scanning. A data treatment of scatter correction (SNV and detrend), second derivative gap size of 4, with no data smoothing was adopted for all calibrations. The Savitzky-Golay filtering routine (SG) was applied to spectra acquired in the assessment of 40° lamp angle. MPLS regressions were performed with all data, and with removal of outlier data as identified using the WINISI chemometric package critical global Mahalanobis distance (GH) of 3. The MMS1 was used with a shroud, 4 lamps illuminating the fruit at 45°, and 4 scan averaging in the spectrometer comparison exercise. Spectral data over the wavelength range 700 – 1,050 nm, 700 - 1,700, and 700 – 2,300 nm was used from the MMS1, Perten DA 7000 and NIR Systems 6500 spectrometers, respectively.

| Attributes | n | All Data Terms | $R_c^2$ | RMSEC | RMSECV | Outliers Removed Terms | Outliers | $R_c^2$ | RMSEC | RMSECV |
|---|---|---|---|---|---|---|---|---|---|---|
| Lamp angle (°) | | | | | | | | | | |
| 20 | 40 | 1 | 0.14 | 0.99 | 1.15 | 1 | 2 | 0.19 | 0.97 | 1.11 |
| 40** | 40 | 3 | 0.64 | 0.65 | 1.18 | 3 | 0 | 0.64 | 0.65 | 1.18 |
| 40SG | 40 | 3 | 0.05 | 1.16 | 1.40 | | | | | |
| 60 | 40 | 4 | 0.76 | 0.52 | 1.96 | 4 | 2 | 0.82 | 0.43 | 0.84 |
| 80 | 40 | 1 | 0.15 | 0.99 | 1.26 | 4 | 1 | 0.38 | 0.84 | 1.03 |
| | | | | | | | | | | |
| Number of Lights | | | | | | | | | | |
| 1 | 210 | 7 | 0.64 | 0.71 | 0.90 | 7 | 3 | 0.68 | 0.68 | 0.86 |
| 2 | 210 | 2 | 0.43 | 0.92 | 0.95 | 6 | 11 | 0.61 | 0.74 | 0.85 |
| 4* | 208 | 6 | 0.63 | 0.73 | 0.88 | 6 | 5 | 0.66 | 0.69 | 0.83 |
| | | | | | | | | | | |
| No. of Scans | | | | | | | | | | |
| 1 | 210 | 2 | 0.45 | 0.89 | 0.92 | 2 | 3 | 0.44 | 0.89 | 0.92 |
| 2 | 210 | 2 | 0.43 | 0.92 | 0.95 | 2 | 6 | 0.46 | 0.88 | 0.91 |
| 4* | 208 | 6 | 0.63 | 0.73 | 0.88 | 6 | 5 | 0.66 | 0.69 | 0.83 |
| 16a | 212 | 5 | 0.57 | 0.78 | 0.87 | 6 | 5 | 0.63 | 0.72 | 0.82 |
| 16b | 210 | 6 | 0.60 | 0.76 | 0.84 | 6 | 6 | 0.61 | 0.74 | 0.81 |
| | | | | | | | | | | |
| Shroud/ pathlength | | | | | | | | | | |
| Shroud on* | 208 | 6 | 0.63 | 0.73 | 0.88 | 6 | 5 | 0.66 | 0.69 | 0.83 |
| Collar on | 210 | 6 | 0.70 | 0.66 | 0.81 | 7 | 7 | 0.75 | 0.58 | 0.67 |
| Shroud off – fixed path | 213 | 2 | 0.45 | 0.89 | 0.91 | 2 | 9 | 0.47 | 0.87 | 0.89 |
| Shroud off – variable path | 196 | 4 | 0.50 | 0.84 | 0.93 | 3 | 15 | 0.51 | 0.82 | 0.87 |
| | | | | | | | | | | |
| Spectrometer Comparison | | | | | | | | | | |
| MMS1** | 40 | 3 | 0.64 | 0.65 | 1.18 | 3 | 0 | 0.64 | 0.65 | 1.18 |
| 6500 (- 2300) | 40 | 1 | 0.21 | 0.95 | 1.07 | 2 | 4 | 0.51 | 0.79 | 1.06 |
| 6500 (- 1100) | 40 | 4 | 0.59 | 0.69 | 0.97 | 3 | 3 | 0.71 | 0.55 | 0.84 |
| Perten (- 1700) | 40 | 3 | 0.64 | 0.65 | 1.18 | 3 | 0 | 0.64 | 0.65 | 1.18 |
| Perten (- 1050) | 40 | 1 | 0.97 | 0.19 | 1.07 | 1 | 0 | 0.97 | 0.19 | 1.07 |

## *Calibration mathematics*

The optimal mathematical treatment of spectral data is expected to be specific to the instrumentation and the application. For example, scatter correction routines (standard normal variance and detrend in the WINISI software) are typically applied to reflectance spectra of samples with a rough, light scattering surface. First and second derivative procedures are useful to remove changes in spectral baseline level and slope, and to highlight spectral features. The optimal value for the 'gap' (number of data points) over which the derivative is calculated will depend on the bandwidth of the spectral feature of interest, and the noise and wavelength resolution of the instrumentation used. Data smoothing routines can also be useful in the reduction of noise and elimination of redundant spectral information. An empirical 'test it and see'

approach is generally used to establish the best mathematical treatment for a given application.  For example, Guthrie and Walsh (1997b) established that a mathematical treatment involving second derivative over 4 data points and smoothing over 4 data points was optimal in the calibration of sugar content in intact pineapple fruit using Foss NIRSystems 6500 reflectance spectra.

Calibration mathematical treatment was optimised for the populations used in the comparison of optical geometry, with a standard treatment adopted (as used in Table 2).  The optimal optical configuration was then used to collect spectra of a large number of fruit (see below).  We report here (Table 3) an exercise in comparison of mathematical treatments on this larger population, which yielded similar conclusions but more marked differences than obtained with the smaller population sets.

Using the WINISI chemometric package, outlier spectra were detected and removed from the data population.  Approximately 5% of spectra were removed from populations, resulting in a consistent improvement in calibration performance (Table 2).  However, as removal of outlier data results in inconsistent population structure, this option was not employed when evaluating mathematical treatments (Table 3). The scatter correction routines (standard normal variance and detrend) decreased calibration performance in terms of RMSECV and $R_c^{\,2}$ (Table 3).  We attribute this result to the optical geometry of the system employed, which was effectively a transmission rather than a reflectance system.

**Table 3.** The effect of data treatments (derivative condition and gap size, smoothing, scatter correction) on the calibration of Zeiss MMS1 spectral data (700 – 1,050 nm) to melon flesh Brix. Spectral data is of a population of 1,991 fruit (mean 8.1, standard deviation 1.26° Brix). Values marked with an asterisk represent cases where the smoothing window was larger than the derivative gap window.

| Derivative | Data Treatment | | Scatter Correction (SNV and Detrend) Treatment | | | | Nil Scatter Correction | | | |
| | Gap Size | Smooth | Terms | $R_c^2$ | RMSEC | RMSECV | Terms | $R_c^2$ | RMSEC | RMSECV |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1* | 15 | 0.55 | 0.85 | 0.86 | 16 | 0.61 | 0.78 | 0.79 |
| 0 | 0 | 4* | 16 | 0.52 | 0.87 | 0.87 | 16 | 0.55 | 0.84 | 0.86 |
| 0 | 0 | 10* | 15 | 0.44 | 0.94 | 0.95 | 15 | 0.48 | 0.91 | 0.94 |
| 0 | 0 | 20* | 15 | 0.31 | 1.00 | 1.10 | 16 | 0.39 | 0.98 | 0.99 |
| 1 | 4 | 1 | 15 | 0.58 | 0.81 | 0.84 | 16 | 0.64 | 0.76 | 0.79 |
| 1 | 4 | 4 | 15 | 0.50 | 0.89 | 0.90 | 15 | 0.55 | 0.85 | 0.85 |
| 1 | 4 | 10* | 16 | 0.50 | 0.89 | 0.91 | 15 | 0.49 | 0.90 | 0.91 |
| 1 | 4 | 20* | 15 | 0.34 | 1.02 | 1.04 | 15 | 0.40 | 0.98 | 0.99 |
| 1 | 10 | 1 | 15 | 0.52 | 0.87 | 0.90 | 15 | 0.55 | 0.84 | 0.87 |
| 1 | 10 | 4 | 16 | 0.50 | 0.89 | 0.91 | 15 | 0.49 | 0.90 | 0.91 |
| 1 | 10 | 10 | 15 | 0.42 | 0.96 | 0.97 | 15 | 0.44 | 0.94 | 0.96 |
| 1 | 10 | 20* | 14 | 0.30 | 1.05 | 1.07 | 16 | 0.39 | 0.98 | 1.00 |
| 2 | 4 | 1 | 13 | 0.59 | 0.81 | 0.84 | 15 | 0.65 | 0.74 | 0.77 |
| 2 | 4 | 4 | 13 | 0.54 | 0.85 | 0.88 | 14 | 0.59 | 0.80 | 0.83 |
| 2 | 4 | 10* | 13 | 0.49 | 0.90 | 0.91 | 14 | 0.54 | 0.86 | 0.88 |
| 2 | 4 | 20* | 12 | 0.34 | 1.02 | 1.03 | 14 | 0.41 | 0.97 | 0.98 |
| 2 | 8 | 1 | 13 | 0.50 | 0.89 | 0.91 | 14 | 0.54 | 0.86 | 0.87 |
| 2 | 8 | 4 | 12 | 0.44 | 0.94 | 0.95 | 14 | 0.49 | 0.90 | 0.92 |
| 2 | 8 | 8 | 15 | 0.44 | 0.94 | 0.96 | 14 | 0.45 | 0.93 | 0.95 |
| 2 | 8 | 10* | 12 | 0.39 | 0.99 | 0.99 | 14 | 0.43 | 0.95 | 0.96 |
| 2 | 8 | 20* | 14 | 0.29 | 1.06 | 1.07 | 14 | 0.30 | 1.05 | 1.06 |
| 2 | 10 | 1 | 14 | 0.50 | 0.89 | 0.91 | 16 | 0.56 | 0.83 | 0.85 |

Various signal filtering routines may be used to remove noise from the acquired spectra (e.g. Fourier transform, box car averaging, Butterworth filter). Osborne *et al.* (1996) reported the use of a Butterworth filter improved the development of calibrations of kiwifruit Brix, based on spectra acquired using a Zeiss MMS1 spectrometer. As the Savitzky-Golay signal filter was available within the Zeiss MMS1 test software, we filtered spectra acquired at a lamp / detector angle of 40°. The calibration developed with this data was severely degraded in terms of regression coefficient and RMSEC (Table 2), and this treatment option was not considered further.

Calibration performance was marginally enhanced by the use of derivatised data (second derivative superior to first or no derivative). The best calibrations achieved using 'raw', first and second derivative data yielded a RMSECV of 0.79° Brix (no smoothing), 0.79° Brix (derivative gap of 4 data points, no smoothing) and 0.77° Brix (derivative gap of 4 data points, no smoothing), respectively. We attribute the relative lack of calibration response to the use of derivatives to the transmission optical geometry of the system employed. Derivatives should offer more value in reflectance systems in which baseline shifts between samples can be large.

Calibration performance was degraded as the gap over which the derivative was calculated was increased from 4 to 8 to 10 (Table 3). Calibration performance was also degraded by the smoothing of data. The Zeiss MMS1 has a pixel resolution of 3.3 nm, and thus smoothing or derivative calculated over 4 data points involves averaging of data over a 13 nm spectral range, equivalent to the wavelength resolution of the instrument. Further averaging will involve loss of spectral information. Also, the MMS1 has a high S/N ratio, such that the effect of signal averaging (smoothing) to signal precision may not contribute to improved calibration performance. Smoothing over a greater gap than used in the calculation of the derivative is expected to result in a loss of information, and a slight degradation in calibration performance was observed.

Based on the above observations, we recommend a data treatment of second derivative over a gap size of 4, with no scatter correction or smoothing, for this instrumentation and optical configuration.

*Design of field unit for consideration of calibration robustness*

In summary, the following instrument design is proposed for the non-invasive assessment of melon using a low cost, commercially available spectrometer module: the Zeiss MM1 spectrometer used with a shroud (45 mm diameter, 100 mm length) between the detector fibre optic and the fruit surface, and 4*50 Watt tungsten halogen lamps and lamps mounted at 90° intervals, positioned 100 mm above the fruit and aligned with the approximate centre of the fruit (i.e. angle of 45° between light incidence and detected area of fruit). These features have been adopted in a 'luggable' or 'at-line' NIR system that can be transported between pack-houses. This unit incorporates a spring-loaded platform to keep the fruit firmly against the detector shroud, while allowing for ease of fruit change over. The outside dimensions of the unit are 400 mm (width) by 400 mm (depth) by 550 mm (height), accommodating 1 melon at a time, and locating the spectrometer and other electronics above the sample chamber.

The following operational parameters are recommended for the use of this hardware: an integration time of 200 ms, as required to achieve a detector response at ca. 50% of saturation, with detector and lamp powered up 2 hours before use to ensure instrument stability. Averaging of 4 scans per spectra is recommended to improve S/N ratio. A 'default' data treatment of second derivative calculated over 4 data points, without further data pre-treatment, is suggested. We have subsequently developed a LabView (National Instruments, Sydney, Australia) based spectral acquisition and analysis package which allows application of calibrations to give predictions in a real-time basis.

This system described was benchmarked in terms of calibration performance against 2 commercial research-grade NIR spectrometers, operated over the full NIR

wavelength capability of these instruments and in reflectance mode (Table 2). The

purpose built instrumentation supported a calibration inferior to that achieved using

the Perten DA7000 (Springfield, USA) and Foss NIRSystems 6500 (Silver Spring,

USA) operated over a similar wavelength range. However, the purpose built

instrumentation is recommended as a field-suitable, low cost alternative. The

relatively good performance of the purpose built instrumentation is ascribed to the

low S/N ratio of the MMS1 detector and optimisation of the system in terms of optical

geometry of incident light, sample and detector. We will employ this system to

collect spectra of melons of various cultivars, growing districts and seasons to further

develop the consideration of Guthrie *et al*. (1998) of calibration robustness.

Harvested rockmelon fruit vary between 6 and 14° Brix, with 10° Brix commonly

accepted as a quality standard. Low cost instrumentation which supported a standard

error of prediction of less than 1° Brix on intact melons would find ready acceptance

in the Australian melon industry for pack-house grading of fruit. Further, equipped

with a robust calibration, this instrumentation should be useful in physiological,

agronomic and breeding programs targeting melon fruit soluble solids content.

# 3

# CALIBRATION MODEL DEVELOPMENT FOR MANDARIN FRUIT INTERNAL QUALITY ATTRIBUTES[2]

## ABSTRACT

The utility of near infra-red spectroscopy as a non-invasive technique for the assessment of internal eating quality parameters of mandarin fruit (*Citrus reticulata* var. 'Imperial') was assessed.  The calibration procedure for the attributes of total soluble solids (TSS) and dry matter (DM) was optimised with respect to a reference sampling technique, scan averaging, spectral window, data pre-treatment (in terms of derivative treatment and scatter correction routine) and regression procedure.  The recommended procedure involved sampling of an equatorial position on the fruit with 1 scan per spectrum, and modified partial least squares model development on a 720 to 950 nm window, pre-treated as first derivative absorbance data (gap size of 4 data points) with standard normal variance and detrend scatter correction.  Calibration model performance for the attributes of TSS and DM content was encouraging (typical $R_c^2$ of > 0.75 and 0.90 respectively; typical root mean square standard error of calibration of < 0.4 and 0.6% respectively),while that for juiciness and total

---

acidity were unacceptable. The robustness of the TSS and DM calibrations across new populations of fruit is documented in a companion study (Guthrie *et al.* 2005b).

## INTRODUCTION

Near infra-red spectroscopy (NIRS) has been applied to the sorting of intact fruit on total soluble solids (TSS) and dry matter (DM) content (Walsh *et al.* 2004), with commercial application to pack-house fruit sorting lines for the sorting of sweetness of citrus, apples, pears and peaches at 3 pieces per second per lane commencing in Japan in the mid 1990s (Kawano 1994a). Commercial application within pack-houses of Western countries is nascent.

Various statistical terms and abbreviations have been used by authors working with NIRS technology. In this manuscript the following terms and abbreviations have been employed: bias (*bias*) is the difference between mean of actual and predicted values, standard deviation (SD) of the reference method values, coefficient of determination on calibration data population ($R_c^2$), coefficient of determination on validation data population ($R_v^2$), root mean square error of calibration (RMSEC); root mean square error of cross validation (6 groups; without *bias* correction) (RMSECV), root mean square error of prediction (without *bias* correction) (RMSEP), and RMSEP corrected for *bias* (RMSEP(C)). The standard deviation ratio (SDR) is calculated as SD/(RMSECV or RMSEP).

Application of NIRS technology to a given fruit commodity requires an appreciation of the distribution of the attribute of interest within the fruit, as fruit is not internally homogenous. The mandarin fruit consists of an exocarp (skin) with numerous oil glands, a mesocarp (white pith), and an endocarp that produces extensions (juice sacs) that occupy space within the carpels. The juice sacs form the

primary edible material of the fruit. Miyamoto and Kitano (1995) noted Satsuma

mandarin TSS was greatest in the distal apex of the fruit, decreasing towards the

proximal (pedicel) end. The coefficient of variation of TSS within a single orange

fruit was reported as 10.2, 1.8 and 5.6% in the proximal to distal, around the fruit

circumference (at an equatorial position), and radial (from centre to skin, at an

equatorial position) orientations, respectively (Peiris *et al.* 1999b). This variation was

greater than that monitored in a peach and an apple fruit, but less than that in a melon

fruit (Peiris *et al.* 1999a). Near infra-red spectroscopic assessment of citrus fruit at an

equatorial position is therefore logical.

The use of NIRS to assess mandarin TSS has been reported by a number of

researchers. Kawano *et al.* (1993) developed multiple linear regression (MLR)

models using Satsuma mandarin in a transmittance sample geometry, and reported $R_c^2$

up to 0.98 and a RMSEC of 0.28% TSS, based on a population with SD of 1.75%

TSS. Miyamoto and Kitano (1995) also developed MLR models based on

transmittance spectra, and reported typical calibration statistics of $R_c^2$ of 0.88 and

RMSEC of 0.5% TSS, based on a population with SD of 1.50% TSS. Ou *et al.* (1997)

developed models using Ponkan mandarins in an interactance geometry, and reported

calibration statistics of $R_c^2$ of 0.52 to 0.74 and RMSEC of 0.41 to 0.64% TSS, for a

given harvest area (SD was not reported). Greensill and Walsh (2002) developed

partial least squares regression (PLS) models using Imperial mandarins in an

interactance geometry, and reported typical calibration statistics of RMSECV of

0.26% TSS using a population of SD 0.45% TSS. The $R_c^2$ (calculated in this instance

as 1 - (RMSECV/SD)$^2$) for these values was 0.67. McGlone *et al.* (2003) explored

the use of several optical configurations and wavelength windows for prediction of

TSS in Satsuma mandarin, reporting best results ($R_v^2$ of 0.93 and RMSEP of 0.32%

TSS) for a transmittance methodology operating between 700 and 930 nm. Results for an interactance geometry ($R_v^2$ of 0.85 and RMSEP of 0.47% TSS) were superior to that for a reflectance geometry ($R_v^2$ of 0.75 and RMSEP of 0.63% TSS).

Shiina *et al.* (1993), Onda *et al.* (1994), Sohn *et al.* (2000) and Schmilovitch *et al.* (2000) have reported various levels of success in measuring total acidity (TA) of intact pineapple, plum, apple and mango, respectively. With Ponkan mandarin, Ou *et al.* (1997) obtained $R_v^2$ of between 0.5 and 0.8 and RMSEP of 0.13 – 0.27% for TA, using a calibration developed on fruit from 1 district and used to predict TA for fruit from another 2 districts. Similarly, Miyamoto *et al.* (1998) used NIR transmittance spectra of intact Satsuma mandarins to achieve prediction of TA in separate populations (origin of each population not stated), with $R_v^2$ of 0.83, *bias* of 0.02% TA and RMSEP of 0.15% TA. However, McGlone *et al.* (2003) concluded that robust TA prediction was not possible in Satsuma mandarin, although $R_v^2$ of up to 0.65 and RMSEP of 0.15% TA could be achieved within a given population through a correlation with skin chlorophyll (fruit maturity).

Calibrations on fruit DM have been reported for kiwifruit and mango (McGlone and Kawano 1998; Guthrie and Walsh 1997b), fruit which store starch, with conversion to sugar at ripening. Typical calibration model statistics were $R_c^2$ of 0.96 and RMSEC of 0.79% DM. While citrus fruit do not store starch, DM content may be a useful index of certain internal quality defects. For example, Peiris *et al.* (1998a) reported the use of the second derivative of absorbance at 768 and 960 nm to identify fruit with section dryness disorder. These wavelengths are relevant to water absorption.

For spectra collected using a transmission optical geometry, it is expected that calibration model performance will be affected by fruit size. To address this issue,

Kawano *et al.* (1993) identified absorption at 844 nm as related to fruit (Satsuma mandarin) diameter, and normalised second derivative of absorbance data for all wavelengths used in the MLR model by dividing by the second derivative of absorbance at 844 nm. However, Miyamoto and Kitano (1995) reported that this procedure hindered prediction ability when sample temperature was varied. These authors reported that it was possible to compensate for optical pathlength by including the second derivative of absorbance data at, or near, the wavelengths of 740 or 840 nm as part of a 4 wavelength MLR calibration.

There are at least 6 prior reports on the application of NIRS to the assessment of TSS, 3 studies on TA and 1 study relevant to DM, in intact mandarins. However, these reports vary in the chemometric approaches used (MLR, PLS), data pre-processing techniques, spectral window, optical geometry, etc. In the current study we seek to confirm the utility of the NIRS method to the assessment of these characters, and optimise these variables in the development of a calibration model, prior to a companion study of robustness of the model across new populations (varying in growing location, time within a season and across seasons).

## MATERIALS AND METHODS

### Plant material and reference analyses

Mandarin fruit (Imperial variety) were sourced following commercial harvest from orchards in Mundubbera (25.6º S, 151.6º E), Bundaberg (24.9º S, 152.3º E) and Dululu (23.8º S, 150.3º E), Queensland. Fruit were sourced from 3 separate farms on 1 day, from separate harvests of 1 tree over a 14-day period, and from 1 pack-house over 4 seasons. In all, 20 populations of Imperial mandarins (each of approximately

100 fruit), obtained over different seasons, growing districts and different harvest times from the 1 tree, were used for spectra acquisition and then assessed for TSS. In addition to this, DM and juiciness of 6 separate populations and TA of 1 population were assessed. All populations were alphabetically named in chronological order.

Three populations of fruit were assessed in consideration of the distribution of TSS, DM and juiciness within fruit. The distribution of TSS and juiciness of the inner and outer section of each segment of 3 fruit was assessed (Population 1). Total soluble solids and DM of inner and outer of proximal, equatorial and distal parts of each of a further 5 fruit were assessed (Population 2). Finally, TSS and juiciness of the proximal and distal halves of each of 99 fruit were assessed (Population 3).

All fruit were halved, juiced, and TSS determined by refractometry (Bellingham and Stanley RMF 320). Total acidity was assessed by titration of a 5 mL sample of juice against 0.1M NaOH using phenolphthalein as an indicator. Dry matter of fruit halves (with skin), was determined by drying at 70°C to constant weight in a forced convection oven over 48 hours. Juiciness was estimated from the weight of juice expressed from a fruit half by a commercial juice extractor (juiciness % = weight of juice/weight of fruit * 100). Data were analysed using ANOVA to determine differences in attribute distribution, with testing of significance conducted at the 10% level.

## Spectroscopy

Spectra were collected over the wavelength range 306 – 1,130 nm using a NIR enhanced Zeiss MMS1 spectrometer and a 100 Watt tungsten halogen light in the interactance optical configuration reported by Greensill and Walsh (2000) ($0°$ angle between illumination and detected light rays, with detection probe viewing a shadow

cast by the probe onto the fruit).  Spectra were collected from 1 side of each fruit, on

the equator of the fruit, equidistant from proximal and distal ends with an integration

time of 30 milliseconds.  For comparison purposes, 1 population was also assessed

using the partial transmittance optical configuration used by Walsh *et al.* (2000) ($45^o$

angle between illumination source and detector, relative to the fruit centre, with

detector probe in contact with the fruit surface).  Various levels of spectral averaging

(1, 2, 4 and 32 scans per spectrum) were also undertaken on this population.  Spectra

were also collected from 1 population at 3 different fruit temperatures (10, 22 and

30ºC).

Spectra were collected as raw analogue to digital counts (15 bit), and converted

to absorbance values using an in-house developed software package.  These data were

then ported to the WINISI (ver. 1.04a) chemometric software package for derivative

calculations.  Examples of analogue to digital counts, absorbance and second

derivative spectra are given in Fig. 1.
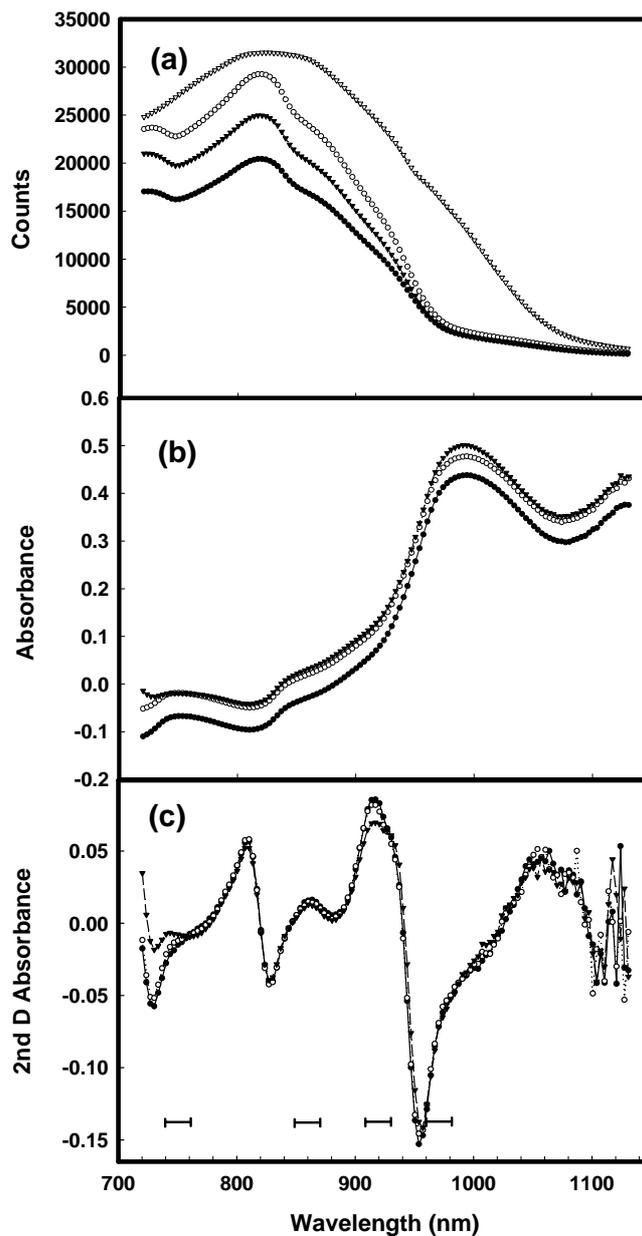
**Figure 1.** Near infra-red spectral data of 3 intact mandarins displayed as (a) analogue to digital counts including the white reference (open triangle), (b) absorbance, and (c) second derivative absorbance data. The mandarin fruit were chosen at random, representing fruit with high (12.7%, closed circle), medium (10.2%, open circle), and low (7.6%, closed triangle) total soluble solids. Horizontal bars show important spectral areas.

## Chemometrics

The software package WINISI (ver. 1.04a) was used for all chemometric analysis except as stated. This package calculates a derivative as a 'Norris regression' using start, central and end points only over a user definable 'gap' ($g$, wavelength range). The 'Norris regression' is calculated by the formula, $a - 2b + c$, where $b$ is the absorbance at wavelength, $\lambda$, and $a$ and $c$ are absorbances at wavelengths $\lambda$-$g$ and $\lambda$+$g$, respectively. In the data smoothing option, the absorbances at the 3 wavelengths used in the derivative calculation can be averaged with a user defined number of neighbouring absorbances. Available scatter corrections include standard normal variance (SNV) and detrend. Standard normal variance weights the absorbance at each wavelength by the SD of the calibration population. Detrending fits a least squares quadratic regression to successive wavelength windows. This curve is then subtracted from the spectrum to give the residual spectrum that is used in the subsequent calibration. Calibrations were developed using both step-wise multiple linear regression (MLR) and modified partial least squares regression (MPLS). Calibration performance was assessed in terms of $R_v^2$, RMSEP, SDR, slope and *bias* of the validation populations.

The Matlab PLS and WINISI MPLS techniques gave equivalent model results for a given data population (data not shown). The effect of spectral window on PLS calibration model performance for TSS and DM was optimised in terms of RMSEC using a PLS interval algorithm, developed in Matlab (ver. 7.0) – PLS toolbox (ver. 3.5 by Eigenvector). First derivative (WINISI gap size 4 with SNV and detrend scatter correction) absorbance data interpolated to 3 nm steps were used in this exercise. The wavelength range of the spectral windows varied in starting position from 700 – 930 nm, with an end position of 800 – 1,020 nm, in increments of 3 nm. The

combined populations J and K for TSS and population T for DM were used in this exercise.

Data pre-treatment procedures are generally optimised for a given application, with a range of derivative and scatter correction techniques trialled, and results (e.g. RMSEC, $R_c^2$, RMSECV, *bias* and RMSEP) 'eyeballed'. As an advance on this situation, Fearn (1996) recommended a protocol involving testing the significance of differences between both the RMSEP and *bias* of different models. Derivative condition, derivative gap size and data smoothing were considered in this study with reference to the use of both transmission and absorbance data. The procedure of Fearn (1996) was employed in this study, facilitated by a spreadsheet (see Appendix F) which 'automated' the procedure. A significance level of 95% was used in these tests.

# RESULTS

### *Attribute distribution*

In general, TSS content increased marginally (less than 1% TSS), albeit significantly, from proximal to distal ends of the fruit, and decreased from skin to core of the fruit (Table 1). Variation in mean TSS, however, differed among populations. In Population 1, the external region TSS (8.9%) significantly exceeded that of the internal region (8.3%). In Population 2, external region TSS was marginally greater than the internal region at the proximal end (8.4 cf. 8.3%), was not different at the equatorial region, and was less at the distal end (8.3 cf. 8.5%). The maximum difference in mean TSS among the combinations of proximal, equatorial and distal, and internal and external portions in Population 2 was only 0.2% while in Population 3, TSS at the distal end was 0.7% units greater than at the proximal end. The

coefficient of variation for TSS (over 10 segments in a single fruit) was 1.2 and 2.1%
in the proximal-distal and equatorial circumference orientations, respectively.

Dry matter varied by less than 1% DM between internal and external regions
(Population 2, Table 1). Dry matter did not differ significantly among proximal,
equatorial and distal positions for external tissue, while for internal tissue DM at the
proximal position was less than the other positions and less than DM in external
tissue. The coefficient of variation for DM was approximately 3% in the proximal-
distal orientation, and 2% in the equatorial-circumference orientation. Juiciness did
not vary between proximal and distal ends (Population 3), but varied between internal
and external regions of the fruit (Population 1). The % DM content was not
correlated to % juiciness (correlation coefficient of 0.02 for a combination of 5
populations, $n = 379$).

**Table 1.** **Spatial distribution of total soluble solids, dry matter and juiciness for 3 populations of mandarin fruit. For Population 1 (3 fruit with an average of 10 segments per fruit), each segment was cut longitudinally into external and internal sections. For Population 2 (5 fruit), each segment was cut longitudinally into external and internal portions, and into proximal, equatorial and distal sections. For Population 3 (99 fruit), each fruit was cut transversely through the equator of the fruit, into proximal and distal halves. Means within a population and attribute not followed by a common letter are significantly different (P = 0.05).**

| Attribute | Population 1 | Population 2 | | | Population 3 | |
|---|---|---|---|---|---|---|
| | | prox | equat | distal | prox | distal |
| **TSS (%)** | | | | | | |
| **External** | 8.9 a | 8.4 a | 8.4 ab | 8.3 a | | |
| **Internal** | 8.3 b | 8.3 b | 8.4 ab | 8.5 c | | |
| **Combined** | | | | | 9.2 a | 9.9 b |
| | | | | | | |
| **Juiciness (%)** | | | | | | |
| **External** | 48.2 a | | | | | |
| **Internal** | 38.8 b | | | | | |
| **Combined** | | | | | 55 a | 55.6 a |
| | | | | | | |
| **DM (%)** | | | | | | |
| **External** | | 9.7 b | 9.6 b | 9.7 b | | |
| **Internal** | | 9.1 a | 9.7 b | 9.5 b | | |

### *Instrumentation and sample presentation*

For 1 population of fruit, spectra were collected using 2 optical geometries, as used by Walsh *et al.* (2000) and Greensill and Walsh (2000). The TSS calibration model developed for the 45ºgeometry was not significantly better in terms of RMSECV than any of the 0º geometries (Table 2) and therefore the 0º geometry was used in all other studies reported here.

Increasing the number of scans averaged per spectra from 1 to 32 did not significantly improve TSS calibration performance in terms of RMSECV when the fruit were positioned equatorially with a 0º optical geometry, although there was a trend towards improved performance (e.g. RMSECV decreased from 0.32 for 1 scan to 0.26 for 32 scans) (Table 2). Calibration model performance on TSS was degraded

if spectra were collected from the proximal end of the fruit, but were similar for equatorial and distal positions (Table 2).

**Table 2.**　　　**The influence of scan averaging, optical geometry and fruit positioning on calibration statistics (coefficient of determination of calibration population ($R_c^2$); root mean square error of cross validation (RMSECV); and standard deviation ratio (SDR)) for total soluble solids (TSS).  A single population of fruit ($n = 97$, mean = 9.6% TSS and standard deviation = 0.77% TSS) was scanned with either 1, 2, 4 or 32 scans averaged per spectrum.  Letters following RMSECV values refer to significance testing at a 95% probability level, relative to the $0^o$ 1 scan treatment.**

| Optics | Fruit position | No. of scans | $R_c^2$ | RMSECV | Terms | SDR |
|--------|---------------|--------------|---------|--------|-------|-----|
| $0^o$ | Equatorial | 1 | 0.87 | 0.32 a | 7 | 2.4 |
| $0^o$ | Equatorial | 2 | 0.88 | 0.32 a | 7 | 2.4 |
| $0^o$ | Equatorial | 4 | 0.87 | 0.34 a | 7 | 2.4 |
| $0^o$ | Equatorial | 32 | 0.92 | 0.26 a | 7 | 2.9 |
| $45^o$ | Equatorial | 4 | 0.91 | 0.39 a | 9 | 2 |
| $0^o$ | Proximal | 4 | 0.68 | 0.63 b | 5 | 1.2 |
| $0^o$ | Distal | 4 | 0.88 | 0.30 a | 6 | 2.6 |

*Spectral window selection*

The effect of spectral window on PLS calibration model performance for TSS and DM was optimised in terms of RMSEC using a PLS interval algorithm.  Low RMSEC values for both TSS and DM were obtained for a window beginning between 703 and 850 nm, and finishing between 906 and 950 nm (Fig. 2).  The minimum RMSEC for TSS (0.26%) was recorded for a start wavelength of 703 nm and a finish wavelength of 911 nm.  The minimum RMSEC for DM (0.34%) was recorded for a start wavelength of 703 nm and a finish wavelength of 920 nm.  All work reported in this manuscript used the region 720 – 950 nm and therefore was similar to the observed optimal spectral window.
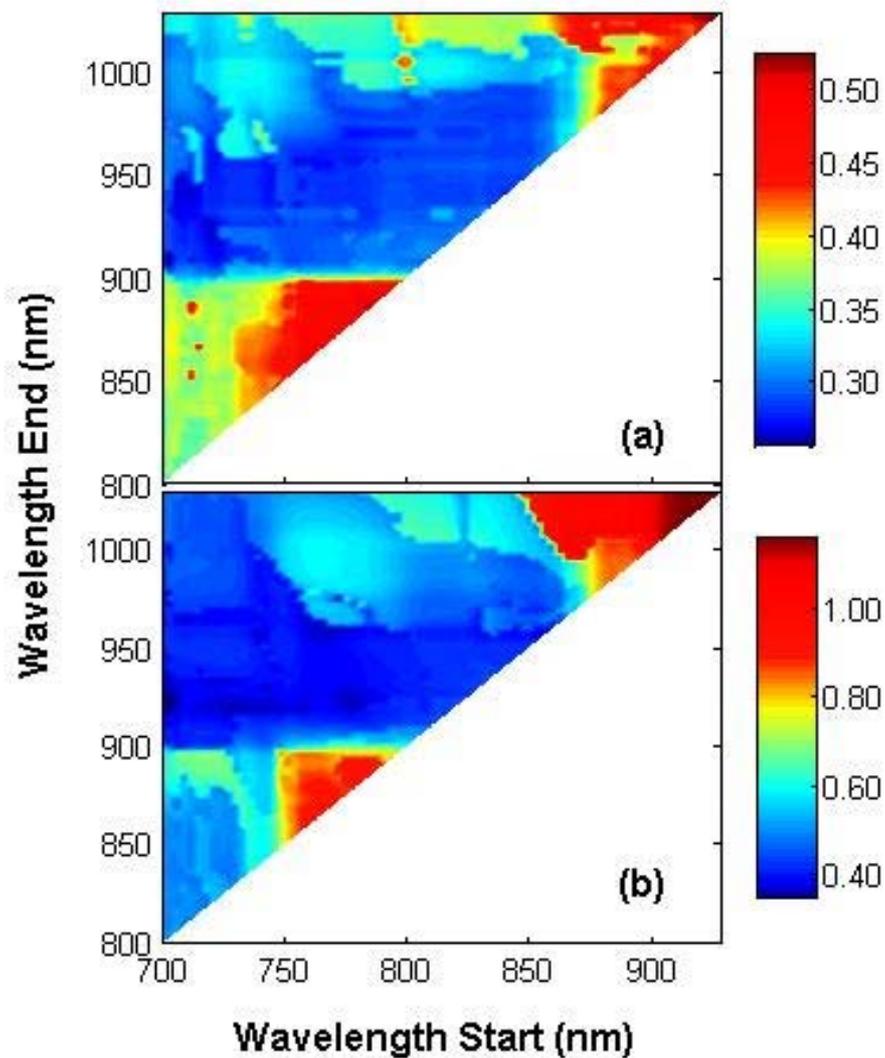
**Figure 2.** Calibration model performance as assessed by root mean square standard error of calibration (RMSEC) for varying spectral windows (varying start and end wavelengths). Partial least squares calibration models for (a) total soluble solids (TSS) for populations J and K combined and (b) dry matter for population T. The colour code to the RMSEC values (% TSS) is shown in the bar scale to the right.

*Spectral data treatment for MPLS*

The data pre-treatment method was optimised for MPLS regression in terms of data type (transmission or absorbance), derivative condition (nil, first or second order), derivative treatment (gap size), smoothing interval and scatter correction for both TSS and DM. Comparison was made on the basis of calibration statistics ($R_c^2$, RMSECV, number of terms and SDR) and prediction statistics (variance, RMSEP and *bias*). The calibration models were used to predict TSS of an independent population drawn from the same harvest as the calibration group (Table 3), TSS for 5 independent populations (Populations A, C, J, L and M; Table 4) and DM for 5 independent populations (Populations T, S, X, W and V; Table 5).

The highest $R_c^2$ and lowest RMSECV (0.95 and 0.35% TSS, respectively) for absorbance was recorded with a second derivative, gap size of 4 and no smoothing, while low RMSEP values were recorded for first derivative absorbance data with a small gap size (say $\leq 5$) and smoothing (Table 3). Overall however, calibration models developed on raw or derivative transmission or absorbance data were similar in calibration and validation performance (Table 3). Calibration model performance on TSS was remarkably insensitive to variations in the gap size used in derivatising spectral data (from 3 to 30 data points, or approximately 10 to 100 nm, either side of the data point) (Table 3). In this respect, second order derivatives appeared more sensitive than first order, with a decrease in calibration and validation performance at gap sizes greater that 15 data points. Model performance was also relatively insensitive to smoothing interval (Table 3).

**Table 3.** The influence of derivative treatment (order of derivative), derivative gap size and smoothing on calibration model performance for total soluble solids (TSS) as assessed by coefficient of determination of calibration population ($R_c^2$), root mean square error of cross validation (RMSECV), the number of terms in the model and standard deviation ratio (SDR), and validated by root mean square error of prediction (RMSEP) and *bias*. Calibration models were developed from Population L ($n = 71$, mean = 9.7% TSS, SD = 1.00% TSS) and validated on a subset ($n = 10$) of samples drawn from the original population prior to calibration development. No outliers were removed during calibration development.

| Data type | Derivative | Gap | Smoothing interval | $R_c^2$ | RMSECV | Number of terms | SDR | RMSEP | *bias* |
|---|---|---|---|---|---|---|---|---|---|
| *Transmission* | 0 | | 1 | 0.89 | 0.42 | 7 | 2.4 | 0.49 | 0 |
| | 1 | 4 | 1 | 0.91 | 0.39 | 7 | 2.6 | 0.43 | -0.64 |
| | 2 | 4 | 1 | 0.86 | 0.43 | 3 | 2.3 | 0.54 | 0.11 |
| | | | | | | | | | |
| *Absorbance* | 0 | | 1 | 0.86 | 0.44 | 6 | 2.3 | 0.52 | 0.11 |
| | 1 | 4 | 1 | 0.88 | 0.43 | 6 | 2.3 | 0.45 | 0.06 |
| | 2 | 4 | 1 | 0.88 | 0.41 | 4 | 2.4 | 0.47 | 0.04 |
| | | | | | | | | | |
| *Absorbance* | 1 | 3 | 1 | 0.89 | 0.43 | 6 | 2.3 | 0.47 | 0.08 |
| | 1 | 4 | 1 | 0.88 | 0.43 | 6 | 2.3 | 0.45 | 0.06 |
| | 1 | 5 | 1 | 0.87 | 0.44 | 6 | 2.3 | 0.44 | 0.08 |
| | 1 | 7 | 1 | 0.89 | 0.41 | 7 | 2.4 | 0.45 | 0.05 |
| | 1 | 9 | 1 | 0.88 | 0.43 | 7 | 2.3 | 0.49 | 0.02 |
| | 1 | 10 | 1 | 0.86 | 0.43 | 6 | 2.3 | 0.45 | 0.11 |
| | 1 | 15 | 1 | 0.84 | 0.44 | 7 | 2.3 | 0.49 | 0.14 |
| | 1 | 20 | 1 | 0.9 | 0.44 | 9 | 2.3 | 0.5 | 0.04 |
| | 1 | 30 | 1 | 0.85 | 0.49 | 8 | 2.1 | 0.48 | 0.15 |
| | | | | | | | | | |
| | 2 | 3 | 1 | 0.9 | 0.4 | 4 | 2.2 | 0.57 | 0.09 |
| | 2 | 4 | 1 | 0.95 | 0.35 | 7 | 2.5 | 0.53 | -0.01 |
| | 2 | 5 | 1 | 0.9 | 0.39 | 6 | 2.2 | 0.44 | 0.08 |
| | 2 | 7 | 1 | 0.9 | 0.38 | 5 | 2.5 | 0.5 | 0.01 |
| | 2 | 9 | 1 | 0.91 | 0.36 | 8 | 2.7 | 0.53 | -0.06 |
| | 2 | 10 | 1 | 0.92 | 0.37 | 7 | 2.6 | 0.44 | 0 |
| | 2 | 15 | 1 | 0.88 | 0.39 | 8 | 2.5 | 0.47 | 0.05 |
| | 2 | 20 | 1 | 0.89 | 0.43 | 8 | 2.2 | 0.53 | 0.25 |
| | 2 | 30 | 1 | 0.82 | 0.48 | 6 | 2 | 0.72 | 0.36 |
| | | | | | | | | | |
| | 1 | 4 | 1 | 0.88 | 0.43 | 6 | 2.3 | 0.45 | 0.06 |
| | 1 | 4 | 2 | 0.87 | 0.42 | 6 | 2.4 | 0.44 | 0.07 |
| | 1 | 4 | 3 | 0.89 | 0.42 | 7 | 2.4 | 0.42 | 0.05 |
| | 1 | 4 | 4 | 0.86 | 0.43 | 6 | 2.3 | 0.44 | 0.08 |
| | 1 | 4 | 5 | 0.88 | 0.43 | 7 | 2.4 | 0.43 | 0.07 |
| | 1 | 4 | 6 | 0.88 | 0.41 | 8 | 2.4 | 0.44 | 0.05 |
| | 1 | 4 | 10 | 0.87 | 0.42 | 7 | 2.4 | 0.45 | 0.05 |
| | 1 | 4 | 20 | 0.91 | 0.42 | 12 | 2.4 | 0.45 | 0.11 |

**Table 4.** Optimisation of data pre-treatment in terms of derivative treatment (none, first or second order) and 4 scatter correction routines (none, standard normal variance (SNV), detrend, or SNV and detrend combined) (a total of 12 treatments) for total soluble solids (TSS) calibration. Model performance is reported in terms of prediction of 5 independent populations of mandarin fruit (Populations A, C, J, L and M - fruit harvested on different days or locations to that used in the calibration). Calibration population statistics were $n$ = 81, mean = 9.6% TSS, standard deviation (SD) = 1.03% TSS and range of 8.2 - 12.3% TSS.

For each treatment prediction group within a population, the treatment with the lowest overall RMSEP was selected and the RMSEP presented in bold (but not underlined). The corresponding *bias* was also bolded. The lowest RMSEP within the other 2 derivative treatments was then selected and presented, along with the corresponding *bias*, with an underline (but not bolded). The RMSEP (or *bias*) in bold was then compared with each underlined RMSEP (or *bias*) in the population. If the values were significantly different at a 95% probability level then the underlined value was bolded so that bold and underline values differed significantly from the lowest value.

| | Variance (RMSEP) | | | *bias* | | |
|---|---|---|---|---|---|---|
| **Scatter Correction** | **Absorb.** | **1st Deriv.** | **2nd Deriv.** | **Absorb.** | **1st Deriv.** | **2nd Deriv.** |
| *Population A* | | | | | | |
| None | 0.398 | 0.433 | 0.411 | 0.059 | -0.171 | -0.031 |
| SNV | <u>0.386</u> | 0.409 | 0.357 | **<u>0.092</u>** | -0.096 | -0.045 |
| Detrend | 0.492 | 0.402 | 0.36 | -0.176 | -0.087 | -0.02 |
| SNV & Detrend | 0.424 | <u>0.385</u> | **0.332** | 0.082 | <u>-0.011</u> | **-0.048** |
| *Population C* | | | | | | |
| None | 0.762 | 0.888 | 0.77 | -0.485 | -0.725 | -0.561 |
| SNV | 0.973 | 0.867 | 0.76 | -0.638 | -0.674 | -0.491 |
| Detrend | 0.96 | 0.849 | <u>0.648</u> | -0.728 | -0.605 | <u>-0.295</u> |
| SNV & Detrend | **<u>0.707</u>** | **0.611** | 0.686 | **<u>-0.432</u>** | **-0.335** | -0.432 |
| *Population J* | | | | | | |
| None | 0.605 | 0.799 | 0.628 | 0.346 | 0.651 | 0.432 |
| SNV | 0.999 | 0.754 | 0.505 | 0.856 | 0.611 | 0.226 |
| Detrend | **<u>0.579</u>** | 0.476 | 0.532 | **<u>0.352</u>** | 0.193 | 0.304 |
| SNV & Detrend | 0.64 | **0.415** | <u>0.45</u> | 0.448 | **0.001** | **<u>0.128</u>** |
| *Population L* | | | | | | |
| None | 0.635 | 0.693 | 0.66 | 0.477 | 0.505 | 0.468 |
| SNV | 0.695 | 0.607 | 0.752 | 0.543 | 0.434 | 0.581 |
| Detrend | **0.501** | 0.648 | 0.713 | **0.17** | 0.448 | 0.579 |
| SNV & Detrend | 0.631 | <u>0.526</u> | **0.606** | 0.474 | **<u>0.355</u>** | **<u>0.424</u>** |
| *Population M* | | | | | | |
| None | **<u>0.655</u>** | 0.732 | 0.498 | **<u>0.413</u>** | 0.503 | -0.102 |
| SNV | 0.786 | 0.781 | <u>0.475</u> | 0.593 | 0.618 | <u>-0.039</u> |
| Detrend | 0.938 | 0.542 | 0.494 | 0.753 | 0.196 | 0.119 |
| SNV & Detrend | 0.688 | **0.462** | 0.5 | 0.54 | **0.007** | **-0.248** |

80

**Table 5.** Optimisation of data pre-treatment in terms of derivative treatment (none, first or second order) and 4 scatter correction routines (none, standard normal variance (SNV), detrend, or SNV and detrend combined) (a total of 12 treatments) for dry matter (DM) calibration. Model performance is reported in terms of prediction of 5 independent populations of mandarin fruit (Populations T, X, S, W and V - fruit harvested on different days or locations to that used in the calibration). Calibration population statistics were $n = 106$, mean = 14.7% DM, standard deviation = 1.83% DM and range of 14.7 - 19.2% DM.

For each treatment prediction group within a population, the treatment with the lowest overall RMSEP was selected and the RMSEP presented in bold (but not underlined). The corresponding *bias* was also bolded. The lowest RMSEP within the other 2 derivative treatments was then selected and presented, along with the corresponding *bias*, with an underline (but not bolded). The RMSEP (or *bias*) in bold was then compared with each underlined RMSEP (or *bias*) in the population. If the values were significantly different at a 95% probability level then the underlined value was bolded so that bold and underline values differed significantly from the lowest value.

| | Variance (RMSEP) | | | *bias* | | |
| --- | --- | --- | --- | --- | --- | --- |
| Scatter Correction | Absorb. | 1st Deriv. | 2nd Deriv. | Absorb. | 1st Deriv. | 2nd Deriv. |
| *Population T* | | | | | | |
| None | 1.075 | 0.937 | 0.909 | 0.814 | -0.416 | 0.074 |
| SNV & Detrend | <u>0.749</u> | <u>0.761</u> | 0.792 | **<u>0.224</u>** | **<u>-0.188</u>** | -0.202 |
| SNV | 0.777 | 0.786 | **0.744** | 0.396 | -0.34 | **-0.029** |
| Detrend | 0.958 | 0.889 | 0.766 | -0.327 | 0.147 | -0.148 |
| *Population X* | | | | | | |
| None | 0.842 | 0.693 | 0.681 | -0.324 | -0.231 | -0.033 |
| SNV & Detrend | <u>0.675</u> | **0.614** | <u>0.642</u> | **<u>0.208</u>** | 0.042 | **<u>-0.038</u>** |
| SNV | 0.757 | 0.669 | 0.669 | -0.325 | -0.235 | -0.054 |
| Detrend | 0.712 | 0.744 | 0.716 | 0.266 | 0.035 | -0.248 |
| *Population S* | | | | | | |
| None | 1.057 | 0.885 | 0.918 | -0.607 | -0.408 | -0.484 |
| SNV & Detrend | <u>0.806</u> | **0.805** | 0.904 | **<u>0.038</u>** | -0.043 | -0.417 |
| SNV | 0.947 | 0.837 | **<u>0.897</u>** | -0.464 | -0.153 | **<u>-0.402</u>** |
| Detrend | 0.832 | 0.858 | 0.909 | -0.048 | -0.31 | -0.447 |
| *Population W* | | | | | | |
| None | 1.031 | 1.181 | 1.062 | 0.314 | -0.411 | -0.021 |
| SNV & Detrend | 0.906 | 1.141 | 0.948 | 0.248 | 0.713 | 0.457 |
| SNV | **0.849** | **<u>0.96</u>** | <u>0.867</u> | **0.081** | **<u>0.334</u>** | <u>0.163</u> |
| Detrend | 1.257 | 1.346 | 1.225 | -0.217 | 0.697 | 0.683 |
| *Population V* | | | | | | |
| None | 5.367 | 4.423 | 4.979 | 5.249 | 4.239 | 4.824 |
| SNV & Detrend | 3.896 | **<u>3.805</u>** | **<u>3.532</u>** | 3.643 | **<u>3.552</u>** | **<u>3.253</u>** |
| SNV | 4.521 | 4.169 | 4.597 | 4.333 | 3.937 | 4.383 |
| Detrend | **2.982** | 4.393 | 4.102 | **2.681** | 4.229 | 3.921 |

The optimal mathematical treatment for the TSS model development based on prediction performance by comparison of RMSEP and *bias* involved a first derivative

absorbance treatment with standard normal variance (SNV) and detrend scatter correction, although there was little difference between first and second derivative treatments (Table 4). There was no clear requirement for derivative or scatter correction for DM calibrations (Table 5). In all further chemometric analysis reported in this study, unsmoothed first derivative absorbance data calculated using a gap of 4 data points, with standard normal variance and detrend scatter correction, were used.

The RMSEP and *bias* for TSS and DM (averaged for the 5 populations of Table 4 and 5, respectively) for models using a mathematical treatment of first derivative, SNV and detrend was 0.48% and 0.14% TSS, and 0.77% and 0.25% DM, respectively.

### *Multivariate regression analysis*

Calibration models, using the data treatment identified above, were developed for 17 populations harvested in 2001, a population each from 1999, 2000 and 2004 and a combination of populations from 2001 (Table 6). Generally, better calibration results ($R_c^2$ and RMSEC) for TSS were achieved using MPLS regression than stepwise MLR. Modified partial least squares calibration model statistics varied among populations, ranging from $R_c^2 = 0.41$ to $R_c^2 = 0.91$, with RMSEC varying between 0.45 and 0.22% TSS (Table 6).

**Table 6.** Calibration model statistics for total soluble solids in each of 17 populations (A to N) of mandarin fruit harvested over 2001, populations A to N combined and populations from 3 other years. Calibration models were developed using modified partial least squares and stepwise multiple linear regression, with a data pre-treatment of first derivative, standard normal variance and detrend.

| Population | Population statistics | | | MPLS calibration | | | | MLR calibration | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n$ | *Mean* | SD | $R_c^{\,2}$ | RMSEC | No. of terms | SDR | RMSEC | $R_c^{\,2}$ | No. of terms | SDR |
| A | 56 | 9.6 | 0.72 | 0.76 | 0.36 | 7 | 1.6 | 0.41 | 0.77 | 6 | 2.1 |
| B | 60 | 9.2 | 0.53 | 0.73 | 0.27 | 7 | 1.4 | 0.31 | 0.66 | 4 | 1.7 |
| C | 41 | 8.5 | 0.78 | 0.66 | 0.45 | 5 | 1.4 | 0.46 | 0.58 | 3 | 1.6 |
| D | 58 | 9 | 0.53 | 0.41 | 0.41 | 5 | 1.2 | 0.48 | 0.17 | 1 | 1.1 |
| E | 78 | 9.6 | 1 | 0.91 | 0.31 | 7 | 2.7 | 0.76 | 0.42 | 2 | 1.3 |
| F | 98 | 9.8 | 0.45 | 0.65 | 0.27 | 7 | 1.5 | 0.34 | 0.53 | 4 | 1.5 |
| G | 94 | 9.7 | 0.6 | 0.86 | 0.22 | 8 | 2.2 | 0.3 | 0.74 | 5 | 2 |
| H | 78 | 9.7 | 0.55 | 0.74 | 0.28 | 7 | 1.5 | 0.34 | 0.61 | 4 | 1.6 |
| I | 91 | 9.3 | 0.54 | 0.63 | 0.33 | 7 | 1.4 | 0.45 | 0.28 | 3 | 1.2 |
| J | 75 | 10.3 | 0.91 | 0.87 | 0.33 | 7 | 2.3 | 0.46 | 0.73 | 3 | 1.9 |
| K | 76 | 9.2 | 0.63 | 0.83 | 0.26 | 6 | 2 | 0.51 | 0.42 | 2 | 1.3 |
| L | 75 | 9.9 | 0.97 | 0.87 | 0.35 | 5 | 2.5 | 0.39 | 0.86 | 4 | 2.7 |
| M | 78 | 9 | 0.74 | 0.87 | 0.27 | 7 | 2.2 | 0.74 | 0 | 1 | 1 |
| N | 72 | 9.1 | 0.57 | 0.68 | 0.32 | 5 | 1.6 | 0.42 | 0.62 | 3 | 1.6 |
| O | 95 | 9.2 | 0.57 | 0.8 | 0.27 | 8 | 1.7 | 0.45 | 0.38 | 2 | 1.3 |
| P | 77 | 9.3 | 0.82 | 0.88 | 0.28 | 6 | 2.7 | 0.32 | 0.85 | 5 | 2.6 |
| Q | 89 | 9.5 | 0.66 | 0.84 | 0.27 | 7 | 2.1 | 0.33 | 0.76 | 4 | 2 |
| A to N | 770 | 9.4 | 0.87 | 0.84 | 0.35 | 10 | 2.4 | 0.5 | 0.69 | 9 | 1.8 |
| 1999 | 199 | 10.6 | 0.96 | 0.88 | 0.33 | 9 | 3.2 | 0.34 | 0.87 | 8 | 2.8 |
| 2000 | 100 | 8.4 | 1.05 | 0.88 | 0.36 | 9 | 2.9 | 0.38 | 0.87 | 6 | 2.8 |
| 2004 | 100 | 10.4 | 1.32 | 0.91 | 0.39 | 8 | 3.4 | 0.4 | 0.91 | 7 | 3.3 |

**Table 7.** Prediction statistics for the validation of modified partial least squares (MPLS), multiple linear regression (MLR) and 'forced' MLR (FMLR) models on mandarin total soluble solids (TSS), developed on populations A to N combined, and A and E individually (see Table 3), on 3 independent validation populations (O, P and Q). In the FMLR model, the model was restricted to using wavelengths of 760, 884 and 910 nm.

| Validation populations | Regression analysis | Calibration populations | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | $R_v^{2}$ | | | bias | | |
| | | A to N | A | E | A to N | A | E |
| O | MPLS | 0.59 | 0.47 | 0.49 | 0.48 | 0.19 | 0.06 |
| | MLR | 0.57 | 0.38 | 0.25 | 0.15 | -0.17 | -0.82 |
| | FMLR | 0.54 | 0.33 | 0.45 | 0.62 | 0.56 | 1.06 |
| P | MPLS | 0.81 | 0.78 | 0.78 | 0.13 | -0.08 | 0.61 |
| | MLR | 0.71 | 0.29 | 0.26 | 0.67 | -0.51 | 0.2 |
| | FMLR | 0.68 | 0.73 | 0.72 | 0.14 | 1.09 | 0.64 |
| Q | MPLS | 0.73 | 0.73 | 0.68 | 0.23 | 0 | 0.76 |
| | MLR | 0.57 | 0.25 | 0.1 | 0.84 | -0.33 | 0.16 |
| | FMLR | 0.6 | 0.55 | 0.64 | 0 | 0.86 | 0.55 |

Calibration models developed on the combined (A to N) populations and on 2 individual populations (A and E, chosen for low RMSEC and $R_c^{2}$) were validated with independent populations (O, P and Q). Modified partial least squares calibration models were also superior (in terms of $R_v^{2}$) to MLR calibration models in the prediction of TSS (Table 7) for the individual calibration populations (A and E) and the combined population (A to N). Multiple linear regression model validation performance ($R_v^{2}$) was generally improved (Table 7) by restricting model development to spectral windows of relevance to sugar ('forced' MLR (FMLR), using 760, 884 and 910 nm wavelengths). Adding calibration groups together marginally improved MPLS and MLR model validation ($R_v^{2}$) of new populations. No method was consistently better in terms of minimising the *bias* of the validated values.

*Calibration model performance for the attributes of TSS, DM, Juiciness and TA*

Typical calibration model statistics ($R_c^2$ and RMSEC) for TSS in a given population were > 0.75 and < 0.4%, respectively (Table 6), and for DM were 0.9 and 0.6%, respectively (data not shown). In contrast, model performance with respect to TA was poor, with $R_c^2$ of 0.3, and a RMSEC of 0.2% recorded for a population of mean 0.67% and SD of 0.19% (data not shown). Calibration model performance over 5 populations (A, C, J, L and M) was also poor with respect to % juiciness ($R_c^2 < 0.2$, RMSEC > 5.0%, for population means ranging from 47 to 52% and SD from 4 to 9%).

Typical MPLS model *b* coefficients (regression coefficients for the model) for TSS and DM models are illustrated in Fig. 3. Stepwise MLR coefficients for models developed on the same data were based on 860, 870 and 900 nm wavelengths for TSS, and 907, 890, and 780 nm wavelengths for DM.
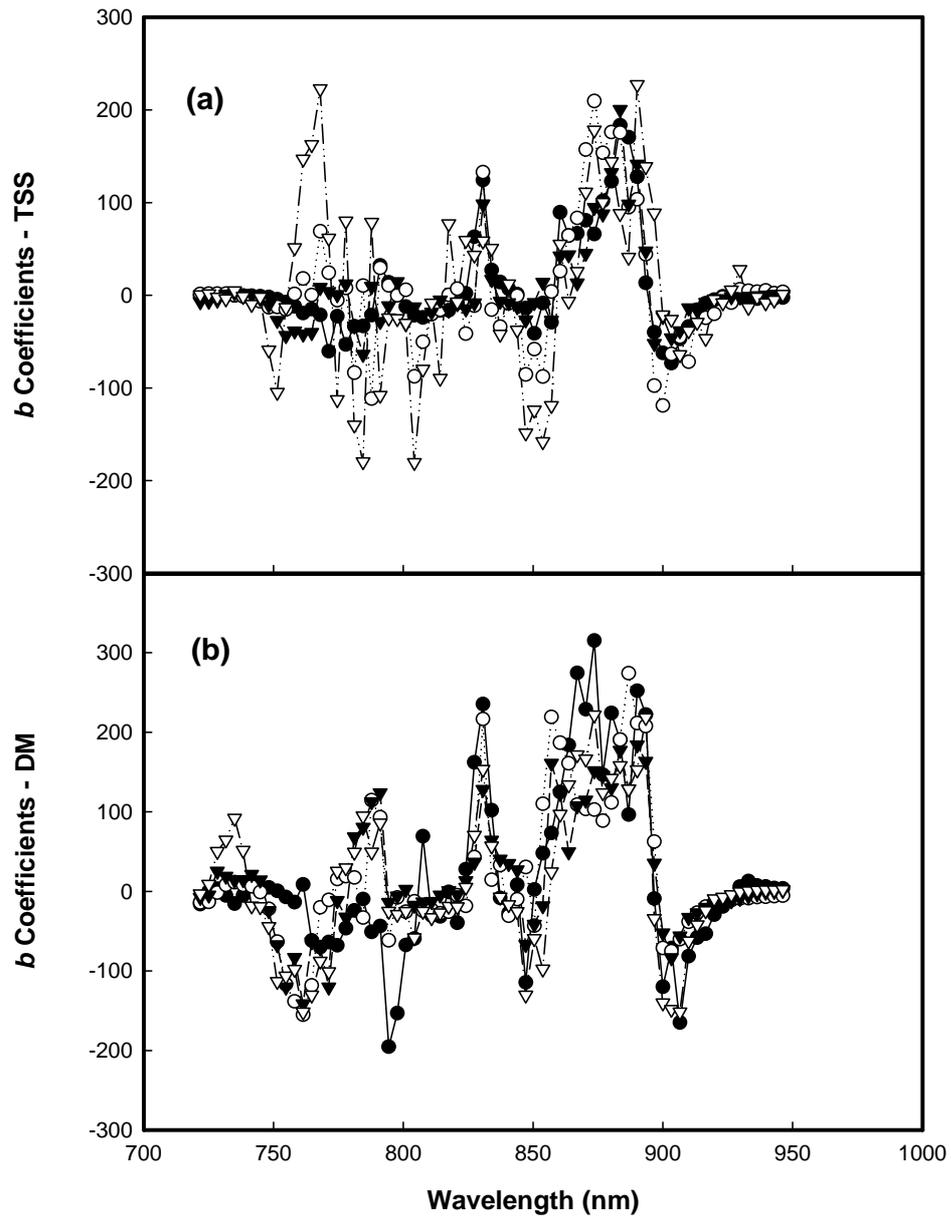
**Figure 3.** Modified partial least squares calibration *b* (regression) coefficients for (a) total soluble solids (TSS) and (b) dry matter (DM) models using second derivative of absorbance with a gap size of 4 points. Model coefficients for populations E, G, L, and P are shown for TSS, and those for populations R, S, T and X are shown for DM.

# DISCUSSION

## *Sample orientation*

The distribution and level of attributes, such as TSS, within a fruit may differ with maturation of the fruit, growing conditions of the fruit (e.g. position within the canopy), and/or size of the fruit. The TSS, juiciness and DM tended to be higher in external (relative to internal) and in distal (relative to proximal) tissue (Table 1). However, the absolute variation in any attribute level with reference to position within the fruit was low (maximum difference of 0.7% TSS and 0.6% DM). There was less variation at the equatorial than proximal or distal positions. The coefficient of variation (CV) for TSS around the equator of the fruit (outer tissue) was 2.1%, similar to that of 1.8% reported by Peiris *et al.* (1999b) for a single orange and grapefruit. However, Peiris *et al.* (1999b) reported greater CV values for proximal to distal variation (10.2% for orange and 12.4% for grapefruit (single fruit in each case)) than we noted for mandarin (2.1%). Mandarin fruit are apparently more homogenous than oranges or grapefruit.

Calibration model performance was poorer when based on spectra acquired from the proximal end compared with the equatorial and distal ends of the fruit (Table 2). This result is not surprising, given the variation in proximal end morphology (variable pedicel removal).

Given the above consideration of attribute distribution and spectral acquisition, it is recommended that optical and reference sampling should occur at any position around the equator of the fruit in order to best represent the entire fruit.

*Spectroscopy*

The short wave NIR spectra of attributes that contribute to TSS (predominantly sucrose, but also glucose, fructose, organic acids, pectins, etc.) and DM (e.g. soluble sugars, starch, cellulose, lignin, proteins, lipids, and, by negative correlation, water) relate to second and third overtones of OH and CH stretching, and related combination bands. These bands are characteristically broad and overlap, and thus spectra are 'featureless'. For example, the chemical environment of each OH and CH bond in water and sugar molecules is different, so that the effective absorption bands are wide. Derivative spectroscopy is used to tease out differences from such spectra, and multivariate calibration is used to tease out relationships between the spectra and the attribute of interest.

Spectral features relevant to sugar and water in the 700 to 950 nm spectral region include the second and third overtones of OH stretching vibrations at around 960 and 760 nm, respectively, and the first and second overtone of OH combination bands at around 840 and 1,180 nm, respectively (Golic *et al.* 2003). Spectral features relevant to sugar CH groups include second order stretching bands between 1,100 and 1,200 nm, and a third overtone band around 910 nm. Miyamoto and Kitano (1995) found the key wavelengths for a MLR calibration model on mandarin fruit TSS to be 770 and 905 nm for intact and peeled fruits respectively. These authors considered absorption at 770 and 905 nm to be associated with the fourth overtone of $CH_2$ and the third overtone of CH and $CH_2$, respectively, while absorption at 840 to 855 nm was related to fruit diameter (at a constant temperature).

The use of the 760, 884 and 910 nm wavelengths in the FMLR TSS models was therefore an attempt to use wavelengths related to 2 overtones of OH stretching and the third overtone of CH stretching. Using the WINISI stepwise MLR, models

generally defaulted to 4 to 9 wavelengths, including wavelengths close to 900, 870 and 860 nm for TSS and 907, 890, and 780 nm for DM.

McGlone *et al*. (2003) noted that "vast numbers of different spectral windows can be created over the wavelength range and all possible options could not be investigated in any reasonable time". These authors limited their investigation to 7 spectral windows chosen on the basis of "prior experience and intuition", concluding that a wavelength range of 750 to 1,100 nm was optimal for interactance spectra. In the current study all combinations of start and end wavelengths were considered for the development of a PLS model. Calibration model performance was relatively stable across the broad spectral window of 703 to 950 nm for both TSS and DM (Fig. 2).

Calibrations for TSS based on second derivative absorbance contained points of inflection for the calibration *b* coefficients at around 760, 860 and 905 nm, while for DM points of inflection were around 760, 810, 850 and 910 nm (Fig. 3). However, the plots are disconcertingly difficult to interpret in terms of spectroscopic relevance, in contrast to the experience of Golic *et al.* (2003) who worked in the same (short wave NIR) wavelength region but with model sugar-water solutions. The high weighting of features not directly related to the attribute of use should make the calibration less robust for an independent validation population. In practice however, the MPLS calibrations were more robust than the MLR or FMLR calibrations where specific wavelengths are selected. Presumably overlap between bands allows shoulder regions to hold more useful information than regions of the absorption peaks. We conclude that use of the whole wavelength region, 720 – 950 nm, is warranted for development of both TSS and DM models. This includes the spectral region associated with the second overtone CH stretch of sugar (910 nm).

Organic acids (titratable acidity) are present in intact fruit at relatively low levels (ca. 1.0%). As such, detection using NIRS is unlikely and we agree with the assessment of McGlone *et al.* (2003) that previous reports of calibration on this attribute are likely to represent secondary correlations on attributes related to fruit maturity.

Another quality defect for mandarins is apparent dryness, commonly assessed by % juiciness of fruit. This characteristic was not modelled successfully with NIR spectral data (correlation coefficient < 0.01, data not shown). This result is consistent with observation that the dryness defect does not correlate with water content (DM). Presumably, some of the water is present in the fruit in a bound (gelled) form. This result is in contrast to that of Peiris *et al.* (1998) but the 'section dryness' defect considered by Peiris may well have been a different type of defect (e.g. frost damage, in which juice sacs dehydrate following damage).

*Spectral collection*

The calibration results for the 0º (interactance) and 45º (transmittance) geometries (e.g. $R_c^2$ of 0.91, RMSECV of 0.4% TSS) were not significantly different. The 0º (interactance) geometry was expected to produce a poorer calibration model than the 45º (transmittance) geometry due to increased detection of specularly reflected radiation and/or the shorter path length of light through the fruit in the 0º geometry (as reported by McGlone *et al.* (2003)). In practice, the degradation in performance was marginal (not significant in terms of RMSECV), indicating that little specular light was detected and that a representative volume of the fruit was optically sampled using the 0º geometry.

The 45º geometry was applied with the detector probe in contact with the fruit to exclude specular reflection.  The 0º geometry was applied without physical contact between the detector probe and the fruit, in contrast to the application of McGlone *et al.* (2003).  The shadow cast by the detector probe in the 0º geometry minimises detection of specular light relative to reflectance spectroscopy.  The separation of probe and fruit allows for rapid in-line sorting, outweighing any disadvantage in terms of a marginally poorer calibration performance due to increased detection of specularly reflected radiation or a shorter path length of light through the fruit.  This conclusion is similar to that of Greensill and Walsh (2000).

Most literature reports employ averaging of multiple scans (e.g. Guthrie and Walsh (1997)).  Increasing the number of scans should improve signal to noise by the square root of the number of scans averaged.  In practice the increase in calibration model performance with 32 scans, compared with 1, 2 or 4 scans, was minimal (no significant differences among RMSECV values).  On a commercial pack-line, operating at a belt speed of 1 m.s$^{-1}$, there is sufficient time for only 1 scan.

We recommend the use of the 0º geometry with a single scan per spectrum as appropriate for use with mandarin fruit.

### *Calibration data treatment*

Models in which the coefficients more highly weight spectroscopically significant wavelengths (i.e. wavelengths related to the band assignments associated with the analyte of interest) should perform better in terms of validation on independent populations (i.e. there should be less risk of over-fitting the model).  Multiple linear regression models were developed in which the regression was based on 4 to 6 wavelengths anywhere in the 720 – 950 nm region, and in which the

regression was 'forced' to use data between 860 – 890 nm and 900 – 933 nm, wavelengths relevant to sucrose band assignments. The 'forced' MLR models were generally better in validation than MLR models, but not MPLS models (Table 7). This result was not related to outlier detection and removal routines as no outlier removal was undertaken in these validation exercises. Thus, while there is a greater potential to overfit MPLS than MLR models, this did not occur, as, in general, MPLS models were better than MLR models in both calibration development and validation on independent populations (Tables 6 and 7). We therefore recommend use of the MPLS procedure in preference to MLR.

Calibration model performance was relatively insensitive to the 'gap' size of derivation. This result is consistent with the wavelength resolution of the Zeiss MMS1 (peak width at half maximum for a line light source of 13 nm, Walsh *et al.* (2000)) and the relatively broad absorption bands for sugar and water occurring in the short wavelength near infra-red region.

For TSS, the optimal derivative and scatter correction condition differed among validation populations (Table 3) but, in general, a first derivative with SNV and detrend routines supported superior model performance. This is consistent with an optical geometry that involves some reflectance (thus sensitive to changes in the sample surface). For DM, no method was consistently superior to other methods. Hence, the use of first derivative (gap size of 4 data points), SNV and detrend procedures are considered appropriate mathematical treatments for calibration model development for mandarin fruit.

## CONCLUSIONS

In this exercise we have attempted to rationalise the NIR calibration procedure for determination of TSS and DM in intact mandarin. The recommended procedure involved sampling of an equatorial position on the fruit using either 0° interactance or 45° partial transmittance optics using 1 scan per spectrum, with partial least squares model development on a 720 to 950 nm window, pre-treated as first derivative absorbance data (gap size of 4 data points) with standard normal variance and detrend scatter correction. A lack of robustness is obvious, however, in terms of the ability of the models to predict attribute levels in new populations. In a companion manuscript we consider sources of variation between populations and calibration model updating procedures.

## ACKNOWLEDGEMENTS

# 4



# CALIBRATION MODEL ROBUSTNESS FOR MANDARIN FRUIT INTERNAL QUALITY ATTRIBUTES[3]

## ABSTRACT

The robustness of multivariate calibration models, based on near infra-red spectroscopy, for the assessment of total soluble solids (TSS) and dry matter (DM) of intact mandarin fruit (*Citrus reticulata* var. 'Imperial') was assessed. Total soluble solids calibration model performance was validated in terms of prediction of populations of fruit not in the original population (different harvest days from a single tree, different harvest localities, different harvest seasons). Of these, calibration performance was most affected by validation across seasons (signal to noise statistic on root mean square error of prediction of 3.8, compared with 20 and 13 for locality and harvest day, respectively). Procedures for sample selection from the validation population for addition to the calibration population ('model updating') were considered for both TSS and DM models. Random selection from the validation

group worked as well as more sophisticated selection procedures, with approximately 20 samples required. Models that were developed using samples at a range of temperatures were robust in validation for TSS and DM.

## INTRODUCTION

In a companion manuscript (Guthrie *et al.* 2005a) the reference sampling procedure and data pre-processing techniques were optimised for the development of partial least squares (PLS) calibration models on intact mandarin fruit for total soluble solids (TSS) and dry matter (DM), using short wavelength (720 – 950 nm) near infra-red (Yermiyahu *et al.* 1997) spectra acquired in an interactance mode (Greensill and Walsh 2000). Chemometric descriptive terms were also defined and will be used in the current manuscript.

The performance of a NIR calibration model in terms of predictive ability is impacted by a range of factors, including instrumentation changes, sample temperature and changes in sample (fruit) size and chemistry.

The application of near infra-red spectroscopy (NIRS) to a given fruit commodity requires an assessment of the robustness of the calibration model across populations of fruit grown under differing conditions. Different growing conditions may result in differences in physical (e.g. trichome density, intercellular space content) and chemical (e.g. water content) properties of a fruit, resulting in altered fruit optical properties and band assignments. Unfortunately, the majority of reports on the application of NIRS to fruit sorting detail the use of a single harvest population, divided into a calibration population and a validation population. For example, McGlone *et al.* (2003) acquired data on mandarins from 3 orchards at approximately weekly intervals over 7 weeks, resulting in 20 data populations for

subsequent analysis. A calibration population was assembled from 75% of each population and used to predict a population consisting of the remaining samples from each of the populations. This procedure allowed an estimate of prediction error, however it did not involve testing the calibration against fruit from independent populations (e.g. different harvest dates or localities). We are aware of only 3 relevant reports involving the use of separate harvest populations of fruit for calibration and validation. These studies involved mandarin and peach fruit.

Ou *et al*. (1997) reported the use of a calibration developed in 1 fruit growing region to predict TSS of Ponkan mandarin fruit from that region and from 2 other regions, with $R_v^2$ of 0.72, 0.44 and 0.30, and RMSEP of 0.68, 1.16 and 1.28% TSS, respectively. A calibration based on data combined across regions performed better, with $R_v^2$ of 0.76 and RMSEP of 0.92% TSS. Miyamoto and Kitano (1995) reported the use of a calibration developed in 1 year to predict TSS content of intact Satsuma mandarins in the subsequent 2 seasons. Prediction statistics were similar to that for calibrations developed within a given season (RMSEP of < 0.6% TSS and *bias* of ≤ 0.4% TSS).

Miyamoto and Kitano (1995) also reported calibration validation across 3 seasons for peach. Prediction statistics for a calibration developed across data from all years (RMSEP of 0.60% TSS, *bias* of < 0.1% TSS) were better than for a calibration developed in any 1 year (RMSEP of 0.64% TSS, *bias* up to 0.34% TSS). Also using peaches, Peiris *et al*. (1998b) reported that a calibration developed on a population drawn from 3 seasons predicted better on a combined season validation population (RMSEP of 0.9 – 1.3% TSS and *bias* of 0.2 – 0.4% TSS) than that developed from populations drawn from a single season (RMSEP of 0.90 – 1.4% TSS and *bias* of 0.2 – 2.1% TSS).

The comparisons of model validation between independent populations are usually difficult as population attributes (e.g. SD) vary. The standard deviation ratio (SDR), expressed as the ratio of SD to RMSECV (for calibration data populations) or RMSEP (for validation data populations), or RPD (ratio of the SD to RMSECV(C) or RMSEP(C) of the data) (Williams and Sobering 1993), is sometimes presented as a gauge of the utility of the technique. Other indices have also been used. Ou *et al*. (1997) reported a form of the coefficient of variation (CV) statistic (CV = RMSEP/mean of the prediction population), while Miyamoto and Kitano (1995) reported an evaluation index (EI = 2 * RMSEP/range * 100), in an attempt to compare model performance across populations. Another approach, suggested by Wortel *et al*. (2001), is based on the Taguchi concept of process control, in which the variation of RMSEP among validation populations of a given condition (e.g. populations drawn from different harvest regions) is quantified in a signal to noise (S/N) statistic (S/N = 20 * $\log_{10}$ [mean RMSEP/$SD_{RMSEP}$], where mean RMSEP is the average of the RMSEP across all validation populations, and $SD_{RMSEP}$ is the SD of all the RMSEP values).

In other industries (e.g. cereal, oilseed) NIRS based models are extended by inclusion of samples from the validation population (e.g. from a new variety of wheat or a new season of oilseed production). The decision on when to add new samples to the calibration population is generally based on an assessment of the dissimilarity of the calibration and validation populations based on principal component analysis (PCA) or partial least squares analysis (PLS) scores. The Mahalanobis distance *D*, (Mahalanobis 1936) is such a measure. The chemometric software package WINISI (ver. 1.04a) uses mean centred score data in the calculation of *D*. Further, *D* is normalised to *f* to create the global H (GH) statistic, as follows:

$$GH = \frac{D^2}{f}$$

where $f$ is the number of PCA/PLS factors in the model.

Shenk and Westerhaus (1991) advocate the use of the GH value and a 'nearest neighbour' Mahalanobis distance (NH, Mahalanobis distance from any given sample to its nearest neighbour in principle component space) for the selection of outliers and for sample addition. However, the choice of how many and which samples from the validation population should be added to the calibration population is vexatious. Typically high leverage samples, which are not outliers, will be chosen, with the number required defined through trial and error (e.g. Wang *et al.* 1991).

Calibration model performance is affected by sample temperature primarily through the strong effect of temperature on H bonding and thus on the absorption bands related to OH (Golic *et al.* 2003). Model robustness should therefore be considered with respect to this variable. We hypothesize that calibration models for DM would be more sensitive to temperature than models for TSS, as DM models may reflect water content.

Kawano *et al.* (1995) noted that as sample fruit (peach) temperature increased, so did absorption at 841 and 966 nm (water bands), resulting in a *bias* in the prediction of TSS. Miyamoto and Kitano (1995) noted that when a calibration developed from spectra collected from mandarin fruit at 20°C was used to predict the same fruit at 6, 15 and 25°C, RMSEP (presumably RMSEP(C)) was constant but *bias* increased linearly with temperature. These researchers developed MLR models using 3 wavelength regions, noting 900 - 910 nm to be directly relevant to sugar, 740 – 755 nm or 840 - 855 nm to compensate for the optical path-length of the fruit, and 794 or 835 nm to compensate for the influence of fruit temperature. Both reports

concluded that if the calibration model was developed with sample temperatures covering the range of future sample temperatures, then prediction accuracy was high. Sanchez *et al.* (2003) also noted that the influence of spectrometer and fruit (apple) temperature was mainly on *bias*, not RMSEP(C). However the effect of spectrometer temperature on *bias* was more than twice that of fruit temperature.

The 'repeatability' file option of WINISI (ver. 1.04a) software may represent an alternative procedure for developing robustness in the model with respect to sample temperature. This procedure was developed for the calibration transfer between instruments, and depends on the collection of spectra of a few samples on different instruments. However, rather than add spectra directly to the calibration, the 'repeatability' file adds 'difference' spectra (for each sample, scanned under different temperatures), with corresponding reference values of zero (Shenk and Westerhaus 1991). The calibration algorithm is modified to give extra weight to these spectra.

In the current study we report on the robustness of NIRS models for the evaluation of attributes related to eating quality (% TSS, % DM) of intact mandarin, and on procedures to select samples for addition to the calibration population. Calibration robustness across harvest time, location and seasons for prediction of TSS, employing the assessment methodology suggested by Wortel *et al.* (2001), is considered. Calibration robustness for prediction of TSS and DM is also considered with respect to sample temperature. Calibration performance across instruments (e.g. as reported for Imperial mandarins by Greensill and Walsh (2002)) will not be considered here.

## MATERIALS AND METHODS

### *Plant material, reference analyses and spectroscopy*

Mandarin fruit (Imperial variety) were sourced following commercial harvest from orchards in Mundubbera (25.6º S, 151.6º E), Bundaberg (24.9º S, 152.3º E) and Dululu (23.8º S, 150.3º E), Queensland.

Populations used in this study are the same as those used in the companion study for the 2001 season, with populations alphabetically named in chronological order as described in Guthrie *et al*. (2005a). Additional populations from the 1999 and 2000 seasons were used in Fig. 1.
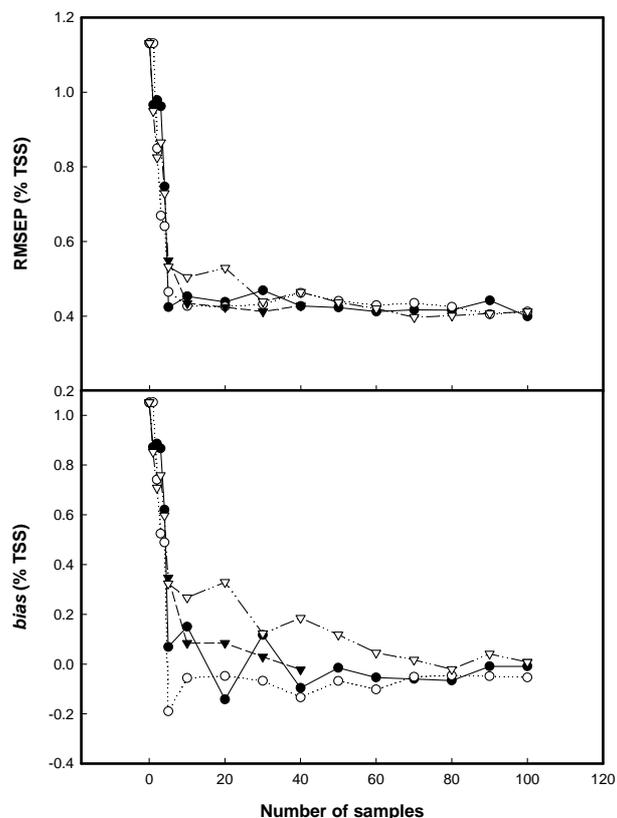
**Figure 1.** Prediction statistics (root mean square error of prediction (RMSEP) and *bias*) for a modified partial least squares (MPLS) calibration model for mandarin fruit total soluble solids (TSS), using 4 methods for sample selection from the prediction population for addition to the calibration population (combined population from years 1999 and 2000 (n = 307, mean of 9.8 and SD of 1.42% TSS). The independent validation population from year 2000 (n = mean of 14.2 and SD of 1.05% TSS) was divided into 2 equal parts, with 1 part used for validation and the other part used for sample addition. The 4 methods were (a) every 'i[th]' sample based on ranking by Mahalanobis Global H statistic (GH), closed triangle; (b) samples with greatest GH less than 3, open circle; (c) samples chosen randomly, closed circle; and (d) samples with the greatest Mahalanobis Neighbourhood H statistic (NH), open triangle.

Model robustness across harvest day, location and season was evaluated for TSS using populations gathered from a single tree over 2 weeks, from 3 different locations and from 4 seasons (from different locations) (Table1). For TSS, the calibration population was a combined J and K with validation populations of M, E, G and L. For DM, the calibration population was T with validation populations of V, R and S. Additionally, a model developed on a combined population made up of 2 populations per year from years 1999 and 2000 ($n = 307$, mean 9.9 and SD of 1.44% TSS) was used to predict on a separate population from year 2000 (mean 14.2 and SD 1.05% TSS).

Total soluble solids content of extracted juice and DM of fruit halves was determined as described in Guthrie *et al.* (2005a) and Chapter 3. The procedures used to acquire spectra, were also described in Guthrie *et al*. (2005a). Briefly, spectra were collected over the wavelength range 720 – 950 nm using a NIR enhanced Zeiss MMS1 spectrometer and a 100 Watt tungsten halogen light in the interactance optical configuration reported by Greensill and Walsh (2000) ($0^{o}$ angle between illumination and detected light rays, with detection probe viewing a shadow cast by the probe onto the fruit).

**Table 1.** Calibration (Cal) and validation (Val) statistics for modified partial least squares (MPLS) and multiple linear regression (MLR) calibration models on mandarin total soluble solids (TSS), with validation across several populations varying in (a) time (Crisosto *et al.* 2000) of harvest, (b) harvest location, and (c) season of harvest. Wavelength range was from 720 to 950 nm. Variation in prediction performance is reported in terms of the Taguchi signal to noise (S/N) value, root mean square error of prediction (RMSEP) and average RMSEP. Population identifiers (letters in brackets) refer to Table 6 in the companion manuscript (Guthrie *et al*. 2005a).

| Fruit Population | SD (% TSS) | MPLS | | | MLR | | |
|---|---|---|---|---|---|---|---|
| | | $R_c{}^2$ | RMSECV/ RMSEP (% TSS) | bias (% TSS) | $R_c{}^2$ | RMSEP(C) (% TSS) | bias (% TSS) |
| *Time* | | | | | | | |
| **Cal** (J - K) (Day1 & 3) | 0.92 | 0.9 | 0.332 | | 0.86 | 0.36 | |
| **Val** | | | | | | | |
| Day 5 (L) | 1.04 | 0.86 | 0.406 | 0.06 | 0.79 | 0.54 | -0.07 |
| Day 7 (M) | 0.74 | 0.56 | 0.533 | -0.13 | 0.64 | 0.48 | 0.03 |
| Day 9 (N) | 0.68 | 0.56 | 0.478 | 0.11 | 0.49 | 0.56 | 0.02 |
| Day 10 (P) | 0.84 | 0.84 | 0.512 | 0.39 | 0.8 | 0.95 | 0.85 |
| Day 13 (Q) | 0.67 | 0.68 | 0.509 | 0.33 | 0.63 | 0.87 | 0.75 |
| S/N RMSEP | | | **19.8** | | | **10.1** | |
| Av RMSEP | | | **0.49** | | | **0.68** | |
| *Location* | | | | | | | |
| **Cal** (J-K) | 0.92 | 0.9 | 0.332 | | 0.86 | 0.36 | |
| **Val** | | | | | | | |
| A (E) | 0.99 | 0.75 | 0.589 | 0.31 | 0.6 | 0.79 | 0.39 |
| B (F) | 0.49 | 0.35 | 0.806 | 0.66 | 0.14 | 1.24 | 1.09 |
| C (G) | 0.6 | 0.53 | 0.951 | 0.85 | 0.3 | 1.31 | 1.15 |
| S/N RMSEP | | | **12.7** | | | **12** | |
| Av RMSEP | | | **0.78** | | | **1.12** | |
| *Seasons* | | | | | | | |
| **Cal** (1999) | 0.92 | 0.93 | 0.273 | | 0.87 | 0.34 | |
| **Val** | | | | | | | |
| Year 1 (1999) | 1.05 | 0.35 | 4.944 | 3.21 | 0.32 | 3.7 | 1.51 |
| Year 2 (2000) | 1.05 | 0.83 | 2.099 | 2.05 | 0.78 | 2.13 | 2.07 |
| Year 3 (2001) | 0.78 | 0.03 | 6.756 | -3.58 | 0.02 | 4.74 | -0.35 |
| Year 4 (2004) | 1.32 | 0.74 | 0.77 | -0.31 | 0.21 | 2.74 | 1.95 |
| S/N RMSEP | | | **3.82** | | | **1.5** | |
| Av RMSEP | | | **3.64** | | | **3.33** | |

*Chemometrics*

The software package WINISI (ver. 1.04a) was used for chemometric analysis. Calibrations were developed using both step-wise multiple linear regression (MLR) and modified partial least squares regression (MPLS). The data pre-treatment options of first derivative, standard normal variance and detrend scatter correction, as recommended in the companion study (Guthrie *et al*. 2005a), were adopted throughout the current study. The 'repeatability' file option in WINISI was also considered as a method to improve prediction statistics across the different sample temperatures.

The criteria of Wortel *et al*. (2001) were applied to evaluate model robustness. This approach involved calculation of an average RMSEP and the S/N statistic for the performance of a given model across a range of validation populations.

A common approach for the improvement of calibration performance on a new validation population involves the addition of samples from the validation population to the calibration population. In this study, we extend the treatments reported by Guthrie and Walsh (2001). Each validation population was initially assessed for outliers as samples with a GH > 3.0 using its own scores and loadings. These outliers were removed and the resulting data divided randomly into 2 groups - 1 group (2 thirds) retained as the validation population and the other group used for selection of samples for addition to the calibration population.

Three approaches were used in the selection of samples from the validation population for addition to the calibration population;

- random - done twice using 2 different seeds to the random number generator;

- selected on GH value, selecting samples with either (a) minimal GH (i.e. spectrally similar to the 'mean' of the calibration population), (b) maximal GH (i.e. spectrally dissimilar to the 'mean' of the calibration population), or (c) spaced evenly on GH ranking (i.e. representative of the 'spread' of the calibration population);

- selection on the basis of NH using 2 methods - (a) NH cut-off (in which only samples with a NH value greater than the 'cut-off' value are chosen; this procedure is available as a WINISI software option, under 'Make and Use Scores', 'Select Samples From a Spectra File'), and (b) NH end (a manual implementation of (a), in which all samples were ranked manually in ascending order of magnitude for NH, with high NH value samples chosen);

Thus in total 6 methods for sample selection were trialled. In these exercises, the GH and NH values were calculated for validation population members, using the scores and loadings of the calibration population. All validation populations were independent of the calibration populations.

Different population updating techniques were compared, as were different numbers of samples for model updating. This was trialled on different calibration and validation populations for the attributes of both TSS and DM (Figs. 1 to 5).
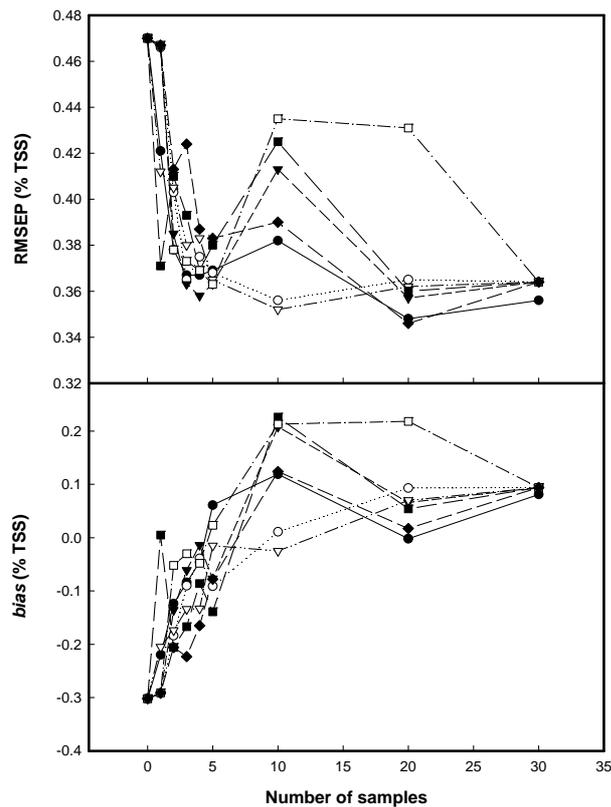
**Figure 2.** Prediction statistics (root mean square error of prediction (RMSEP) and *bias*) for a modified partial least squares (MPLS) calibration model (Populations J and K) for mandarin fruit total soluble solids (TSS), using 6 methods for sample selection from the prediction population for addition to the calibration population. The independent validation population (Population M) was divided into 2 parts, with 2 thirds used for validation and the remainder used for sample addition. The validation population (sub-set of population M) was totally independent of the calibration population (population J and K). The 6 methods were (a) samples with minimum GH values, closed circle, (b) every 'i[th]' sample based on ranking by Mahalanobis Global H statistic (GH), open circle; (c) samples with greatest GH less than 3, closed triangle, (d) samples chosen randomly, open triangle and closed square; (e) samples selected using WINISI sample addition facility, open square; and (f) samples with the greatest NH, closed diamond.
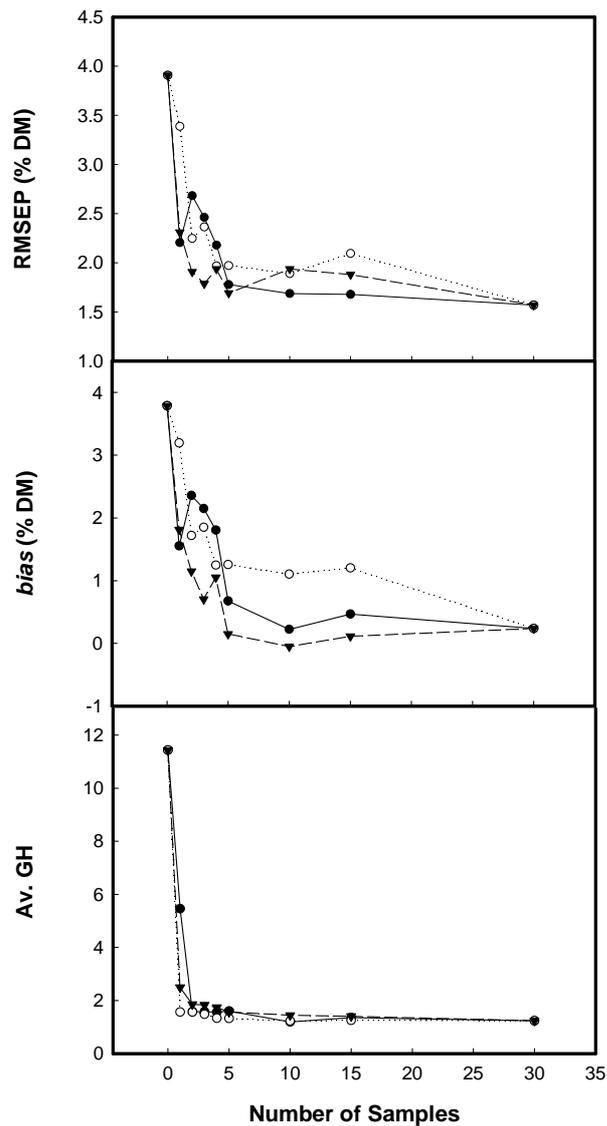
**Figure 3.** Prediction statistics (root mean square error of prediction (RMSEP) and *bias*) for modified partial least squares (MPLS) prediction models for dry matter (DM) of mandarin fruit, using 3 methods for sample selection from the prediction population for addition to the calibration population. The independent validation population was divided into 2 parts, 2 thirds used for validation and the remainder used for sample addition. The validation population (sub-set of population V) was totally independent of the calibration set (population T). The 3 methods were (a) samples with every 'i[th]' sample based on ranking by Mahalanobis Global H statistic (GH), closed circle, (b) samples with greatest GH less than 3, open circle, (c) samples chosen randomly, closed triangle. The average GH of samples in the validation population was calculated using calibration model scores.
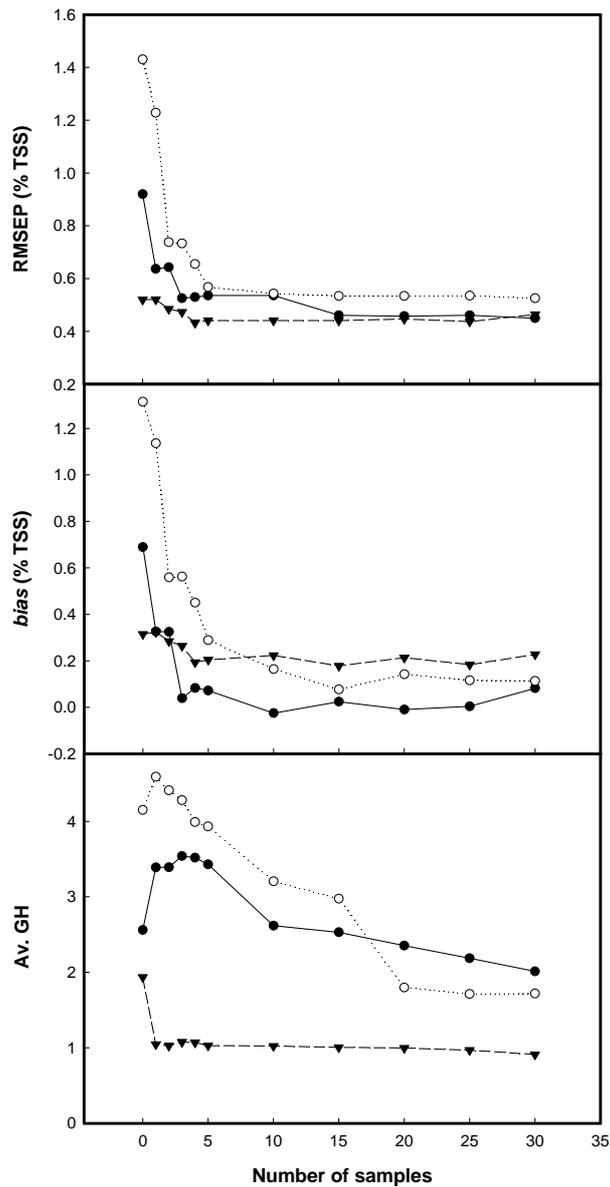
**Figure 4.** Prediction statistics (root mean square error of prediction (RMSEP) and *bias*) for modified partial least square (MPLS) prediction models for total soluble solids (TSS) of mandarin fruit, using 3 independent (of calibration populations) validation populations (Populations E, closed circle; G, open circle; and L, closed triangle). The average Mahalanobis Global H statistic (GH) of samples in the validation population was calculated using calibration model scores. The initial calibration population consisted of populations J and K (as for Fig. 2). Samples were selected randomly from the prediction populations for addition to the calibration population.
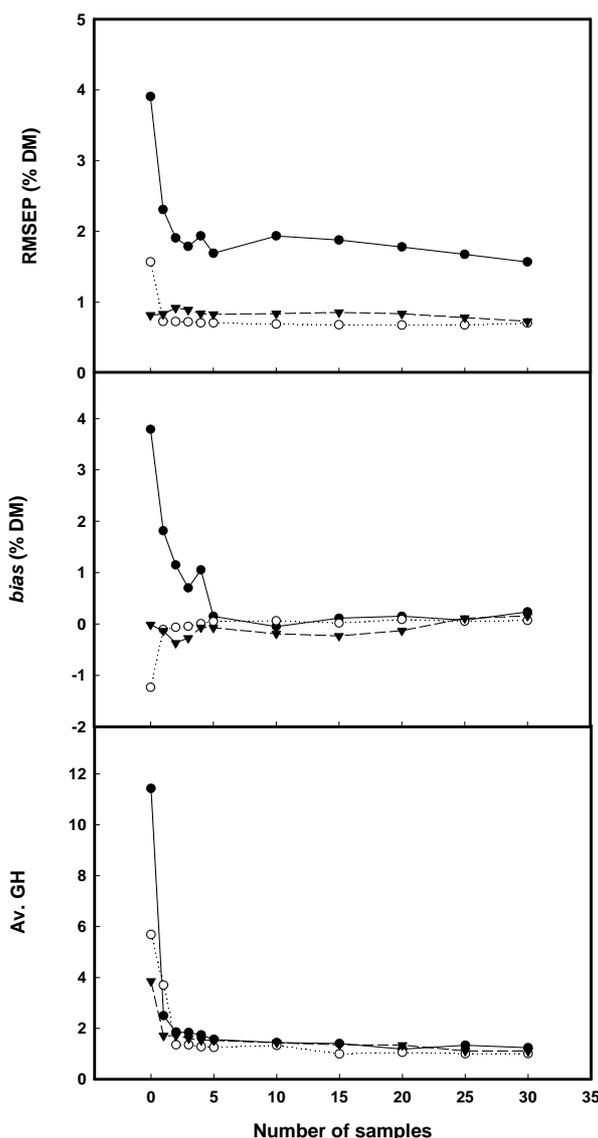
**Figure 5.** Prediction statistics (root mean square error of prediction (RMSEP) and *bias*) for modified partial least squares (MPLS) prediction models for dry matter (DM) of mandarin fruit, using 3 independent (of calibration populations) validation populations (V, closed circle; R, open circle; and S, closed triangle). The average Mahalanobis Global H statistic (GH) of samples in the validation population was calculated using calibration model scores. Samples were selected randomly from the prediction populations for addition to the calibration population. The initial calibration population was population T.

For 1 population of mandarin fruit (population T), spectra were collected of fruit at room temperature (22°C) and then the fruit equilibrated to 10°C and 30°C and rescanned. These fruit were assessed (separate halves) for both TSS and DM. Calibration models were developed on a population of 70 samples (mean 9.6% and

SD 1.51% for TSS, and mean 14.7% and SD 1.66% for DM), from spectra collected

of these fruit at 10, 22 and 30°C.  The prediction populations were based on a separate

population of 34 samples (mean 9.8%, SD 1.64% for TSS and mean 14.7%, SD

2.03% for DM), again with spectra collected of these fruit at 10, 22 and 30°C.  The

calibration models for TSS involved 5 terms, while those for DM involved 6 or 7

terms.  Calibration models were developed on spectra of fruit at 10, 22, 30, 10 and

22°C, and all 3 temperatures.  The WINISI 'repeatability file' option was also

employed, using all samples or the 4 samples with lowest GH values from the 22°C

validation population.  The significance (P < 0.05) of differences in RMSEP and *bias*

was tested as described by Fearn (1996), using an automated spreadsheet (Guthrie *et*

*al*. 2005) (see Appendix F).


# RESULTS

### *Calibration model robustness*

A given TSS calibration model was used to predict populations over harvest

day, location and season (Table 1).  The model used to predict populations over

harvest day and location was based on the combination of 2 populations (J and K),

while the model for predicting populations over seasons (years) was based on

populations from 1999.  Model predictions were more variable across seasons than

across harvest days or location (in terms of both RMSEP and *bias)*.  This prediction

variability was indexed as an average RMSEP and a S/N on RMSEP following the

procedure of Wortel *et al.* (2001).  The S/N ratio on the RMSEP of the MPLS model

predictions was 20 over harvest days, 13 over location, and 4 over seasons (Table 1).

Modified partial least squares models were more robust than MLR models (MLR

models had lower S/N ratios, being 10, 12 and 2 for harvest days, locations and seasons, respectively).

Model performance in prediction of TSS of an independent population was improved by inclusion of samples from the independent population, regardless of the method used to select the samples for inclusion (Figs. 1 and 2). This was demonstrated for a calibration developed on populations from 1999 and 2000 (Fig. 1), and from 2001 (Fig. 2). Of the 4 methods of sample selection employed (random, every 'i[th]' sample based on ranking by GH, maximum GH, and maximum NH), all behaved similarly (Fig. 1). Model performance improved from 1.1% TSS to 0.45% TSS for RMSEP and from 1.1% TSS to 0.15% TSS for *bias* with the inclusion of only 10 samples (Fig. 1). In the second exercise, where 6 methods of selection were used (as above plus minimum GH and WINISI sample addition facility), all methods again behaved similarly in terms of *bias* (Fig. 2). In terms of RMSEP, all methods behaved similarly with addition of up to 5 samples, but there was some divergence between methods with addition of 10 to 20 samples. The RMSEP values increased with the addition of 10 samples for the 'random' and 'greatest GH' (approaches 1 and 2, respectively) and for the addition of 10 and 20 samples for the WINISI 'sample addition facility' (approach 3). However, a repeat of the 'random' selection approach gave divergent results for the addition of 10 samples. Of course, with the addition of all samples from the 1 third validation population all results will converge (except for where there is a slight difference in the size of the population, (e.g. 30 drawn from a population of 31 or 34).

Model performance (from calibration on Population T) in prediction of DM of an independent population (Population V) was also improved (in terms of both RMSEP and *bias*) by inclusion of samples from the independent population,

regardless of the method used to select the samples for inclusion (Fig. 3). Of the 3 methods of sample selection employed (random, every 'i[th]' sample based on ranking by GH and maximum GH), all behaved similarly, reaching a stable value after the addition of 10 samples.

The effect of sample addition (using the random selection method) on the performance of a TSS model (as used for Fig. 2; based on Populations J and K) was described for a further 3 independent validation populations (Populations E, G and L). A similar activity was undertaken for DM (calibration Population T and validation Populations V, R and S). The GH of the validation population was calculated using scores and loadings of the calibration population, with recalculation after each sample addition. Where the average GH of the validation population was markedly different from the calibration population (e.g. average GH > 3), the improvement in validation was quite dramatic (e.g. RMSEP decreasing from 1.45 to < 0.60% TSS with the addition of only 5 samples). When the average GH of the validation population was similar to the calibration population (i.e. GH < 3), the validation performance, while initially acceptable, showed little improvement (e.g. RMSEP changed from 0.50 to 0.42% TSS with addition of 5 samples for a population with an initial average GH of 2) (Figs. 3 and 4).

### *Sample temperature*

Model statistics (RMSEP) for TSS prediction were not significantly different for calibration models developed using spectra of fruit at either 10 or 22°C, but that for 30°C was inferior to that at 22°C (Table 2). For calibrations developed on DM for these 3 fruit temperatures, calibration model RMSEP was not significantly different for models developed at either 10 or 22°C, but a significantly lower RMSEP was

achieved at 30°C, compared to that at 22°C (Table 2). Relative to models developed using fruit at several temperatures, a model (TSS or DM) developed at a single temperature (22°C) produced an inferior result (in terms of *bias* rather than RMSEP(C)) when fruit temperatures were other than that of the calibration population. For both the attribute of TSS and DM, *bias* was related to fruit temperature ($R^2$ = 0.96, slope = –0.10% TSS/°C ($R^2$ = 0.86, slope = -0.10% DM/°C).

Incorporating samples with different temperatures in the calibration population improved the prediction performance of both the TSS and DM models, in prediction of samples within the temperature range included in the calibration population (Table 2). For example, *bias* was -1.17, -0.15, and 0.07 for TSS models developed at sample temperatures of 22°C only, 10 and 22°C, and 10, 22 and 30°C, respectively, for prediction of a population of samples at 30°C.

The 4 samples with the lowest GH value from the calibration population scanned at 22°C were identified. The spectra of these samples at all 3 temperatures were included in the 'repeatability' file of WINISI. In a second exercise, all samples (from across all temperatures) were used in the 'repeatability' file. Including spectra of fruit scanned at different temperatures in the 'repeatability file' did not improve calibration model statistics (for either DM or TSS), or model prediction statistics for DM, relative to a model using fruit re-scanned at all temperatures (Table 2). In contrast, the 'repeatability' file option supported better prediction statistics for TSS, in terms of both RMSEP and *bias*, relative to a calibration developed using fruit re-scanned at all temperatures. Using all samples in the 'repeatability' file was, however, better than using only 4 from each scanning temperature in this WINISI option.

**Table 2.**  **Effect of temperature on prediction of dry matter (DM) and total soluble solids (TSS) for mandarin fruit.  Models were developed on a population of 70 samples, mean 9.6% and standard deviation (SD) 1.51% for TSS, and mean 14.7% and SD 1.66% for DM, from spectra collected of these fruit at 10, 22 and 30°C.  The prediction populations were based on a population of 34 samples, mean 9.8%, SD 1.64% for TSS and mean 14.7%, SD 2.03% for DM, again with spectra collected of these fruit at 10, 22 and 30°C.  The calibration models for TSS involved 5 terms, while those for DM involved 6 or 7 terms.  Calibration models were developed on spectra of fruit at 10, 22, 30, 10 and 22°C, and all 3 temperatures.  The WINISI 'repeatability file' option was also employed, using 4 samples and all samples.  The 4 samples with lowest Mahalanobis Global H statistic (GH) values from the 22°C population were used.**

| Sample Temperature (°C) | | *Calibration model statistics* | | | *Prediction model statistics* | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | 10°C | | 22°C | | 30°C | |
| | n | RMSEP | $R_c^2$ | RMSECV | bias | RMSEP(C) | bias | RMSEP(C) | bias | RMSEP(C) |
| **DM** | | | | | | | | | | |
| 10 | 70 | 0.63 a,b | 0.85 | 0.71 | | | | | | |
| 22 | 70 | 0.41 a | 0.94 | 0.51 | 0.54 | 0.79 | 0.04 | 0.52 | -1.48 | 0.79 |
| 30 | 70 | 0.61 b | 0.86 | 0.66 | | | | | | |
| 10+22 | 140 | 0.53 | 0.9 | 0.6 | -0.26 | 0.77 | -0.01 | 0.65 | -0.78 | 0.81 |
| 10+22+30 | 210 | 0.55 | 0.89 | 0.6 | -0.2 | 0.67 | -0.01 | 0.56 | -0.15 | 0.68 |
| 22+repeatability | 70 | 0.57 | 0.88 | 0.6 | -0.11 | 0.84 | -0.08 | 0.8 | -0.14 | 0.86 |
| **TSS** | | | | | | | | | | |
| 10 | 70 | 0.69 a | 0.79 | 0.75 | | | | | | |
| 22 | 70 | 0.73 a | 0.77 | 0.88 | 0.89 | 1.16 | 0.22 | 1.13 | -1.17 | 1.18 |
| 30 | 70 | 0.63 b | 0.82 | 0.76 | | | | | | |
| 10+22 | 140 | 0.69 | 0.79 | 0.62 | -0.08 | 1.11 | 0.17 | 1.11 | -0.15 | 1.1 |
| 10+22+30 | 210 | 0.68 | 0.79 | 0.62 | -0.05 | 1.14 | 0.38 | 1.19 | 0.07 | 1.17 |
| 22+repeatability | 70 | 0.69 | 0.79 | 0.81 | -0.07 | 0.9 | 0.07 | 0.95 | -0.07 | 0.89 |
| 22+repeatability (4 samples) | 70 | 0.79 | 0.73 | 0.89 | -0.15 | 1.09 | 0.16 | 1.13 | 0.03 | 1.12 |

# DISCUSSION

## *Calibration robustness – across seasons, locations and harvest time*

Validation of a model on a population independent of that used in calibration effectively tests for over-fitting of the model.  Where the calibration model has weighted spectral features that represent fruit characteristics that are correlated to the attribute of interest in the calibration population, but not in the validation population, then validation performance will be poor.  An example is a calibration developed for a variety in which skin chlorophyll content (skin greenness) is related to fruit TSS at maturity, which will not predict well with a variety in which there is no such

relationship (unless the wavelength range considered is trimmed to eliminate the spurious correlation).

Calibration performance across harvest days (fruit from 1 tree in the 1 season) was superior to that across locations (fruit from harvests from varying farms in 1 season) (e.g. S/N statistic of 20 and 13, respectively, with an average RMSEP of 0.49 and 0.78% TSS), but performance was dramatically degraded when applied across seasons (S/N of 4, average RMSEP of 3.64% TSS). There was no trend for performance to degrade with increasing time (days) or distance/soil type of harvest (data not shown).

Taguchi descriptors calculated from 3 literature reports differ to those reported here. A S/N statistic of between 15 and 19, with an average RMSEP of approximately 1.1% TSS was calculated from the results of Peiris *et al*. (1998b) for the use of peach TSS calibrations across 3 seasons. The S/N statistic and average RMSEP for the use of a single variety calibration model across other varieties was between 12 and 17, with an average RMSEP of approximately 1.0% TSS. The mandarin TSS predictions of Miyamoto and Kitano (1995) and Ou *et al*. (1997) yield a S/N statistic of 20 and average RMSEP of 0.58 for predictions applied across seasons, and 7 (S/N) and 0.81 (average RMSEP) across locations. Thus previous studies indicate that model performance should be more stable across seasons, for a given variety, than across varieties, in a given season.

The cause of the dramatic decrease in performance of a calibration when applied to fruit across seasons in this study is not clear and could reflect changes in the instrument used as well as changes in the sample (fruit). However, there were no obvious changes in lamp or detector characteristics (i.e. in white reference spectra collected across years). The change in calibration performance between seasons is

therefore more likely to represent changes in fruit optics (e.g. cell size, porosity), with consequent changes in the volume of fruit optically sampled, or in fruit composition (with characters other than the character of interest varying, and absorbing in similar wavelength regions).

### *Sample addition for calibration*

To improve calibration performance on a new validation population, a common strategy is the addition of samples from the new population to the calibration population. The RMSEP and *bias* (Figs. 1 to 5) decreased with addition of validation samples to the calibration population reaching a stable value with the addition of about 20 samples. Several approaches were used in the selection of samples from the validation population for addition to the calibration population, however (surprisingly) all methods performed equally well. This result indicates that the variation within a new population must be small, relative to the difference of that population to the calibration population, such that any sample chosen from within a given population is a useful representative of that population.

The higher the average GH of the validation population when calculated on the calibration population scores (Figs. 4 and 5 for the attributes of TSS and DM, respectively), the greater the improvement to RMSEP and *bias* with the addition of validation population samples. Higher average GH values reflect an increased difference in the spectra of calibration and validation populations, and a greater leverage on the MPLS regression will be gained in sample addition from the validation population.

It is surprising, however, that the reverse was not true, i.e. that the addition of high GH validation samples to the calibration population was not more effective than

116

the addition of low GH validation samples, in terms of improvement to prediction RMSEP and *bias*.

In practice, the level of accuracy required must be established for each sorting task. Higher accuracy requirements will require higher calibration maintenance. This maintenance might involve adjustment of *bias* for new populations, or recalibration with addition of spectra of at least 20 fruit from the new population to the existing calibration population, to recover RMSEP values.

It is not obvious why the inclusion of such a small number of samples to the calibration population can have such an influential effect. It may be partly due to the added samples being so different (separate) from the original calibration population, resulting in 2 'clusters' (original and new), which are basically treated by the calibration as 2 'points' (Fig. 6). However, it is then intriguing that the model predicts so well on the validation population. Regardless, the methodology was observed to work well in a number of circumstances and for a number of populations.
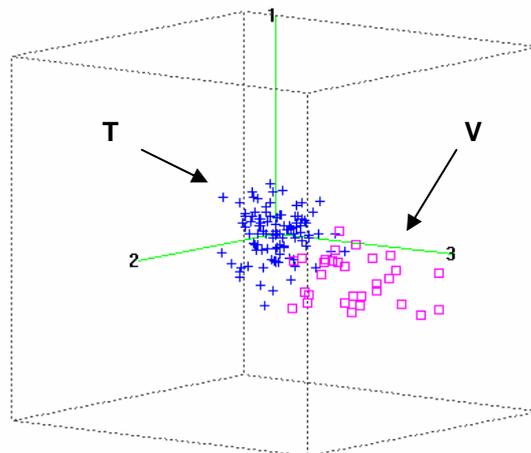


**Figure 6.** **Three-dimensional plot of MPLS scores (1, 2 and 3) of the calibration population T (*n* = 103) and an independent set (sub-set of population V, *n* = 30) calculated using a calibration model for dry matter content.**

*Calibration robustness - temperature*

Mandarin fruit temperature can vary from 5ºC (recommended storage temperature) to over 30ºC (field temperature) during in-line grading in a commercial packing shed.  Temperature impacts on the degree of H bonding, and thus the position and intensity of OH stretching vibration bands.  There are 2 main forms of liquid water, 1 form involving a H bond to another water molecule, and the other form involving more structured water.  The second form dominates at lower temperatures, and absorbs at higher wavelengths relative to the first form.  Golic *et al*. (2003) reported that calibration model statistics for models developed for pure sucrose solutions across a range of sample temperatures were degraded relative to those at a constant temperature (20ºC).  These calibrations resulted in a de-emphasis on those areas of the spectrum associated with OH stretching, favouring those areas associated with other spectral bands of the sugars (e.g. 910 nm CH third overtone).

Where a model was required to predict samples with temperatures outside the range included in the calibration population, *bias* was increased for both DM and TSS models (Table 2).  The RMSEP was affected primarily through an effect on prediction *bias*.  Therefore the following discussion reports on *bias* and RMSEP(C).  In practical application, a *bias* adjustment could be applied for the use of a calibration at temperatures outside of the range included in the calibration population.

Calibrations developed across a wide range of temperatures are expected to be more robust in terms of predicting analyte levels of samples at a range of temperatures, although potentially at the expense of diminished accuracy.  Prediction robustness in terms of *bias* was indeed increased for models developed across several temperatures, but accuracy (RMSEP(C)) was similar to that of single temperature calibration models, for both DM and TSS (Table 2).  Kawano *et al*. (1995) also found

that incorporation of samples across a temperature range in a (MLR) calibration allowed prediction of TSS with a high degree of accuracy and minimal *bias*.

We expected DM calibration models to be more sensitive to temperature than TSS models, given the sensitivity of the water bands to temperature (H bonding status) (Golic *et al.* 2003). This was not so, with TSS and DM similarly sensitive to temperature (Table 2). Presumably this effect reflects the large contribution of sugar OH features which are sensitive to H bonding status, and thus to temperature, in both the TSS and DM calibration models.

The 'repeatability' file option in WINISI was implemented in an attempt to reduce the sensitivity of the calibration to both sample and instrument temperature variations. Wavelengths with less change due to temperature should receive higher PLS scores, thus decreasing emphasis on the remaining wavelengths. For DM, implementation, to the extreme of including all spectra in the 'repeatability' file, was not as successful as incorporating all spectra from the same fruit scanned at a range of temperatures into the calibration population. However, the reverse occurred for TSS.

## CONCLUSIONS

Calibration models were less robust across seasons than across locations and time within a harvest season. In all cases, model updating involving the addition of relatively few samples (approximately 20) was successful in improving prediction of new populations. The method of sample addition was not crucial. Therefore, for ease of operation the random selection approach is the logical choice for sample addition to improve RMSEP and *bias* in the prediction of independent validation populations (for both the attributes of TSS and DM). The higher the average GH of the independent

population with respect to the scores and loadings of the calibration population, the greater the beneficial effect of sample addition.

We conclude, in agreement with Miyamoto and Kitano (1995) and Kawano *et al.* (1995) that samples scanned at a range of temperatures should be included in the calibration population in order for the model to be robust in prediction of samples varying in sample temperature. The issue of calibration population design to incorporate robustness for sample temperature without loss of general validation accuracy (i.e. what proportion of calibration samples should be run at different temperatures, and over what number of temperature steps) requires further consideration. Alternatively, the orthogonal projection method suggested recently by Roger *et al.* (2003) in a consideration of model robustness across instruments may have merit for increasing calibration robustness to temperature variation.

## ACKNOWLEDGEMENTS

# 5
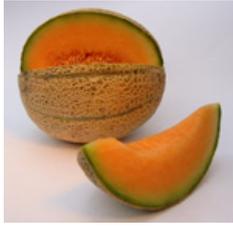


# MODEL DEVELOPMENT AND ROBUSTNESS IN PREDICTION OF MELON FRUIT TOTAL SOLUBLE SOLIDS[4]

## ABSTRACT

The robustness of multivariate calibration models, based on near infra-red spectroscopy in a partial transmittance optical geometry, for the assessment of total soluble solids (TSS) of intact rockmelons (*Cucumis melo*) was assessed. The mesocarp TSS was highest around the fruit equator and increased towards the seed cavity. Inner mesocarp TSS levels decreased towards both the proximal and distal ends of the fruit, but more so towards the proximal end. The equatorial region of the fruit was chosen as representative of the fruit for near infra-red assessment of TSS. The spectral window was optimised at 695 to 1,045 nm, and the data pre-treatment procedures were optimised to second derivative absorbance without scatter correction. The 'global' modified partial least squares (MPLS) regression modelling procedure of WINISI (ver. 1.04) was found to be superior with respect to root mean square error of prediction (RMSEP) and *bias* for model predictions of TSS across seasons,

---

compared to the 'local' MPLS regression procedure. Updating of the model with samples selected randomly from the independent validation population demonstrated improvement in both RMSEP and *bias* with addition of approximately 15 samples.

## INTRODUCTION

Consumer acceptability for rockmelons (*Cucumis melo*) is correlated with sugar concentration at harvest (Lester and Shellie 1992), although the presence of various volatile compounds is also relevant (Yamaguchi *et al.* 1977). Thus high sugar concentrations do not totally define eating quality, however, the absence of high total soluble solids (TSS) makes good eating quality highly unlikely. Mutton *et al.* (1981) established a level of 10% total soluble solids as a minimum standard to ensure adequate eating quality of rockmelons for the Australian fresh fruit market.

The assessment of the TSS of intact melon fruit using near infra-red spectroscopy (NIRS) using a halogen light source was first reported by Dull *et al.* (1989). A reflectance optical geometry was used, with a root mean square error of prediction (RMSEP) of 2.2 % reported (on a single population (variety not stated)) of fruit (standard deviation (SD) not stated). Reflectance optics were also used by Birth *et al.* (1990), again with a RMSEP of 2.2 % (on a single population (variety not stated) of fruit (SD not stated). This group continued this line of work (using second derivative spectral data), reporting a RMSEP of 1.9 % (on a single population of fruit (SD not stated)) with a range of 5.6 – 13.1% TSS (Dull *et al.* 1990). Aoki *et al.* (1996) reported use of transmission optics and a RMSEP of 0.4% TSS (on a population of fruit with SD of 0.76% TSS). Work in Australia followed, with Guthrie *et al.* (1998) using a reflectance geometry with a NIRSystems 6500 to achieve a RMSEP of 0.9 % TSS (on a combined variety population of fruit with a SD of 3.2%

TSS). Greensill and Walsh (2000) reported the use of a non-contact interactance geometry for assessment of melon TSS. In these manuscripts a RMSEP of around 0.7% TSS, on a population with a SD of 1.3% TSS, was reported.

Commercial application of NIRS to fruit sorting was first initiated by the Japanese companies Fantec and Mitsui Metals and Mining in 1999 (Kawano 1994a). Mitsui product literature reports an "accuracy" (interpreted as RMSEP) of 0.5% TSS for peach, apple and Japanese pear. Other companies followed, with Sumitomo Metals and Mining bringing a diode laser based system onto the Japanese market, and Colour Vision Systems (Bacchus Marsh, Australia) bringing a system based on the non-contact interactance geometry mentioned earlier onto the Australian market in 2000. Sumitomo literature quotes a "SEP" of > 0.8 % (no SD reported) with "no bias errors", while Colour Vision Systems claim a RMSEP of < 1%. The Sumitomo product appears to be no longer commercially offered. The demise of this product may reflect the high price of the equipment (approx $A1,000,000 for a single lane pack line incorporating the NIR sensor) or equipment instability problems (laser output instability, temperature control).

The above manuscripts (and commercial product literature) focus on reporting calibration model statistics or a prediction of a subset of the population from which the calibration population is drawn. Model performance in validation of new populations, not included in the calibration population, is a harsher exercise. Equipment manufacturers appear to rely on frequent recalibration to address this issue, with the exception of the Sumitomo laser unit, which was marketed on the basis of calibration model stability.

Model updating is practiced in a wide range of NIR application areas, generally involving selection of representative samples of the new population for addition to the

calibration population (see Chapter 1). In a previous exercise with mandarins (Chapter 4), it was demonstrated that model updating following addition to the calibration population of a relatively small number of samples ($< 20$) chosen by any procedure (e.g. random) from the validation population was a successful strategy.

Model robustness across populations of fruit has been noted to be sensitive to changes in growing condition and variety (see Chapters 1, 4). In Chapter 4, model performance for prediction of intact mandarin TSS was shown to be more sensitive to year than to time within a growing season or growing location. In melon Ito *et al*. (1999) noted that individual varietal calibration models are successful when predicting Brix on the same variety within the same season. He reported on the use of a calibration model developed in 1 season to predict a single lot from the subsequent season. The only published work on TSS model robustness in melons is that of Guthrie and Walsh (1998). In this work stepwise multiple linear regression (MLR) models were developed using between 3 and 5 wavelengths, rather than modified partial least squares (MPLS) regression of the full spectrum, from reflectance spectra. Model performance was reasonable across some, but not all, varieties, and model performance markedly deteriorated across populations of the same varieties harvested at different times within a season. The reflectance technique is likely to optically sample the fruit to a depth of approximately 5 mm only (e.g. Lammertyn *et al*. (2000)). The low coefficient of determination for the calibration population ($R_c^2$), coefficient of determination of the validation population ($R_v^2$) and higher RMSEP obtained for particular varieties (e.g. 'Eastern Star' and 'El Dorado') were attributed to the nature of the skin (irregular and thicker epidermal layers) of these varieties. Further, as there is a poor correlation between skin and inner mesocarp TSS, Guthrie and Walsh (1998) suggested that the correlation of mesocarp TSS with spectral data

may represent a secondary correlation with another constituent of the skin and green flesh layers.  A breakdown in this secondary correlation could be responsible for the lack of model robustness.

Sugiyama (1999) reported that absorbance at 676 nm was closely correlated with TSS, and used this character in imaging of TSS distributed across the cut melon surface.  Absorbance at 676 nm presumably acts as an index of chlorophyll content, with TSS related to chlorophyll content indirectly, through fruit maturity.  However, these models predicted poorly across different varieties and subsequent work focussed on the use of spectroscopically justified wavelengths involving the use of the second derivative at 880 and 910 nm (Tsuta *et al.* 2002).

In this chapter we further examine the issue of rockmelon TSS model robustness, and the use of model updating procedures using small numbers of samples from the new population.  Thus this work extends the consideration of Guthrie *et al.* (1998) using the partial transmittance instrumentation characterised in Chapter 2, and the model updating procedure described in Chapter 4.

## MATERIALS AND METHODS

### Plant material

Rockmelon fruit (*Cucumis melo* (L) varieties 'Eastern Star', 'Hammersley', 'Doubloon', 'Highline', 'Malibu', 'Mission', 'El Dorado', 'Colusa', 'Sterling' and 'Hotshot' were sourced after commercial harvest from growers in 4 growing districts (Bundaberg (24.9° S, 152.3° E), Chinchilla (26.5° S, 150.4° E), Gumlu (19.9° S, 147.7° E) and Kununurra (15.5° S, 128.4° E)).  In all, 22 populations of various varieties of rockmelons (each of approximately 100 fruit), obtained over different

seasons and growing districts, were used for spectral acquisition and then assessed for TSS. A spectrum was collected per fruit from the equatorial region (defined relative to the stem – calyx axis) and the area scanned was then excised from the fruit (60 mm diameter core to a depth of 20 mm), skin removed, crushed to extract the juice and the TSS assessed using digital refractometry (Bellingham and Stanley RMF 320). All populations were alphabetically named within each year (Table 1).

Five rockmelon fruit (variety 'Doubloon') were assessed for TSS distribution within the mesocarp (Fig. 1a). Twelve slices, each of approximately 1 cm thickness, were taken from the distal (calyx) end through to the proximal (stem) end of each fruit. Twelve TSS measurements per slice (6 from the outer and 6 from the inner mesocarp, with sampling at and between each of the 3 vascular tissue areas. Data was normalised to the TSS concentration at slice 6 (equatorial) of the inner mesocarp (Fig. 1b).

The penetration of near infra-red radiation through an intact fruit was visualised by illuminating 1 side of the fruit equator with a 100 Watt halogen lamp and monitoring 180° to this point with a Zeiss MMS1 spectrometer. The fibre optic bundle of the MMS1 was moved over a grid pattern, detecting light at the wavelength of maximum detected intensity (820 nm) from a 1 cm$^2$ area of the fruit flesh surface as the fruit was consecutively cut away in 1 cm thicknesses from the opposite end to the light source (in the plane perpendicular to the lamp – fruit axis) (Fig. 2). The results of the light penetration were plotted using a volumetric-data visualisation software package (Slicer Dicer ver. 3.0).

**Table 1.** **Population and NIR TSS calibration statistics using MPLS 'global' procedure for 22 populations of 10 different varieties of rockmelon fruit harvested over 3 years and from 4 growing districts within Queensland and Western Australia.**

| Population | Variety | Growing location | *n* | *Mean* TSS | SD | MPLS factors | $R_c^2$ | RMSECV | SDR |
|---|---|---|---|---|---|---|---|---|---|
| 1998a | Eastern star | Bundaberg | 75 | 8.0 | 1.42 | 7 | 0.84 | 0.76 | 1.9 |
| 1998b | Hammersley | Chinchilla | 80 | 7.5 | 1.16 | 8 | 0.67 | 0.92 | 1.3 |
| 1999a | Eastern star | Gumlu | 71 | 9.0 | 1.10 | 6 | 0.72 | 0.81 | 1.4 |
| 1999b | Eastern star | Gumlu | 80 | 9.4 | 0.93 | 8 | 0.66 | 0.84 | 1.1 |
| 1999c | Eastern star | Gumlu | 90 | 8.9 | 1.03 | 8 | 0.83 | 0.63 | 1.6 |
| 1999d | Eastern star | Gumlu | 78 | 9.7 | 0.95 | 6 | 0.69 | 0.67 | 1.4 |
| 1999e | Eastern star | Gumlu | 100 | 9.3 | 1.91 | 9 | 0.85 | 1.1 | 1.8 |
| 1999f | Hammersley | Gumlu | 80 | 7.6 | 1.12 | 7 | 0.67 | 0.91 | 1.2 |
| 1999g | Doubloon | Chinchilla | 85 | 8.6 | 1.37 | 6 | 0.63 | 1.23 | 1.1 |
| 1999h | Doubloon | Chinchilla | 86 | 8.3 | 1.25 | 7 | 0.76 | 0.73 | 1.7 |
| 1999i | Doubloon | Chinchilla | 99 | 7.6 | 0.93 | 13 | 0.78 | 0.70 | 1.3 |
| 1999j | Doubloon | Chinchilla | 99 | 8.5 | 1.03 | 8 | 0.56 | 0.96 | 1.1 |
| 1999k | Highline | Chinchilla | 90 | 8.8 | 1.10 | 7 | 0.78 | 0.65 | 1.7 |
| 1999l | Malibu | Chinchilla | 92 | 7.5 | 1.01 | 8 | 0.74 | 0.67 | 1.5 |
| 1999m | Mission | Gumlu | 105 | 9.9 | 1.10 | 7 | 0.74 | 0.71 | 1.5 |
| 1999n | El dorado | Gumlu | 130 | 11.8 | 1.66 | 7 | 0.89 | 0.66 | 2.5 |
| 1999o | El dorado | Gumlu | 100 | 9.4 | 1.70 | 7 | 0.80 | 0.95 | 1.8 |
| 2000a | Colusa | Kununurra | 100 | 9.3 | 1.80 | 9 | 0.91 | 0.68 | 2.7 |
| 2000b | Sterling | Kununurra | 96 | 9.3 | 0.88 | 6 | 0.65 | 0.69 | 1.3 |
| 2000c | Hotshot | Gumlu | 80 | 8.3 | 1.32 | 4 | 0.82 | 0.68 | 1.9 |
| 2000d | Hotshot | Kununurra | 100 | 10.3 | 1.00 | 8 | 0.74 | 0.67 | 1.5 |
| 2000e | Hotshot | Kununurra | 99 | 9.5 | 0.95 | 7 | 0.66 | 0.71 | 1.3 |
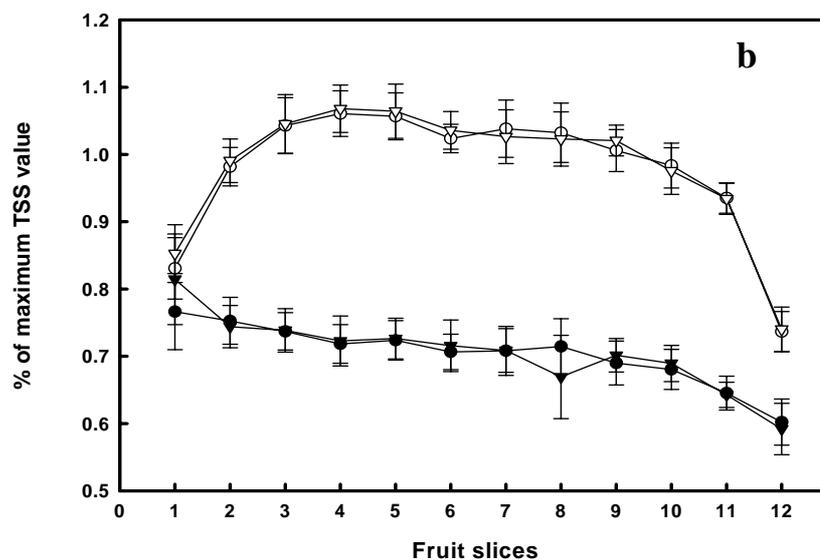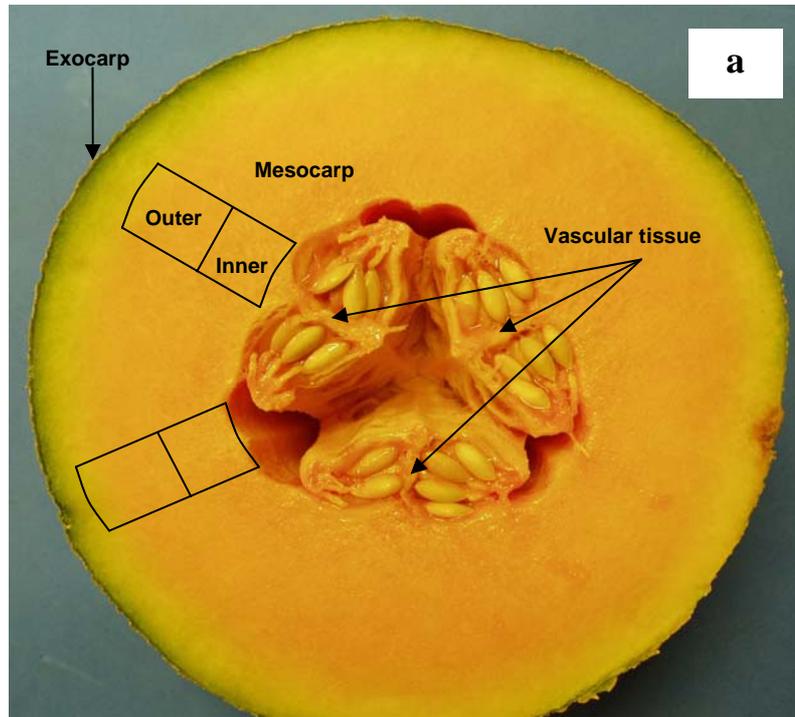
**Figure 1.** Spatial distribution of the attribute of total soluble solids (TSS) in rockmelon fruit (variety 'Doubloon'). (a) Image of fruit indicating location of sampling. (b) TSS distribution. Fruit were cut transversely into 12 slices. Each slice was assessed for % TSS in the inner (open symbols) and outer mesocarp (solid symbols), adjacent (triangles) and between (circle) vascular tissue (12 locations per slice). Data points represent the average of 30 values (6 positions per fruit, 5 fruit), normalised to the TSS concentration at slice 6 (equatorial) of the inner mesocarp. Slices are numbered from proximal to distal end of the fruit.
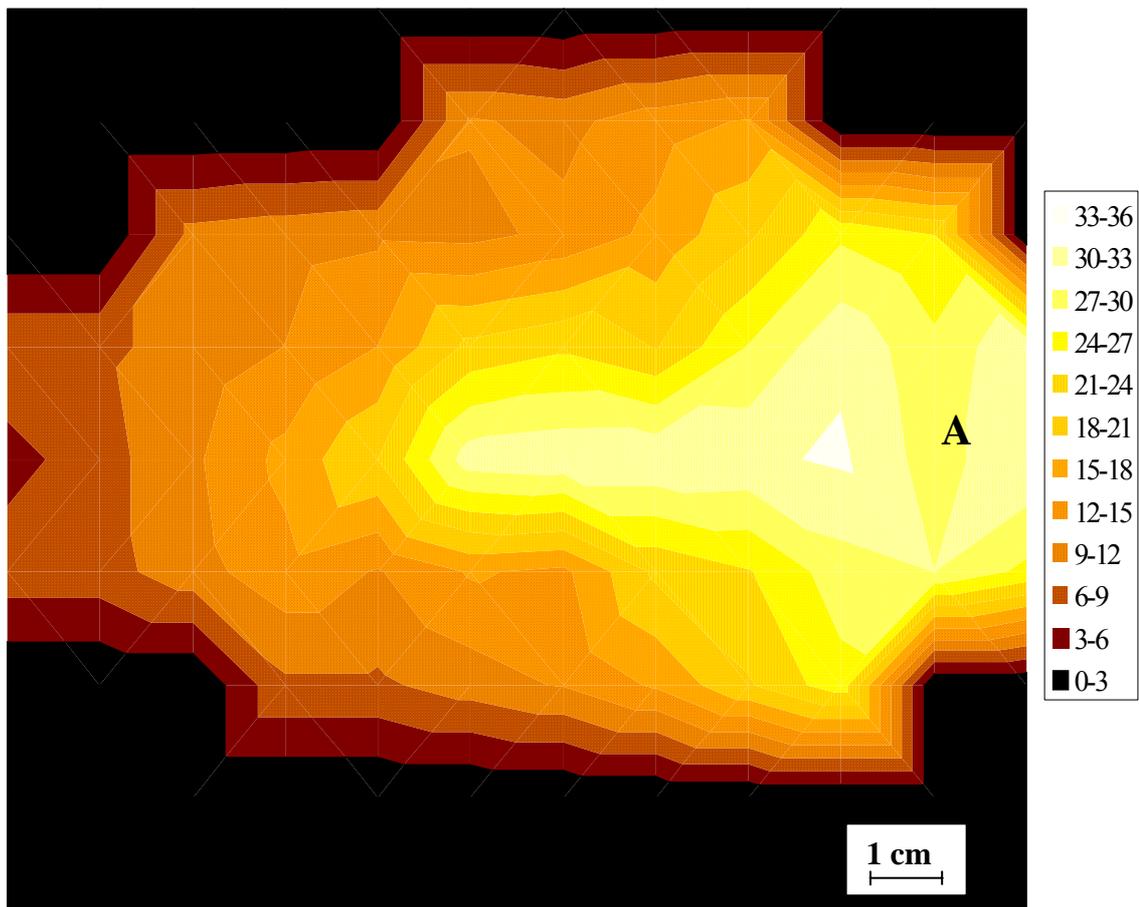
**Figure 2.** **Light penetration through a rockmelon fruit. Light (100 Watt tungsten-halogen lamp) illuminated position A on the equator of the fruit. Data of a virtual longitudinal slice in the plane of the lamp – fruit axis (for a fruit position with stem end uppermost) is presented. Scale bar is in arbitrary units of detector analogue to digital counts.**

## Spectral acquisition

Spectra were collected over the wavelength range 306 – 1,130 nm using a NIR enhanced Zeiss MMS1 spectrometer (photo-diode array, comprising 256 silicon detectors with a resolution of approximately 3.3 nm) and 4 tungsten halogen lamps of 50 Watt each, in a partial transmission optical configuration as described in Walsh *et al.* (2000) (45° angle between illumination source and detector, relative to the fruit centre, with the detector probe in contact with the fruit surface). White (teflon tile)

and dark reference measurements were taken at the start of each experimental run. Spectra were averaged over 4 scans at an integration time of 200 ms per scan.

## Chemometrics

The software package WINISI (ver. 1.04a) (ISI International, USA) was used for all chemometric analysis. WINISI employs a modified partial least squares regression (MPLS) procedure in which the reference data and the spectral data at all wavelengths are scaled to give a standard deviation of 1.0 before calculation of each subsequent MPLS factors (Shenk and Westerhaus 1991). Calibrations were developed using MPLS in both 'global' and 'local' WINISI procedures (based on the combined 1998 and 1999 populations).

In the 'global' procedure a single MPLS calibration model is developed using all the samples in the calibration population (excluding 'outliers'). The 'global' WINISI models were optimised in terms of wavelength interval, derivative condition and scatter correction technique. The significance of differences between both the RMSEP and *bias* of the different models were tested according to the procedures of Fearn (1996), using a significance level of 95%.

In the 'local' procedure, a separate MPLS model is developed for every sample to be predicted. This model is based on a small number of spectra from the calibration population selected on the basis of similarity to the unknown sample (based on a regression of the scores of the unknown sample against those of all members of a sample library) (Shenk *et al* 1997). This procedure results in a model 'over-fitted' in the sense that the model is fit only for use in prediction of 1 unknown sample. The number of samples used (20 − 100, steps of 20), the number of MPLS factors used (5 − 25, steps of 5), and the number of MPLS factors removed (1 − 5,

steps of 1) (125 combinations) was optimised on the basis of lowest RMSEP(C) for validation across the 5 populations from the year 2000 (as used in Table 1) (data not shown). The WINISI software forced the removal of at least 1 MPLS factor, on the basis that the first MPLS factor is indicative of the scattering of light in the sample due to particle size and not the attribute of interest.

Calibration performance for both 'local' and 'global' MPLS procedures was assessed in terms of $R_v^2$, RMSEP, standard deviation ratio (SDR = SD/RMSECV or RMSEP), slope and *bias* of the validation populations as per Guthrie *et al*. (2005a). The SDR facilitates comparison of different populations with differing SD's. The best data pre-treatment (second derivative with gap size of 4 data points either side and no scatter correction, data not shown) was used in all subsequent model development. The RMSECV, RMSEP and *bias* values were tested for significance (P = 0.05) using the procedure of Fearn (1996) (Appendix F).

In a separate exercise, the effect of spectral window on partial least squares (PLS) calibration model performance for TSS was optimised in terms of root mean square error of calibration (RMSEC) using a moving PLS interval algorithm, developed in Matlab (ver. 7.0) – PLS toolbox (ver. 3.5 by Eigenvector) (Guthrie *et al*. 2005a). The combined populations of years 1 and 2 (17 populations comprising 1,467 spectra) were used, employing second derivative (WINISI gap size 4 without scatter correction) absorbance data interpolated to 3 nm steps. The start and end wavelengths of the spectral window were varied from 650 to 950 nm, and 750 to 1,050 nm, respectively, in increments of approximately 3 nm and a map produced that involved 20,200 separate PLS models. For each model, the number of PLS factors was determined on the basis of a 3% difference between RMSEC and RMSECV values. Model performance is reported in terms of RMSEC in this exercise.
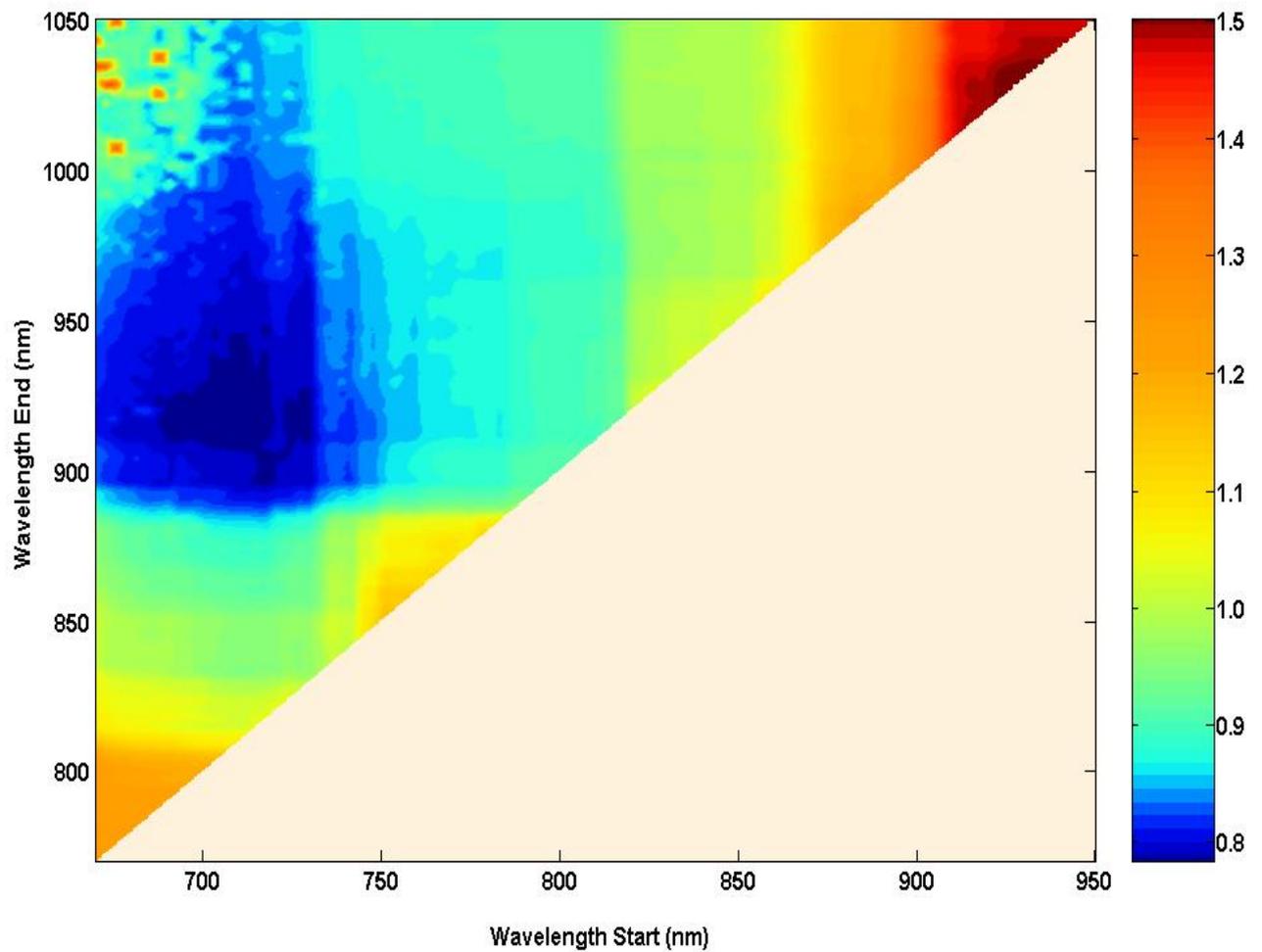
**Figure 3.** Calibration model performance (root mean square standard error of calibration (RMSEC)) for varying spectral windows (start and end wavelengths varied). Partial least squares calibration models were based on the combined populations from year 1998 and 1999 (17 populations) for total soluble solids (mean = 8.9% TSS and SD = 1.73% TSS).

**Table 2.** Optimisation of data pre-treatment in terms of derivative treatment (none, first or second order) and 2 wavelength regions (695 – 1,046, 722 –945 nm) for TSS calibration. Model performance is reported in terms of prediction of 5 independent populations of rockmelon fruit(year 3 harvest). The model was developed using the combination of 17 populations (year 1 and 2 harvests, $n = 1,467$, mean = 8.9%, SD = 1.73% and range of 4.8 - 15.2% TSS).

For each population, the treatment with the lowest overall RMSEP was presented in bold (but not underlined). The corresponding *bias* was also bolded. These values were then compared with that related to the lowest RMSEP in the other wavelength window, and that related to the highest RMSEP value in the other derivative condition of the same wavelength region (value underlined). These values were tested for significance (95% probability level), with significantly different values shown in bold.

| Math treatment | Variance (RMSEP) | | | *bias* | | |
|---|---|---|---|---|---|---|
| | Absorb. | 1st Deriv. | 2nd Deriv. | Absorb. | 1st Deriv. | 2nd Deriv. |
| *Population A* | | | | | | |
| 695-1046 nm | 1.0 | 1.1 | **1.0** | 0.24 | 0.22 | 0.04 |
| 722-945 nm | 1.1DT | *1.3* | **1.0DT** | -0.42 | **0.86** | **0.28** |
| *Population B* | | | | | | |
| 695-1046 nm | 0.7 | 0.6 | 0.6 | 0.18 | 0.05 | -0.09 |
| 722-945 nm | 0.8 | *0.9* | **0.6** | -0.53 | **0.69** | **0.10** |
| *Population C* | | | | | | |
| 695-1046 nm | **1.1** | 1.1 | 1.3 | -0.58 | -0.69 | -0.87 |
| 722-945 nm | 1.1 | **0.9** | *1.4* | -0.68 | **-0.28** | **-1.05** |
| *Population D* | | | | | | |
| 695-1046 nm | 1.2 | 1.2 | 1.0 | 0.85 | 0.84 | **0.70** |
| 722-945 nm | **0.8** | 1.6 | 1.2 | **0.16** | 1.5 | 0.92 |
| *Population E* | | | | | | |
| 695-1046 nm | 0.7 | 0.7 | 0.7 | 0.19 | 0.22 | **0.08** |
| 722-945 nm | 0.8 | 1.1 | **0.7** | -0.40 | **0.92** | **0.37** |

## Model updating

The use of a model to predict attribute levels of a totally independent population is often compromised. The inclusion of a number of samples from the new population into the calibration population, with subsequent redevelopment of the MPLS model ('model updating') is a well accepted procedure to encompass new variation in the

sample. However, there is no rule defining the number of samples required to 'update' the model. Guthrie and Walsh (2001) and Guthrie *et al.* (2005b) reported on the use of a range of sample selection protocols for updating of mandarin TSS models, and the recommended procedures are trialled in the current study. . In those studies, the use of 20 samples for updating was recommended.

Each validation population was initially screened for outliers (defined as samples with a Mahalanobis distance (GH) > 3.0 using the scores and loadings from the validation population (Shenk and Westerhaus 1991)). These outliers were removed (for populations A, B and C – 6, 5 and 7 outliers, respectively) and the resulting data were divided randomly into 2 sets. One set (2/3) was retained as a validation population and the other set (1/3) used for selection of samples for addition to the calibration population. Two validation populations were chosen that exhibited a low and high average GH value (calculated from the scores and loadings of the calibration population), respectively. That is, these 2 populations will each represent a new population similar to, and markedly different from, the calibration population. For addition to the calibration population (comprising 17 populations from years 1999 and 1998), new samples were chosen on a random basis from the subset of the validation population (1/3). Calibration models were then developed using MPLS regression in both 'global' and 'local' WINISI procedures.
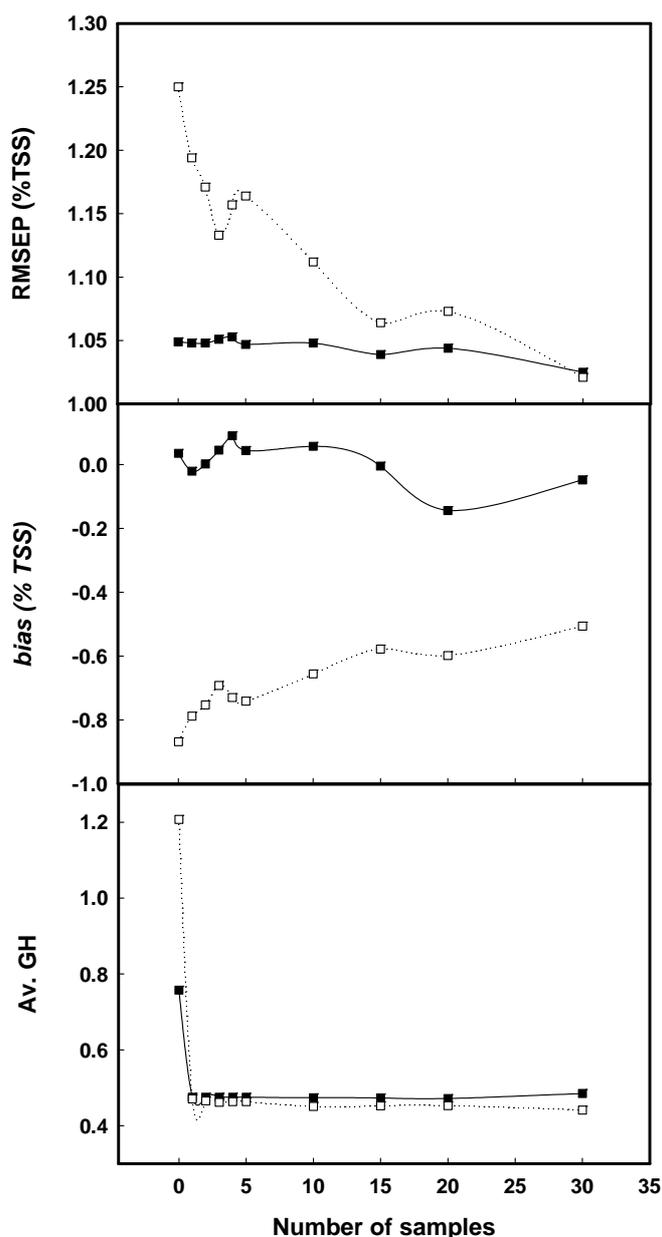
**Figure 4.** Prediction statistics for 'global' WINISI modified partial least squares regression models in prediction of total soluble solids in rockmelon fruit for two independent (of calibration populations) validation populations. The average of global Mahalanobis distance (GH) of samples in the validation population was calculated using calibration model scores. The initial calibration population was comprised of the combined population of years 1998 and 1999 (17 populations). Populations from year 2000 (2 populations, A and C) were used as validation populations, less the 1 third randomly selected samples which were used for 'model updating' (addition to the calibration population). Samples for 'model updating' were chosen at randomly from this population.
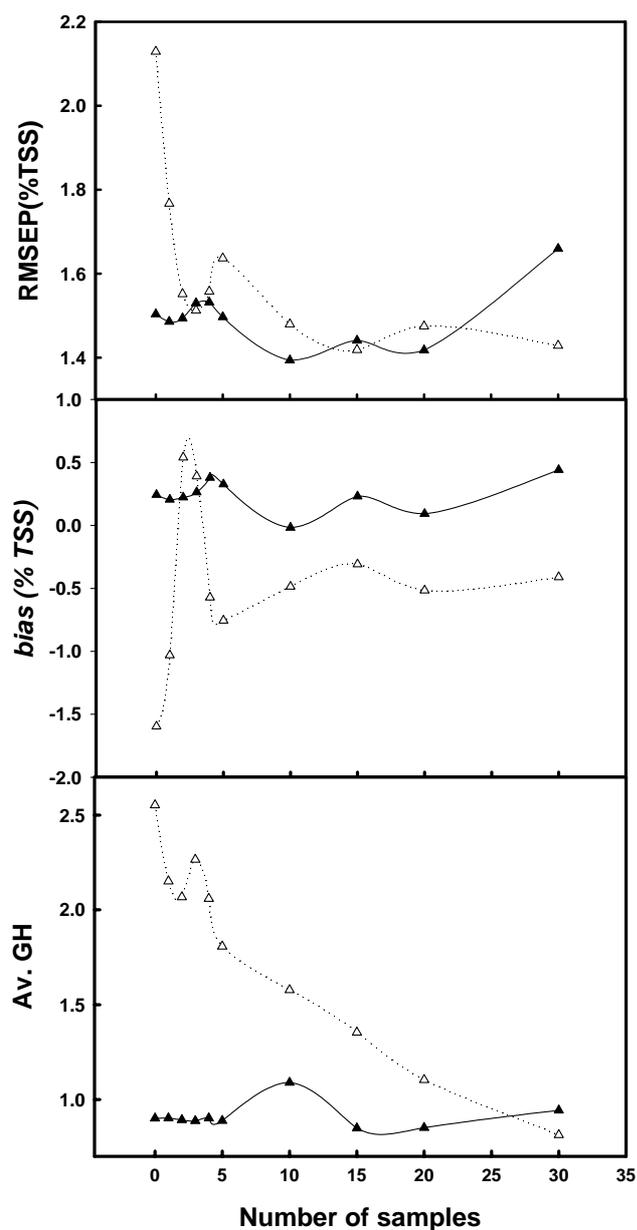
**Figure 5.** Prediction statistics for 'local' WINISI modified partial least squares regression models in prediction of total soluble solids in rockmelon fruit for 2 independent (of calibration populations) validation populations. The data populations used and treatments undertaken were the same as in Fig. 4.

# RESULTS AND DISCUSSION

## Attribute distribution and light penetration

In the inner mesocarp, TSS content was highest around the equator of the fruit and lessened towards both ends, but slightly more so towards the proximal end.

136

However, in the outer mesocarp there was little difference along the length of the fruit. Within the fruit, TSS concentration increased from the skin to the seed cavity and was highest around the vascular bundles near the seed cavity (Fig. 1b). There was little variation in TSS around the equator of the fruit, even when that part of the fruit in contact with the ground was sampled. The difference in TSS levels between various parts of an individual fruit's flesh was as great as 6 to 7% TSS and averaged approximately 4% TSS overall. These observations reinforce the need to take both spectroscopic and reference values from the same area on the fruit.

Ito *et al*. (1999) chose to base spectroscopic assessment of melons on the calyx end because "the flesh in the calyx end of melon fruit is thinner than that in other parts". Certainly the chlorophyll containing mesocarp layer is thinner in this region, and the mesocarp is thinner overall. However, the variation between the inner and outer mesocarp is similar to that in other regions of the fruit (Fig. 1b). To facilitate in-line fruit grading, we chose to assess fruit in an equatorial position of the fruit.

Light penetrated further through the centre of the fruit (i.e. through the seed cavity) than through the mesocarp layers (Fig. 2). Light penetrated through the bulk of the fruit, propagating through the mesocarp layers but was absorbed strongly by the exocarp layers. With a 100 Watt lamp, light was detectable in transmission through the fruit once the skin (5 mm depth) was removed.

We conclude that the skin of the intact fruit presents the greatest obstacle to the penetration of NIR radiation. This is more so than in other fruit (e.g. apples, stone fruit) because of the relative thickness and often irregular (e.g. netted) surface of rockmelon fruit.

## Spectral window and data treatment

The spectral window used in PLS calibration model development for TSS was optimised in terms of RMSEC (Fig. 3) using the moving PLS interval algorithm described above. Low RMSEC values for TSS were obtained for a window beginning between 702 and 709 nm, and finishing between 914 and 920 nm, although acceptable results were obtained up to 1,050 nm (Fig. 3). The minimum RMSEC for TSS (0.26%) was recorded for a start wavelength of 709 nm and a finish wavelength of 917 nm, surrounded by an area (or island) of low RMSEC values, suggesting a stable spectral region for TSS model development.

Although the RMSEC was used to compare the relative spectral regions, the final 'proof' of a model is in its validation statistics rather than the calibration statistics. Using a model developed utilising a combined population of years 1 (1998) and 2 (1999) (17 populations across varieties and growing districts) to predict the 5 populations from year 3 (2000), the best data pre-treatment routine was second derivative absorbance data without scatter correction (Table 2). The use of derivative spectra is a very effective method for removing both the baseline offset and the slope from a spectrum (Norris 1982). The full short wave NIR region (695 to 1,045 nm) was not significantly better than the restricted region (722 – 945 nm) in terms of RMSEP, but was significantly better (5% level) in regard to *bias* in 4 out of the 5 independent validation populations (Table 2). Both windows encompass the spectrally significant 910 nm third overtone of CH stretching (which will primarily be due to sucrose in melon fruit). Therefore, second derivative absorbance data without scatter correction, over a 695 – 1,045 nm wavelength range, was subsequently used in all calibration model development.

Calibration model statistics varied across the 22 populations, with root mean square error of cross validation (RMSECV) ranging from 0.63 to 1.2% TSS and the SDR from 1.1 to a high of 2.7 (Table 1). In prediction of independent data sets, the models gave similar or better results (in terms of RMSEP, see Tables 2,3) than that obtained by other researchers [e.g. Dull *et al*. (1990), RMSEP of 1.9% TSS; Guthrie *et al*. (1998), RMSEP of 0.9% TSS and Greensill and Walsh (2000), RMSEP of 0.7% TSS] but were inferior to those obtained by Aoki *et al.* (1996), RMSEP of 0.4% TSS). However, Aoki *et al*. (1996) employed a population with a lower SD (0.76%TSS) than any population in the current study.

**Table 3.** **Prediction statistics (RMSEP and *bias* % TSS) for validation populations A and C from year 2000 predicted by the calibration from the combined population of years 1998 and 1999 (17 populations) using both 'global' and 'local' WINISI MPLS procedures.**

| MPLS calibration procedure | Population A SD = 1.80% TSS | | Population C SD = 1.44% TSS | |
|---|---|---|---|---|
| | RMSEP | *bias* | RMSEP | *bias* |
| **Global** | 1.1 | 0.04 | 1.3 | -0.87 |
| **Local** | 1.5 | 0.24 | 2.1 | -1.60 |

## Regression method

The settings for the 'local' procedure were optimised at 100 samples, 25 MPLS factors and 1 MPLS factor removed. As noted above the apparent model 'over-fitting' inherent in the use of 25 MPLS factors is not an issue because of the use of the model for the prediction of the 1 sample only. However, a limitation of the WINISI 'local' procedure is that it forces the elimination of the first MPLS factor. This step may be logical for reflectance spectra in which the first factor can be associated with

explanation of scattering properties of the sample. However, this step is not logical for partial transmission spectra as used in this study. Using the MPLS 'global' procedure the residuals (predicted – actual) for the prediction of populations C were independent of analyte level, while those for population A increased with analyte levels (data not shown). It was therefore expected that the 'local' procedure might improve prediction in the case of population A. The prediction results of the same independent validation populations (Populations A and C from 2000) obtained from calibration models developed with the WINISI MPLS 'global' procedure were better than those obtained from the 'local' procedure (Table 3). The use of alternative to 'global' MPLS, including 'local' neural networks and support vector machines should be further explored.

## Model updating

As expected, the prediction performance of a model on an independent population was determined by the average GH of that population (calculated using calibration model scores and loadings) (Fig. 4). Better initial prediction occurred when the average GH value was low. Model updating, using a small number of samples from the validation population, offers an alternative to the development of a entirely new model.

The effect of model updating to improve the prediction performance of an independent validation population depends on the average GH value of the validation population A greater improvement in both prediction RMSEP and *bias* values is demonstrated in both the WINISI procedures of 'global' and 'local' MPLS calibration models following sample addition (model updating) when the validation population exhibited a higher initial GH value (Figs. 4 and 5). A stabilisation of both the

RMSEP and *bias* values occurred with the addition of approximately 15 samples (Figs. 4) for both 'global' and 'local' WINISI procedures (Fig. 5).

This result is consistent with the work reported for mandarins (Chapter 4). The greater the effect of sample addition - model updating for a high GH validation population is ascribed to a greater leverage on the MPLS calibration. However, it is surprising that so few samples (i.e. $n = 15$) can influence a large calibration population ($n = 1,467$).

## CONCLUSIONS

Model performance for prediction of TSS in intact rockmelon fruit was inferior to previous work on mandarins (Chapters 3 and 4). This was ascribed to the heterogenous distribution of TSS (sugars) within the fruit and poor penetration of light through the irregular fruit skin.

Data selection and pre-treatment was optimised in terms of prediction performance. The use of second derivative absorbance data with scatter correction using a spectral window optimised for each application, is recommended. The WINISI MPLS calibration procedure of 'global' was superior in terms of RMSEP and *bias* to the 'local' procedure in prediction of independent validation populations. Prediction statistics (RMSEP and *bias*) can be improved with the addition of approximately 15 samples from the validation population, as found for mandarins (Chapter 4) and for intact macadamia kernels (Chapter 6).

# 6

# MODEL DEVELOPMENT AND ROBUSTNESS IN PREDICTION OF QUALITY ATTRIBUTES OF MACADAMIA KERNELS[5]

## ABSTRACT

Spectral data were collected of intact single and ground kernels using 3 instruments (using Si-PbS, Si and InGaAs detectors), operating over different areas of the spectrum (between 400 and 2,500 nm) and employing transmittance, interactance, and reflectance sample presentation strategies. Kernels were assessed on the basis of oil and water content, and with respect to the defect categories of insect damage, rancidity, discolouration, mould growth, germination, and decomposition. Model performance statistics for oil concentration models were acceptable on all instruments ($R_c^2 > 0.94$; RMSECV < 2.5%, which is similar to reference analysis error), although that for the instrument employing reflectance optics was inferior to models developed for the instruments employing transmission optics. However, these models performed poorly ($R_c^2$ 0.92, RMSECV $\geq$ 4.0%) in prediction of oil concentration of kernels of a second population acquired in a subsequent season. Model updating by addition of less than 10 samples from the validation population improved prediction of the second population with error levels similar ($R_v^2 > 0.92$, RMSEP 4.7%, *bias* 0.6% oil)

---

to the calibration statistics. The spectral positions for calibration coefficients for the oil models were consistent with absorbance due to the third overtones of $CH_2$ stretching. Calibration models for moisture concentration in ground samples were acceptable on all instruments ($R_c^2 > 0.97$; RMSECV <0.2 %), while calibration models for this attribute in intact kernels were relatively poor. Calibration coefficients were more highly weighted around 1,360, 740 and 840 nm, consistent with absorbance due to overtones of OH stretching and combination. Kernels with brown centres or rancidity could be discriminated from each other and from sound kernels using principal component analysis. Part kernels affected by insect damage, discolouration, mould growth, germination, and decomposition could be discriminated from sound kernels. However, discrimination among these defect categories was not distinct and could not be validated on an independent population.

A low cost Si photodiode array instrument is recommended for its potential to assess the oil and moisture concentration of intact macadamia kernels. However, further work is required to examine predictive model robustness across different populations, including growing districts, cultivars and times.


## INTRODUCTION

Payment to growers for macadamia nut-in-shell is based on the weight of nut-in-shell at 10% moisture concentration and on the percentage of sound (defect-free) kernels, determined by the processor on representative samples from each consignment. Penalties are imposed if the percentage of unsound kernels exceeds 3.5%. Sorting of kernels into sound and the various unsound categories for grower payment is currently done by subjective visual evaluation. The industry would benefit

from objective tests that could be used in the processing plant quality assurance laboratory.

A major defect of macadamia nuts is kernel immaturity. Immaturity is currently qualitatively identified using visual criteria for the purpose of grower payments. This defect can be quantified by measuring oil concentration (Ripperton *et al.* 1938); (Himstedt 2002). Other quality defects include insect damage, rancidity, mould growth, decomposition, germination and discolouration. These defects are also assessed visually. Discolouration in this context is that which is due to causes other than insect damage, mould growth, decomposition or germination. Moisture concentration is also an important quality parameter, having a major effect on shelf life (Cavaletto *et al.* 1966).

Kernel moisture content has a large effect on shelf-life. Processing of macadamia nuts requires factory drying from approximately 10% nut-in-shell (4% kernel) moisture to 4% nut-in-shell (1.5% kernel) moisture. Kernel moisture concentration is currently assessed during the drying process by using an oven drying method. Determination of kernel moisture by near infra-red spectroscopy (NIRS) would be much more rapid and would therefore be beneficial to the industry.

Near infra-red spectroscopy is widely employed for oil and moisture determination in the oil seed and grain industries. For example, Tillmann *et al.* (2000) report a $R_c^2$ of 0.95 and a standard error of cross validation (SECV) of 0.83% for oil concentration of whole canola seed, and Williams and Sobering (1993) a $R_v^2$ of 0.95 and standard error of prediction (RMSEP) of 0.35% for moisture concentration of whole barley seed. Further, Ha *et al.* (1998) report a $R_c^2 > 0.9$ on a range of specific fatty acids in sesame seed oil. Near infra-red spectroscopy has also been used to assess various attributes of oil seed quality (e.g. acid value, peroxide value as

indicators of rancidity) (Cho *et al.* (1998) and Ha *et al.* (1998)).  Other workers have used NIRS to assess malting barley grain for fungal contamination resulting in grain discolouration (calibrating against grain brightness – L*) (Fox *et al.* 2000).  The technique can also be used for qualitative purposes (e.g. the detection of insect damage in whole wheat kernels by measuring the amount of reflected versus absorbed light) (Chambers *et al.* 1994).

For oil, strong electromagnetic absorption is reported around 2,200 to 2,400 nm ($CH_2$ stretch bend and combinations), with weaker absorption around 1,750, 1,200 and 900 nm (first, second and third overtones of $CH_2$ stretching) (Osborne *et al.* 1993).  However, shorter wavelengths allow better penetration of biological samples (Kawano *et al.* 1994b), and as such, shorter wavelengths should be useful in assessment of whole macadamia kernels.

The aim of this project was to assess the feasibility of using NIRS as an objective analytical method to replace the existing subjective methods for detection of kernel defects.  Typically, the feasibility of using NIRS for a given application is assessed using laboratory grade instrumentation that is unsuitable for industry use with regard to cost and complexity.  In this project we have drawn on our prior expertise with modular, low cost instrumentation (Greensill and Walsh 2000); (Walsh *et al.* 2000), comparing their performance with laboratory grade instrumentation.  It was also hypothesised that model robustness would be higher for this application (oil concentration of a low moisture product) than for sugar concentration of that of models developed for the intact fresh fruit (high moisture products).

## MATERIALS AND METHODS

## Sampling

Raw kernels were collected over several weeks of the 2002 and 2004 seasons by a commercial processor. Four populations of kernels were assembled for calibration development with respect to oil concentration (1 population in each year), moisture concentration, and defects, respectively. Samples were bulk packed into evacuated, heat-sealed, foil laminate bags and stored at 4ºC between collection and analysis.

Populations of 200 kernels (100 mature and 100 immature and free from other defects) in 2002 and 110 kernels (55 mature and 55 immature and free from other defects) in 2004 were utilised for oil concentration models. The use of both mature and immature kernels greatly increased the range of oil concentrations studied (Table 1), desirable for reliable predictive modelling. A third population ($n = 105$) of sound, mature kernels was used for the moisture concentration study. Kernels with a range of moisture concentrations were obtained by taking nut-in-shell samples at various stages during the factory (approximately 5-8 days) drying process (Table 2). Of this population, 105 kernels were used intact, while 35 kernels were ground to pass a 2 mm screen in a Zyliss CH3250 grinder. Kernels were removed from the bulk packs and packaged individually in heat-sealed low density polyethylene bags (from a single manufacturing batch), and were stored refrigerated in individual sample jars, to maintain kernel moisture levels. Spectra were acquired of kernels in their individual low density polyethylene bags. Spectra were also acquired of ground samples ($n = 35$), in individual low density polyethylene bags packed to an approximate depth of 5 mm. The contribution of the low density polyethylene bags to the acquired spectrum of each sample was assumed to be constant.

**Table 1.** Oil concentration, kernel weight and height of 2 populations of macadamia kernels (half population mature and half immature kernels). Population 1 was assessed in 2002 and population 2 in 2004.

| Attribute | Mature | Immature | Total |
|---|---|---|---|
| **Population 1 (Calibration population)** | | | |
| Oil (%) Mean $\pm$ SD | 75.7 $\pm$ 2.1 | 46.2 $\pm$ 10.7 | 60.7 $\pm$ 16.84 |
| Range | 67.8-81.0 | 18.9-70.2 | 18.9 – 81.0 |
| Weight (g) $\pm$ SD | 2.4 $\pm$ 0.68 | 1.3 $\pm$ 0.24 | 1.85 $\pm$ 0.75 |
| Height (mm) $\pm$ SD | 14.1$\pm$ 1.35 | 11.3 $\pm$ 1.26 | 12.7 $\pm$ 1.93 |
| **Population 2 (Validation population)** | | | |
| Oil (%) Mean $\pm$ SD | 70.3 $\pm$ 4.4 | 47.9 $\pm$ 9.7 | 59.1$\pm$ 15.4 |
| Range | 57.3-77.7 | 29.4-66.9 | 20.1-77.7 |
| Weight (g) $\pm$ SD | 2.2 $\pm$ 0.28 | 1.4 $\pm$ 0.38 | 1.80 $\pm$ 0.51 |
| Height (mm) $\pm$ SD | 14.0 $\pm$ 0.99 | 12.0 $\pm$ 1.44 | 13.0 $\pm$ 1.59 |

**Table 2.** Moisture concentration of 105 intact and 35 ground samples.

| Attribute | Intact | Ground |
|---|---|---|
| **Moisture (%)** | | |
| **Mean $\pm$ SD** | 1.92 $\pm$ 0.33 | 1.98 $\pm$ 0.34 |
| **Range** | 1.4 - 4.2 | 1.5 - 2.9 |

In a separate exercise in 2002, spectra were collected of 20 mature kernels of each of the 8 categories: sound, affected by mould, brown centres, insect damage, decomposition, germination, discolouration, and rancidity; ($n = 160$).

## Instrumentation

Three instruments, operating over different areas of the electromagnetic spectrum (Table 3) were used to collect spectra of intact and ground kernels. Instrumentation was powered on 2 hours before spectral acquisition to ensure operational stability of both light source and detector.

The Si-PbS system was used in 2 differing modes of reflectance. One incorporated the use of the Foss NIRSystems remote reflectance probe (operating between 400 – 1,900 nm) with intact kernels, and the other, a spinning cup module (operating between 400 – 2,500 nm) with ground samples.

The InGaAs and Si systems were used in a full transmission configuration, on intact kernels with respect to oil, moisture, brown centres, and rancidity. A single 50 W Philips 402004 tungsten halogen lamp was mounted at 180º with respect to the sample and the detector fibre optic. Ground samples sealed in plastic bags were placed on an aluminium plate with a 7 mm diameter hole, between detector and light source.

**Table 3.**       **Description of Instrumentation.**

| Detector Type | Instrument | Wavelength (nm) | Optical configuration |
|---|---|---|---|
| Silicon-Lead Sulphide (Si-PbS) | Foss NIRSystems 6500 | 400 – 2500 | Diffuse reflectance (remote reflectance and spinning cup accessories) |
| Silicon (Si) | Zeiss MMS1-NIR Enhanced | 300 –1100 | Transmittance Interactance |
| Indium Gallium Arsenide (InGaAs) | Zeiss MMSNIR | 800 – 1700 | Transmittance Interactance |

The InGaAs and Si systems also used an interactance configuration (Ocean Optics bifurcated optical fibre held 1 mm from kernel surface), on intact kernels affected by mould, brown centres (part kernels), insect damage, decomposition,

germination, and discolouration. This technique is a highly localised and useful in exploratory studies such as this where defect area is small in comparison to non-affected areas.

The Si-PbS instrument was operated using NSAS software, while the Si and InGaAs instruments were operated with in-house developed Labview based software. The NSAS default setting (average of 32 scans) was adopted in collecting each spectrum for the Si-PbS system, while 20 scans were averaged for the InGaAs and Si systems (for a discussion for signal averaging, signal to noise ratio and its effect on calibration model performance, see Guthrie and Walsh (1999) and Greensill and Walsh (2000). Integration times were adjusted to achieve count levels above 50% of detector saturation (20 to 80 ms) for the InGaAs and Si systems.

## Reference analysis (Oil Concentration, Moisture, and Defect)

Reference analysis was undertaken using a micro-soxhlet apparatus (standard error of +/- 3% from the mean) for oil concentration and a TGA-601 Thermogravimetric Analyser (LECO Corporation) for moisture concentration (s.e. of $\pm$ 2% from the mean). Defect samples were supplied by the commercial processor, and samples were assessed for defect category by staff accredited for subjective assessment of macadamia kernel defects by the Australian Macadamia Society.

## Chemometric analysis

WINISI (ver. 1.5) and The Unscrambler (ver. 7.5) software packages were used for chemometric analyses. Calibration models were developed using absorbance, first and second derivatives of absorbance, with or without scatter correction (standard

normal variance and / or detrend) (twelve combinations of treatments, as per Guthrie (1998)). Calibration development was based on the WINISI Modified Partial Least Squares regression technique.

On the population for 2004, partial least squares regressions using the 2 variables of oil concentration (%) and kernel height (mm) singly (PLS1) and jointly (PLS2) as dependent variables were also undertaken using The Unscrambler (ver. 7.5) chemometric software package.

Calibration performance is reported in terms of the $R^2$ of the calibration ($R_c^2$), the $R^2$ of a 6 group cross validation procedure ($R_v^2$), the root mean square of the standard error of the cross validation (RMSECV), and the ratio of population standard deviation to RMSECV (SDR).

In the assessment of different instruments, the same population of kernels was presented to all instruments. However, sample mis-presentation or signal saturation resulted in unequal sample numbers (Table 4).

**Table 4.      Oil and moisture PLS calibration results for 3 instruments. Calibration models were developed for each attribute/instrument using a factorial combination of derivative condition (0, 1, 2 i.e. raw absorbance data, first or second derivative of absorbance data), and scatter correction routines (standard normal variance, detrend or both (SC) with the best results reported here. Calibration results are reported in terms of number of spectra, number of outliers removed, standard deviation (SD) of the assessed population, $R^2$ of the calibration and validation populations ($R_c^2$, $R_v^2$, respectively), RMSECV and SDR (SD/RMSECV).**

| Instrument | Math | #PC | n(#outliers) | SD | $R_c^2$ | $R_v^2$ | RMSECV | SDR |
|---|---|---|---|---|---|---|---|---|
| *Oil %-intact* | | | | | | | | |
| SiPbS | 1 SC | 7 | 199(7) | 16.3 | 0.94 | 0.91 | 5.3 | 3.1 |
| Si | 0 SC | 11 | 199(19) | 15.4 | 0.98 | 0.98 | 2.4 | 6.5 |
| InGaAs | 0 SC | 8 | 199(16) | 15.8 | 0.99 | 0.99 | 1.7 | 9.3 |
| *Moisture %-intact* | | | | | | | | |
| SiPbS | 1 | 6 | 105(5) | 0.25 | 0.89 | 0.79 | 0.11 | 2.2 |
| Si | 2 SC | 9 | 98(11) | 0.25 | 0.78 | 0.59 | 0.16 | 1.6 |
| InGaAs | 2 SC | 4 | 102(4) | 0.34 | 0.37 | 0.29 | 0.29 | 1.2 |
| *Moisture %-ground* | | | | | | | | |
| SiPbS | 1 SC | 3 | 35(5) | 0.28 | 0.97 | 0.95 | 0.06 | 4.4 |
| Si | 2 SC | 1 | 34(1) | 0.30 | 0.54 | 0.45 | 0.22 | 1.3 |
| InGaAs | 0 SC | 2 | 32(1) | 0.33 | 0.58 | 0.31 | 0.28 | 1.2 |

Further, different numbers of outlier spectra – reference values were identified and removed from each calibration population. Therefore comparison of $R_c^2$, $R_v^2$, and RMSECV between calibrations must be tempered by consideration of standard deviation (SD). The standard deviation ratio (SDR) statistic, as a ratio of SD to RMSECV, is useful in this connection.

To achieve a statistically valid comparison between data of populations acquired using 2 instrument platforms, all samples giving rejected and outlier values were eliminated from both population sets, and residuals used to compare RMSECV values at a 95% confidence interval using Fearn's criteria (Fearn 1996). A Microsoft Excel ® spreadsheet was developed to implement this procedure and is available from the author (john.guthrie@dpi.qld.gov.au) (Appendix F).

The WINISI discriminant analysis routine was used in an attempt to differentiate spectra of whole kernels scored for soundness, rancidity and brown centres, and spectra of part kernels affected by insect damage, discolouration, mould growth, germination, and decomposition.

## RESULTS AND DISCUSSION

### Kernel Characteristics

Immature kernels were smaller, with an average weight of only 54 and 64% of mature kernels, and contained only 62 and 68% of the specific oil concentration (%w/w) of mature kernels for populations from 2002 and 2004, respectively (Table 1). Thus immature kernels contained only 34 and 44% of the oil concentration of mature kernels (on a per kernel basis), respectively. The mean diameter of mature kernels used in this study was 16.3 mm and the mean diameter of immature kernels was 9.6 mm.

The current industry standard for kernel maturity is 72% or higher oil concentration (Ripperton *et al.* 1938) and (Himstedt 2002). The supplying processor selected obviously mature and immature kernels for the population acquired in 2002, and a low error rate was achieved (only 3 classified as mature had less than 72% oil concentration, at 68, 69 and 70%), and none of the kernels classified as immature had greater than 72% oil concentration. This selection process is reflected in the clear separation of means for the categories of mature and immature (76 and 46% oil concentration, respectively).

The assessed moisture concentrations (population 2, kernels taken through the drying process) varied between 1.4 and 4.2%.  These moisture concentrations are as expected for Australian macadamia production and processing (Himstedt 2002).

On visual assessment, membership of the mould, decomposition, insect damaged, and discolouration groups appeared overlapped.  Germinating kernels were distinguished by a brown, orange or green coloration on the micropyle.  Insect damaged kernels exhibited a small discoloured and sometimes mouldy area centred on a wound site.  Rancid kernels were not visually distinguishable from sound kernels.

## Oil MPLS

Calibration statistics for oil concentration models for intact kernels were acceptable on all instruments based on SDR (SDR > 3.0, RMSECV similar to reference method error).  However, the calibrations developed on the InGaAs instrument (SDR 9.3) were significantly (based on Fearn's criterion) superior to that from the Si instrument (SDR 6.5) and from the Si-PbS reflectance system (SDR 3.1) (2002 population, Table 4, Fig. 1).  The population assessed in the 2004 season for oil concentration supported a model with higher RMSECV values than the population assessed in 2002.  However, the 2004 based model still had a SDR value of greater than 3.0 (Table 5), using both the Si and InGaAs instruments.
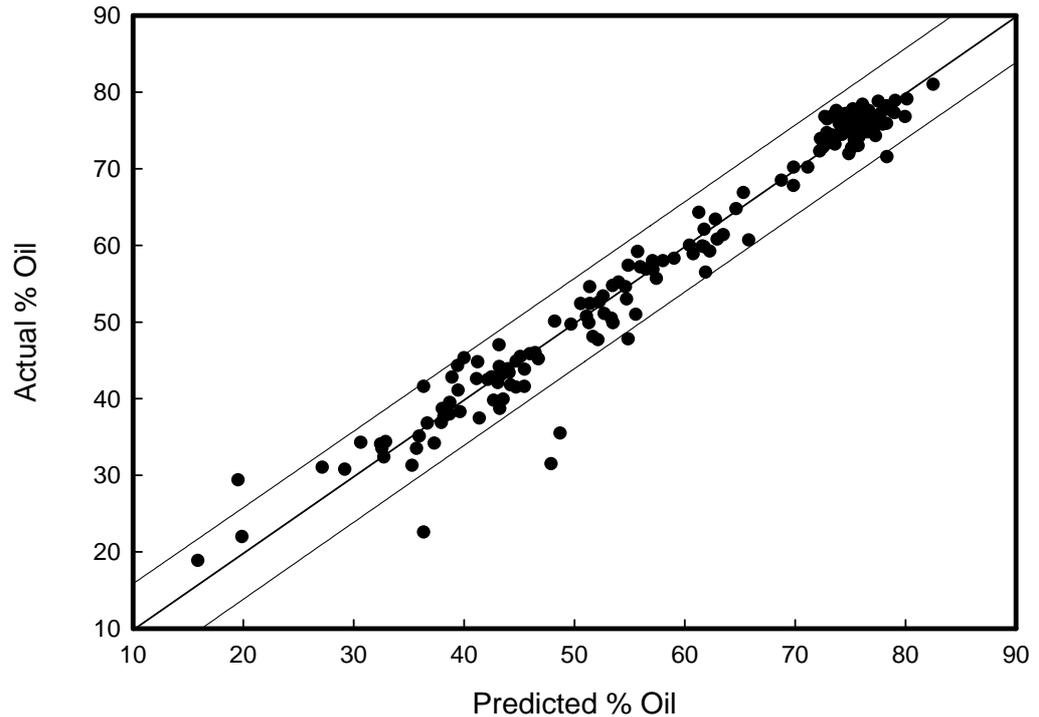
**Figure 1.** **Regression between actual oil concentration and values predicted using the MPLS model for intact macadamia kernels (2002 population), based on spectral data from the Si (Zeiss MMS1) operated in transmission geometry and over the wavelength range 700 – 1,100 nm. Regression performed on absorbance data, pre-treated with SNV and detrend. Dotted lines represent 95% confidence interval. Calibration statistics (from Table 4): $n = 199$, SD = 15.4%, $R_c^2 = 0.98$, RMSECV = 2.4%).**

The better performance of the InGaAs unit, relative to the Si unit, for the assessment of oil concentration was attributed to detection of wavelengths relevant to lower order overtones of the oil $CH_2$ bond. This advantage presumably outweighs the disadvantage of the more limited penetration of wavelengths above 1,000 nm and may indicate that the kernel is relatively homogenous in relation to oil concentration.

The Si-PbS instrument collected data over the full 400 – 2,500 nm range, however reflectance optics were used. The transmission optical path employed with the photo-diode instruments was presumably advantageous for this application.

154

Incorporating the height of the individual kernels in a PLS2 regression for oil prediction with the Si instrument showed no improvement in the calibration statistics for the 2004 population (PLS1 with $R_c^2$ of 0.94 and RMSEC 4.9% compared to the PLS2 result of $R_c^2$ 0.91 and RMSEC of 6.2% oil concentration, data not shown). This can be explained by the fact that the height of the kernel is already included in the PLS1 regression since a wide range of kernel heights (mature and immature) are in the population.

For spectral data collected with the Si photodiode instrument (used in transmission mode), the best model performance for the attribute of oil concentration was achieved using absorbance spectra, treated with detrend (data not shown). Typical statistics for calibration models on oil were: $R_v^2 = 0.98$; RMSECV = 2.4% (Fig.1). (McGlone and Kawano 1998) suggested that a SDR (SDR = SD/RMSECV) of > 3 is adequate to support sorting into 3 classes. The SDR of > 6 reported here is indicative that the calibration would support a useful sorting function.

Calibration loadings were consistent with absorbance due to the third overtones

of CH₂ stretching (e.g. heavy weightings around 930 nm) (data not shown).

**Table 5.** **Calibration statistics for population 2 (2004) and prediction of population 2 (2004) with the population 1 (2002 model) for percentage oil. Calibration models were developed for the Si and InGaAs instruments using a first derivative of absorbance data and scatter correction routine of standard normal variance and detrend. Calibration result is reported in terms of number of spectra, number of outliers removed, standard deviation of the calibration population, $R^2$ ($R_c^2$), RMSECV and SDR (SD/RMSECV). Validation results are reported in terms of $R_v^2$, RMSEP(C) and *bias*.**

| Instrument | Math | # PC | *n* (# outliers) | SD | $R_c^2$ | RMSECV | SDR | $R_v^2$ | RMSEP(C) | *Bias* |
|---|---|---|---|---|---|---|---|---|---|---|
| **Si** | 1:4:1 Detrend | 3 | 103 (12) | 14.21 | 0.92 | 4.2 | 3.37 | 0.85 | 5.8 | 94.9 |
| **InGaAs** | 1:4:1 SC | 3 | 110 (13) | 13.50 | 0.92 | 4.0 | 3.34 | 0.79 | 7.3 | 7.8 |

## Moisture MPLS

Calibration models for moisture concentration of intact kernels were not as

reliable as those for oil concentration. Best results were obtained with the Si-PbS

instrument ($R_v^2$ = 0.79, RMSECV 0.11) (Table 4, Fig.2).
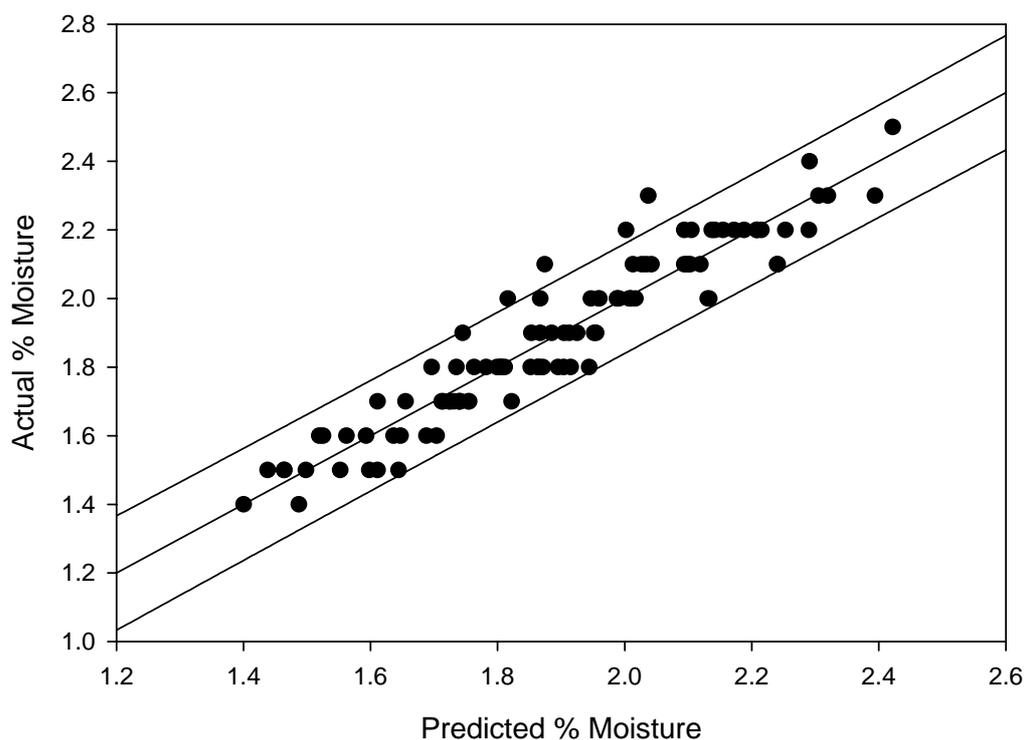
**Figure 2.** **MPLS regression (calibration) for actual and predicted moisture concentration in intact macadamia kernels, based on spectral data from the Si-PbS (Foss NIRSystems 6500) operated in reflectance geometry and over the wavelength range 700 – 1,900nm. Regression performed on first derivative absorbance data, without scatter correction. Lines represent a 95% confidence interval. Calibration statistics (from Table 4):** $n = 105$**, SD = 0.25,** $R_c^2 = 0.89$**, RMSECV = 0.11%).**

Calibration model performance based on spectra acquired using the Si ($R_v^2 = 0.59$, RMSECV 0.16) (Table 4), were superior to that obtained with the InGaAs instrument. The better performance of the Si, relative to InGaAs, photodiode array based unit is consistent with a relatively non-homogenous distribution of moisture concentration in the kernel, with the better relative penetration of SW-NIR (700 – 1,100 nm) radiation, supporting a stronger calibration model. Conversely, best results were obtained with the Si-PbS instrument, which was operated in a reflectance mode. Reflectance optics would optically sample only the surface layers of the kernel. These results are not consistent, and we anticipate further work to resolve this issue.

It was necessary to limit the number of principal components to 3 to avoid over-fitting of data with the ground samples.  WINISI (1998) suggest 1 PC per 10 samples to avoid over-fitting.  Within this constraint, the Si-PbS unit ($R_v^2 = 0.97$, RMSECV 0.06% moisture) (Table 4), operated with the spinning cup reflectance module, supported better model statistics than that obtained from intact kernels with the same unit, operated with a remote reflectance probe (Table 4).  This result is consistent with a level of non-homogeneity of moisture within intact kernels.  Grinding reduces this variation, and the spinning cup module allows scanning of a large proportion of the sample.   In contrast the predictive models developed for the Si and InGaAs units were relatively poor (e.g. $R_v^2 < 0.5$).  The transmission optics employed for the Si and InGaAs units with ground samples allowed spectral assessment of only a small proportion of the sample.  A change to this arrangement is recommended in future work.

Calibration coefficients were weighted around 1,360, 740 and 840 nm (data not shown), consistent with absorbance due to overtones of OH stretching and combination.

## Discriminant Analysis

Sound kernels and those with brown centres and rancidity could be discriminated from each other using principal component analysis of the spectral data obtained in transmission mode (Fig. 3).

Part kernels affected by insect damage, discolouration, mould growth, germination, and decomposition could be discriminated from sound kernels (Fig. 4). However, discrimination among these defect categories was not distinct and could not be validated on an independent population (data not shown).
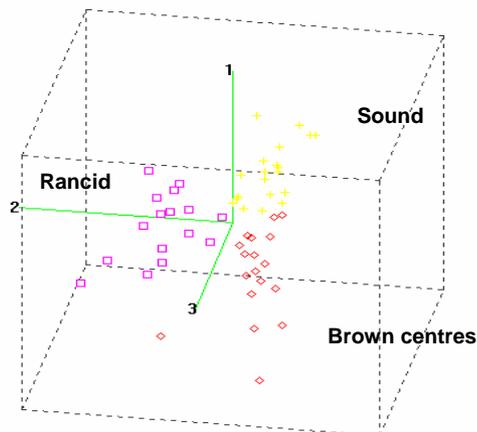
**Figure 3.** Discriminant analysis (PCA) of whole macadamia kernels using the InGaAs (Zeiss MMSNIR) in transmission mode. Plot uses the first 3 principal components.
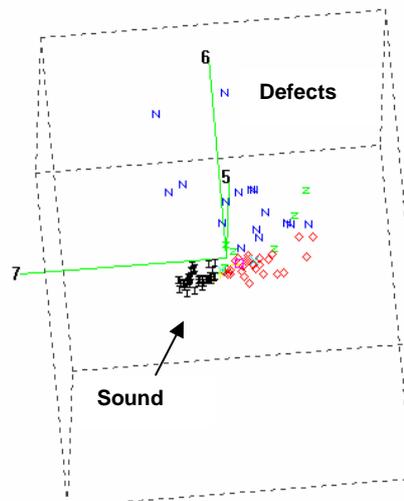


**Figure 4.** Discriminant analysis (PCA) of macadamia kernels using the Si (Zeiss MMS1) in interactance mode. Plot of the last 3 (fifth, sixth, seventh) principal components.

This was probably due to the overlapping of defect categories (e.g. decomposed also exhibited discolouration and possibly mould growth). This study was also based on relatively low sample numbers, and further replication is recommended.

**Oil model robustness**

A calibration model developed from season 2002 kernels performed poorly in prediction on kernels acquired in season 2004, (RMSEP of 94.9 and 10.7% oil concentration for the Si and InGaAs instruments, respectively) but the error was predominantly due to the large *bias* (94.9 and 7.8 % oil concentration for the Si and InGaAs instruments, respectively; Table 5).  Updating of the calibration developed with the 2002 kernels with fewer than 10 samples chosen randomly (Guthrie *et al*. 2005b) from the 2004 kernels resulted in both the RMSEP and *bias* being similar to the results of the 2004 calibration model (Fig. 5).
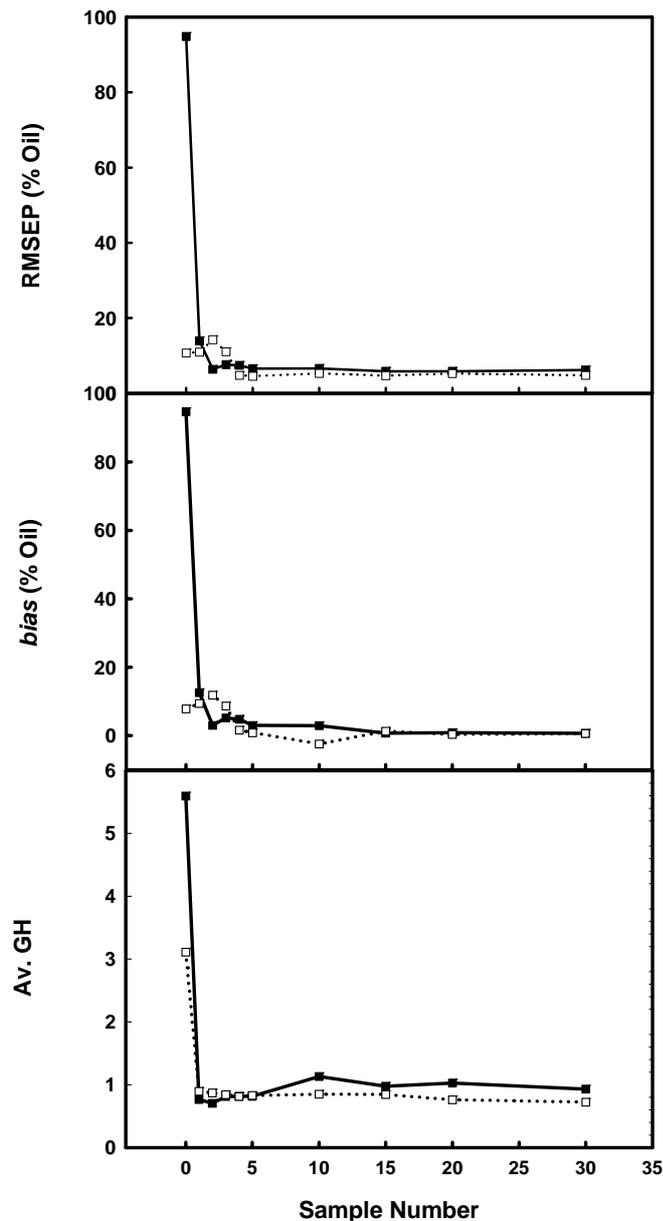
**Figure 5.** Prediction statistics (root mean square error of prediction (RMSEP) and *bias*) for modified partial least square (MPLS) prediction models for oil concentration of intact macadamia kernels, using an independent (of the calibration population) validation population with 2 instruments (Si detector instrument, closed square; InGaAs detector instrument, open square). The average Mahalanobis Global H statistic (GH) of samples in the validation population was calculated using calibration model scores. The initial calibration population consisted of the population assessed in 2002. Samples were selected randomly from the prediction population (population assessed in 2004) for addition to the calibration population.

# CONCLUSION

The low cost Si photodiode array instrument achieved a RMSECV of 2.4 and 4.2% on oil concentration of intact kernels over 2 separate seasons, respectively. However, contrary to the expectation that prediction of oil concentration would be more robust than prediction of sugar, prediction of the second season's kernels with a calibration developed in the first year was poor. This low moisture product also benefited from the model updating procedure outlined in Guthrie *et al*. (2005b). Model updating involving addition of 10 samples or less supported prediction results which were comparable with the second season's prediction on itself.

The Si photodiode instrument achieved a RMSECV of 0.2% on moisture concentration of ground kernels, and also allowed discrimination between kernels affected with a number of defects.

This study has shown encouraging results indicating that this technology should prove useful to the processing industry as an assessment tool of some quality attributes both in the processor laboratory and in-line. Further work should consider optical configurations to optimise sampling of the product, and also the robustness of the calibration models across different growing districts, cultivars and times of harvest.

# 7

# CONCLUSIONS AND FUTURE RESEARCH

The potential of NIR spectroscopy to assess internal quality attributes, notably TSS, of intact fruit is well established in the literature. However, the practical application of this technology for the sorting of intact fruit in commercial packing sheds will be limited by the robustness of the prediction models. Many of the literature reports of model development have been restricted to calibration exercises, and have therefore been limited because of the lack of true validation (i.e. validation of the calibration model on a population truly independent of the calibration population). Robustness should be defined in terms of the ability of the model to predict accurately across populations differing in time, cultivar/variety and growing district.

This thesis has characterised hardware options, scanning position (as determined by commodity attribute distribution) and chemometric approaches (with regard to optimum spectral window, data pre-treatment and regression techniques) for the application of TSS assessment of intact fruit, from the viewpoint of model robustness rather than straight calibration model performance. The utility of relatively inexpensive SW-NIR Si PDA spectrometers for this application was confirmed. Partial least squares regression models were demonstrated to be superior to either MLR or 'local' PLS models as used in the WINISI chemometric software

package. It is recommended that the optical configuration employed be varied between commodities, with thin skinned produce of relatively homogenous composition requiring a different arrangement to that for thick skinned produce, or produce with heterogenous composition.

However, while the hardware and chemometric approaches were optimised, model robustness across new populations remains a performance issue. Robustness of prediction models on independent populations was most impacted by growing season, relative to locality of production, variety or time of harvest.

A simple model updating procedure was proposed to accommodate the variations introduced by the new populations. It was recommended that model updating by the addition of between 10 and 30 samples chosen randomly from the new population be undertaken when the average GH of the new set (calculated using the scores and loadings of the model developed on the calibration population) exceeds 1.0. Further work is required to establish the number of added samples required in relation to population size, both in relation to new populations of fruit and sample temperature. The NH statistic as advocated by WINISI can be used to winnow the calibration population, maintaining a relatively compact population size, without loss of prediction accuracy (e.g. see Appendix C).

The model updating procedure used was, however, relatively labour intensive. To support commercial application it is possible that the model updating could be undertaken automatically by software. In such software, if the average GH of the predictions were to exceed 1.0, 20 samples for which spectra and reference values (e.g. %TSS) were available, randomly selected from the available population, would be merged with the calibration population and the PLS calibration model regenerated. Matlab and PLS Toolbox (Eigenvector) are appropriate packages for such a task.

Thus a hardware (Si photodiode array/halogen lamp) and statistical (PLS) solution to the application of fruit grading has been characterised.  Significant performance improvements, however, are likely to come from changes to different types of hardware and statistical solutions, rather than incremental improvements to the existing solution.  For example, Butz *et al*. (2001) advocated the use of diode lasers operating in the SW-NIR as illumination sources.  They reported on the use of multiple 50 milliwatt laser diodes operating at SW-NIR wavelengths for the assessment of TSS and firmness of apples, although method robustness was not considered.  The Japanese company Sumitomo commercialised such technology for fruit sorting in the 1990s', claiming reduced calibration transfer problems and increased model robustness.  However, this unit is apparently no longer commercially available.  It is likely that power stability/temperature control/laser output stability and safety issues presented significant technical hurdles.

Fourier transform near infra-red spectroscopy (FTNIR) is another technology that holds promise, with claims that model transfer between instruments is straightforward.  This feature is ascribed to high wavelength resolution, and the use of a single detector.  However, model transfer between instruments is a different issue to model robustness across new populations, and this issue deserves study for the application of the assessment of internal quality attributes of fruit.  Further, the technology is based on a Michelson interferometer, involving use of a moving part and a discrete time interval in wavelength scanning.  Thus the applicability of the technology to the sorting of fruit on a pack-line (at 10 pieces per second, in a vibrating environment) should also be considered.

Raman spectroscopy is another technique, generally considered as complimentary to FTIR/NIR spectroscopy as it probes fundamental molecular

vibrational transitions rather than overtones and combination bands. The potential of this technique to assess low levels of specific compounds in fruit should be assessed (e.g. pesticide or fungicide residues in dried, ground samples).

Partial least squares regression became available in the 1970s, and has been adopted as the chemometric method of choice by the NIR community, supplanting MLR. Other methods have been proposed, and may eventually supplant PLS for the application of fruit sorting. For non-linear applications, neural network and genetic algorithms have proven superior to PLS, but for the application of fruit sorting on TSS, these techniques yield results generally equivalent to, or inferior to, PLS. Least squares support vector machine (SVM) is another non-linear approach which has recently been promoted for use in chemometrics. Cogdill and Dardenne (2004) report that for the prediction of apple sucrose concentration from SW-NIR spectra, SVM outperformed neural networks, local PLS and MPLS (RMSEP's of 0.32, 0.33, 0.34 and 0.37% TSS, respectively). Unfortunately model robustness was not considered, although it was noted that the power of this method to model non-linear functions could easily result in model over-fitting, and consequent poor predictive performance. This technique deserves further consideration for this application.

It is also likely that NIRS will be used in the future for prediction of more 'complex' attributes than average TSS or DM concentration. Examples include the detection of insect presence or damage (see Appendix D), internal discolouration (Chapter 6) or predicted shelf life of foods. Near infra-red imaging also offers potential for in-line food processing and in production agriculture.

Finally, the adoption of a new technology is constrained by many factors. At a superficial level, one might expect the ability to sort fruit for eating quality (TSS level) at an affordable price would be rapidly adopted. However, the implementation

of such a capability can be disruptive to existing procedures, constraining uptake. Parallel research is therefore required into breeding (for varieties producing sweeter fruit), agronomic technique (learning to grow sweeter fruit), consumer preferences, human health related issues, et cetera.

# REFERENCES

Aoki H, Kouno Y, Matumoto K, Mizuno T, Maeda H (1996) Nondestructive determination of sugar content in melon and watermelon using near infrared (NIR) transmittance. In 'International Conference on Tropical Fruits'. Kuala Lumpur, Malaysia. (Eds. S Vijaysegaran, M Pauziah, MS Mohamed and S Ahmad) p. 35. (Malaysian agricultural Research and Development Institute.)

Barnes RJ, Dhanoa MS, Lister SJ (1989) Standard normal variate transformation and de-trending of near infrared diffuse reflectance spectra. *Applied Spectroscopy* **43**, 772-777.

Bellon V, Vigneau JL, Leclercq M (1993) Feasibility and performances of a new, multiplexed, fast and low-cost fiber-optic NIR spectrometer for the on-line measurement of sugar in fruits. *Applied Spectroscopy* **47**, 1079-1083.

Bellon-Maurel V, Vigneau JL (1995) NIR fast spectrometer for fruit internal quality assessment. In Harvest and postharvest technologies for fresh fruits and vegetables. Guanajuato, Mexico pp.471-476.

Birth SG, Dull GG, Leffler RG (1990) Non-destructive determination of the solids content of horticultural products. *Optics in Agriculture* **1379**, 10-15.

Bouveresse E, Massart DL (1996) Improvement of the piecewise direct standardisation procedure for the transfer of NIR spectra for multivariate calibration. *Chemometrics & Intelligent Laboratory Systems* **32**, 201-213.

Butz P, Merkel C, Lindauer R, Tauscher B (2001) Application of a multi-wavelength near infrared diode laser array for non-destructive food analysis. In '10th International Conference on Near Infrared Spectroscopy'. Kyonjgu, Korea. (Eds. AMC Davies and RK Cho) pp. 151-154. (NIR Publications).

Cavaletto CA, Cruz Dela E, Ross E, Yamamoto HY (1966) Factors Affecting Macadamia Nut Stability. 1. Raw Kernels. *Food Technology* **20**, 1084-1087,

Chambers J, Ridgway C, van Wyk CB, Baker CW, Barnes RJ (1994) 'The potential of near-infrared spectroscopy for the rapid detection of pests in stored grain.' In Project Report No. 92, Central Science Laboratory (MAFF), London Road, Slough, Berkshire SL3 7HJ.

Cho S, Kim J, Rhee C (1998) Determination of rancidity of soybean oil by near infrared spectroscopy. *Journal of Near Infrared Spectroscopy* **6**, 349-354.

Cogdill RP, Dardenne P (2004). Least-squares support vector machines for chemometrics: an introduction and evaluation. *Journal of Near Infrared Spectroscopy* **12**, 93-100.

Galactic Industries Corporation (1999) 'PLSplus/IQ' chemometric software manual (Galactic Industries Corporation: Salem, NH, USA).

Crisosto CH, Day KR, Johnson RS, Garner D (2000) Influence of in-season foliar calcium sprays on fruit quality and surface discoloration incidence of peaches and nectarines. *Journal of American Pomological Society* **54**, 118-122.

de Noord OE (1994) Multivariate calibration standardization. *Chemometrics and Intelligent Laboratory Systems* **25**, 85-97.

Dull GG, Birth GS, Leffler RG (1989) Exiting energy distribution in honeydew melon irradiated with a near infrared beam. *Journal of Food Quality* **12**, 377-381.

Dull GG, Leffler RG (1992) Nondestructive measurement of soluble solids in fruits having a rind or skin. United States Patent # 5,089,701

Dull GG, Leffler RG, Birth GG, Smittle DA (1990) Instrument for nondestructive measurement of soluble solids in honeydew melons. In '1990 International Winter Meeting'. Chicago, Illinois. pp. 735-737. (The American Society of Agricultural Engineers).

Fearn T (1992) Flat or natural? A note on the choice of calibration samples. In 'Near-Infrared Spectroscopy: Bridging the Gap between Data Analysis and NIR Applications'. (Eds. KI Hildrum, T Isaksson, T Naes, A Tandberg) pp. 61-66. (Ellis Harwood, New York).

Fearn T (1996) Comparing standard deviations. *NIR News* **7**, 5-6.

Flinn PC, Murray I (1987) Potential of near infrared reflectance spectroscopy (NIR) for evaluation of herbage quality in southern Australia. In Temperate pastures - their production, use and management. Australian Wool Corporation/CSIRO. pp. 426-428.

Fox GP, Sulman M, Osborne B, Inkerman A (2000) Evaluation of barley colour pigment by NIR. In '9th Australian Near Infra-red Spectroscopy Conference'. Horsham, Victoria, Australia.

García-Ciudad A, Ruano A, Becerro F, Zabalgogeazcoa I, Vázquez de Aldana BR, B. G-C (1999) Assessment of the potential of NIR spectroscopy for the estimation of nitrogen content in grasses from semiarid grasslands. *Animal Feed Science and Technology* **77**, 91-98.

Gemperline PJ, Webber LD, Cox FO (1989) Raw materials testing using soft independent modelling of class analogy analysis of near-infrared reflectance spectra. *Analytical Chemistry* **61**, 138-144.

Golic M, Walsh KB, Lawson P (2003) Short-Wavelength Near-Infrared Spectra of Sucrose, Glucose, and Fructose with Respect to Sugar Concentration and Temperature. *Applied Spectroscopy* **57**, 139-145.

Greensill CV, Walsh KB (2000) Optimisation of instrumentation precision and wavelength resolution for the performance of NIR calibrations of sucrose in a water-cellulose matrix. *Applied Spectroscopy*, **41**, 426-430.

Greensill CV, Walsh KB (2000) A remote acceptance probe and illumination configuration for spectral assessment of internal attributes of intact fruit. *Measurement Science and Technology* **11**, 1674-1684.

Greensill CV, Walsh KB (2002) Calibration transfer between miniature photodiode array-based spectrometers in the near infrared assessment of mandarin soluble solids content. *Journal of Near Infrared Spectroscopy* **10**, 27-35.

Greensill CV, Wolfs PJ, Spiegelman C, Walsh K (2001) Calibration transfer between PDA-based spectrometers in the NIR assessment of melon soluble solids content. *Journal of Applied Spectroscopy,* **55**, 647-653.

Guthrie J, Walsh K (1997a) NIR analysis of soluble solids in pineapple and dry matter in mango. In 'Proceedings from the Sensors for nondestructive Testing International Conference and Tour Holiday Inn International Drive Resort *Not in the official proceedings*'. Orlando, Florida.

Guthrie JA, Walsh KB (1997b) Non-invasive assessment of pineapple and mango fruit quality using near infra-red spectroscopy. *Australian Journal of Experimental Agriculture* **37**, 253-263.

Guthrie JA, Walsh KB (1999) Influence of environmental and instrumental variables on the non-invasive prediction of Brix in pineapple using near infra-red spectroscopy. *Australian Journal of Experimental Agriculture* **39**, 73-80.

Guthrie JA, Walsh KB (2001) Assessing and enhancing near infrared calibration robustness for soluble solids content in mandarin fruit. In '10th International Conference on Near Infrared Spectroscopy'. Kyonjgu, Korea. (Eds. AMC Davies and RK Cho) pp. 151-154. (NIR Publications).

Guthrie JA, Walsh KB, Reid DJ, Liebenberg CJ (2005a) Assessment of internal quality attributes of mandarin fruit. 1. NIR calibration model development. *Australian Journal of Agricultural Research* **56**, 405-416.

Guthrie JA, Reid DJ, Walsh KB (2005b) Assessment of internal quality attributes of mandarin fruit. 2. NIR calibration model robustness. *Australian Journal of Agricultural Research* **56**, 417-426.

Guthrie JA, Wedding B, Walsh K (1998) Robustness of NIR calibrations for soluble solids in intact melon and pineapple. *Journal of Near Infrared Spectroscopy* **6**, 259-265.

Ha J, Koo M, Ok H (1998) Determination of the constituents of sesame oil by near infrared spectroscopy. *Journal of Near Infrared Spectroscopy* **6**, 371-373.

Himstedt SR (2002) Oil content and other components as indicators of quality and shelf-life of *Macadamia integrifolia*. Ph D Dissertation, University of Queensland.

Hruschka W (1987) Data analysis: wavelength selection methods. In 'Near-Infrared Technology in the Agricultural and Food Industries'. (Ed. P Wa, K Norris) pp. 35-56. (American Association of Cereal Chemists, Inc.: Minnesota 55121, USA).

Ito H, Ippoushi K, K. A, Higashio H (1999) Non-destructive estimation of soluble solids in intact melons by non-contact mode with a fibre optic probe. In '9th International Conference of Near Infrared Spectroscopy'. Verona Italy. (Eds. AMC Davies and R Giangiacomo) pp. 859-862. (NIR Publications).

Jaenisch HM, Niedzwiecki AJ, Cernosek JD, Johnson RB, Seeley JS, Dull GG, Leffler RG (1990) Instrumentation to measure the near-IR spectrum of small fruits. *Optics in Agriculture* **1379**, 162-167.

Kawano S (1994a) Non-destructive NIR quality evaluation of fruits and vegetables in Japan. *NIR News* **5**, 10-12.

Kawano S (1994b) Present condition of nondestructive quality evaluation of fruits and vegetables in Japan. *Japan Agricultural Research Quarterly* **28**, 212-216.

Kawano S, Abe H, Iwamoto M (1995) Development of a calibration equation with temperature compensation for determining the Brix value in intact peaches. *Journal of Near Infrared Spectroscopy* **3**, 211-218.

Kawano S, Fujiwara T, Iwamoto M (1993) Non-destructive determination of sugar content in Satsuma Mandarin using near infrared (NIR) transmittance. *Journal of the Japanese Society for Horticultural Science* **62**, 465-470.

Kawano S, Watanabe H, Iwamoto M (1989) Measurement of sugar contents in intact peaches by NIRS. In '2nd International NIRS Conference'. Tsukuba, Japan. (Eds. Iwamoto M, Kawano S) pp. 343-352.

Kim J, Mowat A, Poole P, Kasabov N (2000) Linear and non-linear pattern recognition models for classification of fruit from visible-near infrared spectra. *Chemometrics and Intelligent Laboratory Systems* **51**, 201-216.

Kojima T, Inoue Y, Tanaka M (1994) Measurement of Brix value in various developing stages of Japanese pear by NIR spectroscopy. *Bulletin of the Faculty of Agriculture, Saga University.* **77**, 1-10.

Korcak RF, Walker V, Norris KH (1990) Measurement of fruit tree total leaf nitrogen by near-infrared reflectance spectroscopy. In 'International Symposium on Diagnosis of Nutritional Status of Deciduous Fruit Orchards'. pp. 241-247. (Acta Horticulturae).

Lammertyn J, Nicolai B, Ooms K, De Smedt V, De Baerdemaeker J (1998) Non-destructive measurement of acidity, soluble solids, and firmness of Jonagold apples using NIR-spectroscopy. *Transactions of the American Society of Agricultural Engineers* **41**, 1089-1094.

Lammertyn J, Peirs A, De Baerdemaeker J, Nicolai B (2000) Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment. *Postharvest Biology and Technology* **18**, 121-132.

Lester G, Shellie KC (1992) Postharvest sensory and physicochemical attributes of honey dew melon fruits. *HortScience* **27**, 1012-1014.

Mahalanobis PC (1936) On the generalized distance in statistics. *Proceedings of the National Institute of Science of India* **2**, 49.

Mark H (1986) Normalized distances for qualitative near-infrared reflectance analysis. *Analytical Chemistry* **58**, 379-384.

Mark HL, Tunnell D (1985) Qualitative near-infrared reflectance analysis using Mahalanobis distances. *Analytical Chemistry* **57**, 1449-1456.

McGlone VA, Fraser DG, Jordan RB, Kunnemeyer R (2003) Internal quality assessment of mandarin fruit by vis/NIR spectroscopy. *Journal of Near Infrared Spectroscopy* **11**, 323-332.

McGlone VA, Kawano S (1998) Firmness, dry-matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy. *Postharvest Biology and Technology* **13**, 131-141.

Miyamoto K, Kawauchi M, Fukuda T (1998) Classification of high acid fruits by partial least squares using the near infrared transmittance spectra of intact satsuma mandarins. *Journal of Near Infrared Spectroscopy* **6**, 267-271.

Miyamoto K, Kitano Y (1995) Non-destructive determination of sugar content in satsuma mandarin fruit by near infrared transmittance spectroscopy. *Journal of Near Infrared Spectroscopy* **3**, 227-237.

Mowat AD, Poole PR (1997) Use of visible-near infrared diffuse reflectance spectroscopy to discriminate between kiwifruit with properties altered by preharvest treatments. *Journal of Near Infrared Spectroscopy* **5**, 113-122.

Mutton LL, Cullis BR, Blakeney AB (1981) The objective definition of eating quality in rockmelons (*Cucumis melo*). *Journal of the Science of Food and Agriculture* **32**, 385-390.

Norris KH (1983) Using gap derivatives as pre-processing for quantitative models - Food Research and Data Analysis. In Proceedings of the 1982 IUFST Symposium. Oslo, Norway. (Ed. H Martens) pp 46-47. (Applied Science Publishing).

Onda T, Tsuji M, Komiyama Y (1994) Possibility of nondestructive determination of sugar content, acidity and hardness of plum fruit by near infrared spectroscopy. *Journal of the Japanese Society for Food Science & Technology* **41**, 908-912.

Oriel (1994) 'Oriel Catalogue' Volumes II and III (Eds. EG Arthurs, D Zbigniew) (Oriel Stratford, Connecticut, USA).

Osborne BG, Fearn T, Hindle PH (1993) Theory of near infrared spectrophotometry. In 'Practical NIR Spectroscopy with Applications in Food and Beverage Analysis'. (Ed. D Browning) pp. 13-35. (Longman Singapore Publishers Ltd.).

Osborne SD, Jordan RB, Kunnemeyer R, (1996) Using near infrared (NIR) light to estimate the soluble solids and dry matter content of kiwifruit. In 'Proceedings of International PostHarvest Science Conference'. Taupo, New Zealand. pp.187-191.

Ou AS, Lin S, Lin T, Wu S, Tiarn M (1997) Studies on the determination of quality-related constituents in Ponkan Mandarin by near infrared spectroscopy. *Journal of the Chinese Agricultural Chemical Society* **35**, 462-474.

Peiris KHS, Dull GG, Leffler RG, Burns JK, Thai CN, Kays SJ (1998a) Non-destructive detection of section drying, an internal disorder in Tangerine. *HortScience* **33**, 310-312.

Peiris KHS, Dull GG, Leffler RG, Kays SE (1999a) Rapid, nondestructive method for determination of processed soluble solids in intact unprocessed tomato fruit using near infrared spectrometry. *Acta Horticulturae* **487**, 413-417.

Peiris KHS, Dull GG, Leffler RG, Kays SJ (1998b) Near-infrared spectroscopic method for non-destructive determination of soluble solids content of peaches. *Journal of the American Society for Horticultural Science* **123**, 898-905.

Peiris KHS, Dull GG, Leffler RG, Kays SJ (1999b) Spatial variability of soluble solids or dry-matter content within individual fruits, bulbs, or tubers: implications for the development and use of NIR spectrometric techniques. *Hortscience* **34**, 114-118.

Peirs A, Tirry J, Verlinden B, Darius P, Nicolai BM (2003) Effect of biological variability on the robustness of NIR models for soluble solids content of apples. *Postharvest Biology and Technology* **28**, 269-280.

Ripperton JC, Moltzau RH, Edwards DW (1938) Methods for evaluating the macadamia nut for commercial use and the variation occurring among seedlings plantings in Hawaii. *Hawaii Agricultural Experimental Station's Bulletin* **79**.

Roger J-M, Chauchard F, Bellon-Maurel V (2003) EPO and PLS external parameter orthogonalisation of PLS - application to temperature independent measurement of sugar content of intact fruits. *Chemometric and Intelligent Laboratory Systems* **66**, 191-204.

Sanchez H, N., Lurol S, Roger JM, Bellon-Maurel V (2003) Robustness of models based on NIR spectra for sugar content prediction in apples. *Journal of Near Infrared Spectroscopy* **11**, 97-107.

Schmilovitch Z, Mizrach A, Hoffman A, Egozi H, Fuchs Y (2000) Determination of mango physiological indices by near-infrared spectrometry. *Postharvest Biology and Technology* **19**, 245-252.

Shenk JS, Westerhaus MO (1991) Population definition, sample selection and calibration procedures for near infrared reflectance spectroscopy. *Crop Science*, **31** 469-474.

Shenk JS, Westerhaus MO (1993) 'Analysis of agriculture and food products by near infrared reflectance spectroscopy.' (Infrasoft International, LLC, Infrasoft International, Port Matilda, PA).

Shenk JS, Westerhaus MO (1997) Investigation of a LOCAL calibration procedure for near infrared instruments. *Journal of Near Infrared Spectroscopy* **5**, 223-232.

Shenk JS, Workman JJ, Westerhaus MO (1992) Application of NIR spectroscopy to agricultural products. In 'Handbook of Near-Infrared Analysis'. (Eds. DA Burns and EW Ciurczak) pp. 383-431. (Marcel Dekker, Inc.: New York).

Shiina T, Ijiri T, Matsuda I, Sato T, Kawano S, Ohoshiro N (1993) Determination of Brix value and acidity in pineapple fruits by near infrared spectroscopy. *Acta Horticulturae* **334**, 261-272.

Slaughter DC, Barrett D, Boersig M (1996) Nondestructive determination of soluble solids in tomatoes using near infrared spectroscopy. *Food Science* **61**, 695-697.

Sohn M-R, Park K-S, Cho SI (2000) Near infrared reflectance spectroscopy for non-invasive measuring of internal quality of apple fruit. *Near Infrared Analysis* **1**, 27-30.

Sugiyama J (1999) Visualization of sugar content in the flesh of a melon by near-infrared imaging. *Journal of Agricultural and Food Chemistry* **47**, 2715-2718.

Swierenga H, Haanstra WG, de Weijer AP, Buydens LMC (1998) Comparison of two different approaches toward model transferability in NIR spectroscopy. *Applied Spectroscopy* **52**, 7-16.

Thomas EV, Ge N (2000) Development of robust multivariate calibration models. *Technometrics* **42**, 168-177.

Tillmann P, Reinhardt TC, Paul C (2000) Networking of near infrared spectroscopy instruments for rapeseed analysis: a comparison of different procedures. *Journal of Near Infrared Spectroscopy* **8**, 103-107.

Tsuta M, Sugiyama J, Sagara Y (2002) Near-infrared imaging spectroscopy based on sugar absorption band for melons. *Journal of Agriculture and Food Chemistry* **50**, 48-52.

Ventura M, de Jager A, de Putter H, Roelofs FPMM (1998) Non-destructive determination of soluble solids in apple fruit by near infrared spectroscopy (NIRS). *Postharvest Biology and Technology* **14**, 21-27.

Walsh KB, Golic M, Greensill CV (2004) Sorting of fruit using near infrared spectroscopy: application to a range of fruit and vegetables for soluble solids and dry matter content. *Journal of Near Infrared Spectroscopy* **12**, 141-148.

Walsh KB, Greensill CV, Guthrie JA (1999) Development and use of an 'at-line' NIR instrument to evaluate robustness of melon Brix calibrations. In '9th International Conference Near Infrared Spectroscopy'. Verona Italy. (Eds. AMC Davies and R Giangiacomo) pp. 859-862. (NIR Publications).

Walsh KB, Guthrie JA, Burney J (2000) Application of commercially available, low-cost, miniaturised NIR spectrometers to the assessment of the sugar content of intact fruit. *Australian Journal of Plant Physiology* **27**, 1175-1186.

Wang Y, Veltkamp DJ, Kowalski BR (1991) Multivariate instrument standardisation. *Analytical Chemistry* **63**, 2750-2756.

Williams PC, Sobering DC (1993) Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds[a]. *Journal of Near Infrared Spectroscopy* 1, 25-32.

Wortel VAL, Hansen WG, Wiedemann SCC (2001) Optimising multivariate calibration by robustness criteria. *Journal of Near Infrared Spectroscopy* 9, 141-151.

Yamaguchi M, Hughes DL, Yabumoto K, Jennings WG (1977) Quality of cantaloupe muskmelons: variability and attributes. *Scientia Horticulturae* 6, 59-70.

Yermiyahu R, Nir S, Ben-Hayyim G, Kafkafi U, Kinraide TB (1997) Root elongation in saline solution related to calcium binding to root cell plasma membranes. *Plant and Soil* **191**, 67-76.

# CURRICULUM VITAE

| | |
|---|---|
| **NAME:** | JOHN AUSTIN GUTHRIE |
| **ADDRESS:** | Parkhurst, Rockhampton Q 4701 |
| **TITLE:** | Principal Food Technologist |
| **QUALIFICATIONS:** | QDAH (1969), D App Sc (Food Tech) (1978), |
| | B App Sc (Food Tech) (1980), M. App. Sc. (1997). |

**PROFESSIONAL EXPERIENCE**

**1969 – 1972                              DAIRY PROCESSING;**

- Market milk, butter, cheese, skim and whole milk powders.

**1972 – 1977                    Food Microbiology Laboratory;**

- Quality Control testing, demonstrating to undergraduate students.

**1977 – 1980                    Food Technologist;**

- officer in charge of pilot plant in tertiary teaching institution, involved in dairy processing, canning and retorting, meat processing/smallgoods, cryogenic freezing and spray drying/freeze drying.

**1980 - 1986                    Regional Dairy Technologist;**

- regional quality assurance with the dairy industry.
- regulatory work for the Queensland Milk Board including milk producers, processors and vendors.

**1986 - PRESENT                         REGIONAL FOOD TECHNOLOGIST;**

- extension, R & D and fee-for-service for the CQ regional food producing and food processing sector.  These food sectors have included eggs, poultry, seafood, meat, condiments, fruit and vegetables.

**PROFESSIONAL MEMBERSHIPS**

Member of the Australian Institute of Food Science and Technology

Member of the Dairy Industries Association of Australia (DIAA)

Member of the NIR users group of the RACI

Member of the Council of NIR Spectroscopy (USA)

**REFEREED PUBLICATIONS** (prior to PhD candidature):

Guthrie JA, Dunlop KJ, Saunders GA (1994) Use of Petrifilm$^{TM}$ 3M to assess coliform numbers on lamb carcasses. *Journal of Food Protection* **57**, 924-927.

O'Connor-Shaw RE, Guthrie JA, Dunlop KJ, Roberts R (1994) Coliforms in processed mango - significance and control. *International Journal of Food Microbiology* **25**, 51-61.

Guthrie JA, Walsh KB (1997) Non-invasive assessment of pineapple and mango fruit quality using near infra-red spectroscopy. *Australian Journal of Experimental Agriculture* **37**, 253-263.


**CONFERENCE PAPERS** (prior to PhD candidature):

Guthrie JA, Rutherford A (1991) 'Grain Sorghum: Developing Opportunities as a Human Food'. Australian Institute of Food Science and Technology (AIFST) Annual National Conference, Hobart, Tasmania, July.

Guthrie JA Walsh KB (1995) 'Development of a non-invasive objective test to measure the eating quality of pineapple and mango fruit by NIR spectroscopy'. Royal Australian Chemical Institute (RACI) NIR Users Conference, Cairns, Queensland, July.

Guthrie JA Walsh KB (1996) 'Non-invasive assessment of pineapple (*Ananas comosus* [L] Merrill) and mango (*Magnifera indica*) fruit quality by near infra-red spectroscopy'. Australian Institute of Food Science and Technology (AIFST) Annual National Conference, Gold Coast, Queensland, May.

Guthrie JA Walsh KB (1996) 'Whole Fruit Analysis by NIR', 1st Annual NIR Workshop for the Food and Packaging Industries, Victorian University of Technology (VUT), Melbourne, Victoria, 31 October - 1 November.

# INFLUENCE OF ENVIRONMENTAL AND INSTRUMENTAL VARIABLES ON THE NON-INVASIVE PREDICTION OF BRIX IN PINEAPPLE USING NEAR INFRARED SPECTROSCOPY[6]

## ABSTRACT

The Brix content of pineapple fruit can be non-invasively predicted from the second derivative of NIR reflectance spectra.  Correlations obtained using a NIRSystems 6500 spectrophotometer through MLR and MPLS analyses using a post-dispersive configuration were comparable to that from a pre-dispersive configuration in terms of accuracy (e.g. coefficient of determination, $R^2$, 0.73; standard error of cross validation, SECV, 1.01 °Brix).  The effective depth of sample assessed was slightly greater using the post-dispersive technique (ca. 20 mm for pineapple fruit), as expected in relation to the higher incident light intensity, relative to the pre-dispersive configuration.  The effect of such environmental variables as temperature, humidity and ambient light, and instrumental variables such as the number of scans averaged to form a spectrum, was considered with respect to the accuracy and precision of the measurement of absorbance at 876 nm, as a key term in the calibration for Brix, and predicted Brix.

---

[6] This appendix has been published in: *Australian Journal of Experimental Agriculture, 1999*, **39**, 73-80, under the title: 'Influence of environmental and instrumental variables on the non-invasive prediction of Brix in pineapple using near infrared spectroscopy'.  Authors were J.A. Guthrie and K.B. Walsh.

The application of post-dispersive near infrared technology to in-line assessment of intact fruit in a packing shed environment is discussed.

# INTRODUCTION

Near infrared spectroscopy (NIRS) is widely used for the identification of organic compounds, and has found increasing use for the non-invasive quantification of organic constituents within biological material. For example, the technique is widely used in the Australian grains, forage and oil seeds industry (e.g. assessment of protein and moisture contents). In Japan (Mitsui Mining and Smelting Corp., Omiya, and Maki Manufacturing Co., Hamamatsu) commercial in-line near infrared sensors are being used in packing sheds to assess the sweetness, ripeness and acidity of relatively smooth and thin skinned temperate fruits (citrus, apples, pears and peaches), at three pieces per second per lane (Kawano 1994). These units are not commercially available in Australia.

The thick, rough skin of pineapple fruit is expected to compromise the application of the technique to the assessment of Brix in this fruit. Shiina *et al*. (1993) and Guthrie and Walsh (1997, 1998) have reported the development of calibrations for Brix in pineapple fruit, under laboratory conditions. The error of prediction (SEP), however, was higher in these studies than those reported for thin skinned fruit.

The application of the NIRS technique to the prediction of pineapple Brix in a packhouse setting, as opposed to a laboratory setting, will involve additional sources of error. For example, the ability of a NIR spectroscopic system to predict the Brix of

fruit will be influenced by the effect of environmental parameters (e.g. light, temperature, humidity) of the sample on the NIR spectrum (a spectroscopic problem), by characteristics of the illumination system (e.g. effective depth of fruit from which information is acquired, use of 'white' or monochromatic incident light), by the effect of environmental parameters such as temperature and humidity on the instrument, and by performance characteristics of the spectrometer (e.g. signal : noise characteristic, reflecting mechanical considerations such as grating positioning, electronic considerations such as detector dark current and A/D conversion of data and optical considerations such as stray light within the spectrometer). The post-dispersive mode should offer advantages over the pre-dispersive mode with in-line applications because light intensity is lost in passage over the diffraction grating (for a concave holographic grating, as used in the NIRSystems 6500, maximum light transmission is typically 30% at the blaze wavelength). Thus the post dispersive system has the advantage of delivering a higher intensity of all wavelengths onto the sample, improving the effected depth of penetration into the sample. Also, external light around the sample will have less effect on the detector response. While NIRS is in commercial use in Japan, there is little published consideration of the effect of such variables for the prediction of fruit quality attributes. In this study, we consider the influence of these parameters on the assessment of pineapple fruit Brix, with a view to the application of NIRS to the sorting of pineapples for Brix content in an in-line setting.

# MATERIALS AND METHODS

## Plant material and constituent analysis

The pineapples (*Ananas comosus* [L] Merrill, var. Smooth Cayenne) used in the experiment were grown commercially on a Yeppoon, Central Queensland farm, and transported to the laboratory on the day of harvest. Spectra were acquired after sample temperature equilibration and within 3 days of harvest, in an air-conditioned laboratory at 22-24$^O$C. After scanning, a 60 mm diameter stainless steel corer was used to excise both skin and underlying flesh to a depth of 20 mm. The skin was subsequently removed and the flesh squeezed manually through a nylon cloth to extract the juice. The extracted juice was measured for Brix content using an Erma digital refractometer (accuracy +/- 0.2 °Brix).

## Near infrared spectrometry

A scanning monochrometer (Model 6500, NIRSystems, Silver Springs, MD, USA) driven by NSAS software (Version 3.3, NIRSystems) was used in 2 configurations. In the 'pre-dispersive' configuration, light from a 75 W tungsten halogen lamp passes through a slit and is dispersed by a moving grating. The dispersed light passes through order sorting filters, with the primary spectrum delivered via a 1.6 m fibre optic cable to a remote reflectance probe (Fig. 1). Thus monochromatic light is incident on the sample, and detectors in the probe (mounted at 45° to the sample surface) monitor the intensity of reflected light. In the 'post-dispersive' configuration, light from a 75 W lamp directly illuminates the sample, and light returning from the sample is delivered via a 1.6 m fibre optic to a spectrometer

(Fig. 1). Light passing a slit is dispersed by a moving grating, and is delivered through order sorting filters onto the detectors (positioned normal to the incident beam).

The protocols of Guthrie and Walsh (1997) for pre-dispersive work were adopted. Briefly, intact fruit were held in a light-proof poly vinyl chloride (PVC) box with a 60 mm diameter window. A laboratory jack held the fruit against the window so that the fruit skin was in direct contact with the quartz glass window of the NIRSystems remote reflectance probe. The same population of pineapple fruit was then scanned with the direct light NIRSystems 6500, with lamp and fibre optic positioned 7 cm above the intact fruit. In both configurations, spectra were obtained at the centre of the fruit's longest dimension, from each half of the fruit (i.e. 2 spectra per fruit), and reference scans were undertaken between each sample spectrum, using a white ceramic tile as the reference.
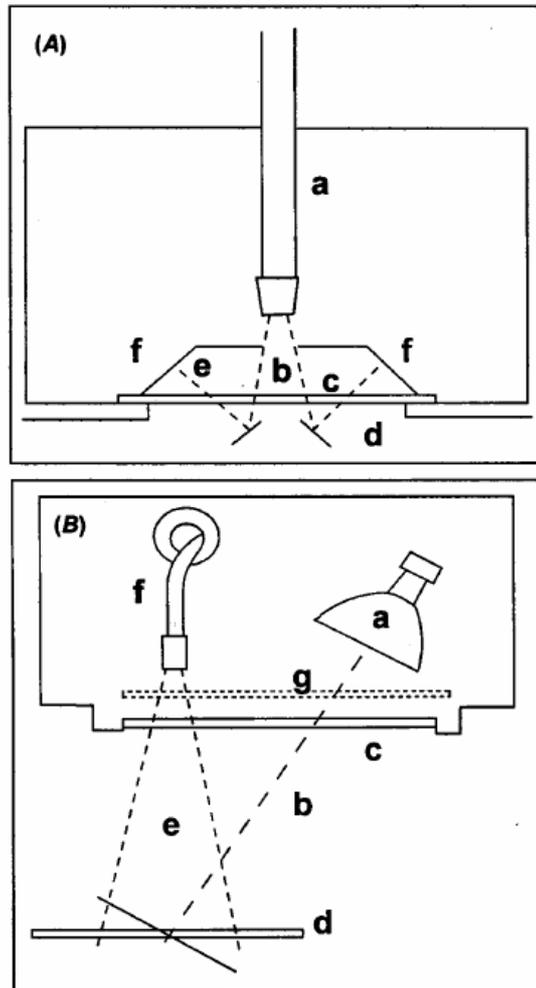
**Figure 1** Schematic diagram of sample presentation to pre- and post-dispersive instrument configurations (NIRSystems 6500). (A) Remote reflectance fibre optic probe employs monochromatic light generated by a scanning monochromator conveyed via a fibre optic bundle (a). Light leaves the fibre optic, radiates at 22° (a function of the numerical aperture of the fibre), passes a slit (b) and a quartz glass window (c) before interacting with the sample (d). The direction of incident light is perpendicular to the sample surface. Diffusely reflected radiation (e) is measured by detectors (f) mounted at 45° to the sample surface. Two lead sulphide detectors and one silicon detector are mounted on each side of the incident light slit. The sample (d) was positioned adjacent to the quartz glass window (c). (B) Direct light sensing head employs white light generated by a 75 W tungsten halogen lamp (a). The incident white light (b) was directed at the sample surface (d) immediately under the fibre optic (f). The sample was positioned 7 cm from the quartz glass window (c). Diffusely reflected light (e) was conveyed via the fibre optic (f) to a scanning monochromator and detector system. A ceramic reference (g) is pneumatically positioned under the lamp and fibre optic between sample measurements.

## Calibration development

Calibration equations were developed on second derivative spectral data using multiple linear regression (MLR) and modified partial least squares (MPLS) analysis with ISI (version 3.0) software. Spectral and Brix analyses were obtained for 208 samples (mean 15.44, range 11.9 - 20.4, s.d. 1.74 $^{\text{o}}$Brix). In the MPLS calibration procedure, the population was partitioned into 6 subgroups, with each group sequentially used as a predicted group and the remainder as a calibration group (such that every sample is predicted once). In contrast, in the MLR procedure, the population was divided into equal groups for use as calibration and prediction sets.

Two procedures were used in dividing the population for the MLR procedure. In the rectangular ('boxcar') distribution method for analyte concentration, the population was first ranked into groups varying by 1°Brix, and then 14 samples (if the division had less than 14, all were used) were taken from each group to make the calibration set. In this method the calibration set is equally weighted for samples across the full Brix range of the population, although the remaining prediction set is overweighted with samples about the mean. In the 'sequential' method, samples were ordered in ascending Brix values, then sequentially paired and a value from each pair randomly split into either calibration or prediction sets. In this procedure both sets are equally weighted for the range of Brix levels present in the population.

## Instrumental and environmental variables

The noise within a spectroscopic condition was characterised with respect to absorption at 876 nm and the predicted Brix value of the assessed fruit. Absorbance

at 876 nm was chosen, as the derivative of absorbance at this wavelength was important in the calibration (MLR, pre-dispersive) developed for pineapple Brix (Table 1). Absorbance values at 876 nm were averaged, and the average divided by the standard error as an estimate of the signal : noise ratio.

### *Number of Scans*

The 'default' option used on the NIRSystems 6500 was to average the spectra of 50 scans per spectrum acquired. With each scan taking about 1 s, spectrum acquisition therefore requires about 1 min. A fewer number of scans are expected to decrease the signal : noise ratio of the measurement. This effect was studied by collecting 5 spectra of a sample (intact pineapple) with paired reference spectra, using either 1, 2, 4, 8, 16, 32, or the maximum allowed, 50 scans per spectra.

### *External Light*

To consider the influence of light, spectra were acquired using the pre-dispersive system under the following 4 conditions: (i) 'standard' practice (e.g. as used by Guthrie and Walsh, 1997), with the remote reflectance fibre optic probe sealed close to the object of interest (a pineapple fruit) in a light proof box; (ii) the lid of this box left open within a room with windows in which the fluorescent overhead lights were turned off; (iii) the lid of the box left open, with room fluorescent lights on; and (iv) a tungsten halogen floodlight directed at the open box. Similarly, spectra were acquired using the post-dispersive system under the following 4 conditions: (i) darkness; (ii) room with unshuttered windows; (iii) room with fluorescent lights on; (iv) floodlight directed at the fruit. Spectra were acquired at 2 locations on 5 fruit (10 spectra) for each of these conditions, for both the pre- and post-dispersive systems.

*Depth of sample assessed*

The effective depth of the sample from which diffusely reflected light was acquired was estimated using the pre- and post-dispersive configurations. Spectra were acquired of filter paper (Whatmans No. 42 – 18.5 cm diameter) soaked in 10% w/v sucrose, within a petri plate on a teflon background. The number of filter papers was varied between 1 and 20. In an alternative consideration, spectra were collected from 10 pineapple fruit in which flesh was sequentially trimmed from the side away from the skin surface facing the NIRS probe.

**Table 1.        MLR and MPLS based calibration of NIRS and pineapple Brix level.  Results are summarised for two protocols for the partitioning of a population into prediction and calibration sets (boxcar and ranked sequentially) for the MLR procedure.**

| | | **Pre-dispersive** | **Post-dispersive** |
|---|---|---|---|
| | **MLR - boxcar calibration (760 - 1300 nm)** | | |
| *Calibration* | | | |
| | n | 105 | 105 |
| | wavelengths (nm) | 740, 764, 788 | 1188, 708 |
| | $R^2$ | 0.740 | 0.705 |
| | SEC | 1.155 | 1.219 |
| *Prediction* | | | |
| | n | 103 | 103 |
| | $R^2$ | 0.28 | 0.38 |
| | SEP | 1.49 | 1.29 |
| | **MLR- ranked sequential  760 - 1300 nm)** | | |
| *Calibration* | | | |
| | n | 104 | 104 |
| | wavelengths (nm) | 876, 764 | 1188, 1244, 708 |
| | $R^2$ | 0.650 | 0.652 |
| | SEC | 1.055 | 0.992 |
| *Prediction* | | | |
| | n | 104 | 104 |
| | $R^2$ | 0.50 | 0.50 |
| | SEP | 1.36 | 1.30 |
| | **MPLS - 6 cross validation groups  (760-2300 nm)** | | |
| | n | 208 | 208 |
| | number of terms | 4 | 3 |
| | $R^2$ | 0.642 | 0.728 |
| | SEC | 1.060 | 0.876 |
| | SECV | 1.112 | 1.011 |

*Temperature and humidity*

To consider the effect of temperature, 5 fruit were varied in temperature between 2 and 60$^{\circ}$C, and spectra acquired using both systems (instruments held at 22-24$^{\circ}$C). Fruit temperature was measured using a thermocouple placed under the skin of the fruit. Reference spectra (ceramic) were taken at laboratory temperature. In a parallel experiment, relative humidity of the environment of the remote reflectance probe and sample was varied, and spectra acquired. Reflectance probe and sample or spectrometer were enclosed within a cabinet and the humidity decreased by recirculating air through a cold trap, and increased by introducing steam into the recirculating air supply, and monitored using a 'Tinytag' datalogger.

The effect of humidity was also considered with respect to the instrument. In this experiment, the instrument was housed in the chamber in which humidity was altered, and the probe and sample maintained under ambient laboratory conditions (i.e. constant humidity and temperature).

# RESULTS AND DISCUSSION

## Calibration for pineapple Brix

MLR calibrations developed using post-dispersive analysis were equivalent to those using pre-dispersive analysis in terms of calibration and prediction regressions (Table 1). For the MPLS-based calibrations, post-dispersive analysis yielded a superior result to the pre-dispersive analysis, in terms of coefficient of determination ($R^2$, standard error of calibration (SEC) and standard error of cross-validation (SECV) (Table 1). Coefficients of determination of 0.73 and 0.64, achieved using post- and pre-dispersive analyses, respectively, will allow grading of articles into 3 grades at a

success rate of 70% and 65% respectively (Shenk and Westerhaus, 1993). High and low samples will be correctly identified with a success rate of 99% and 100%, respectively.
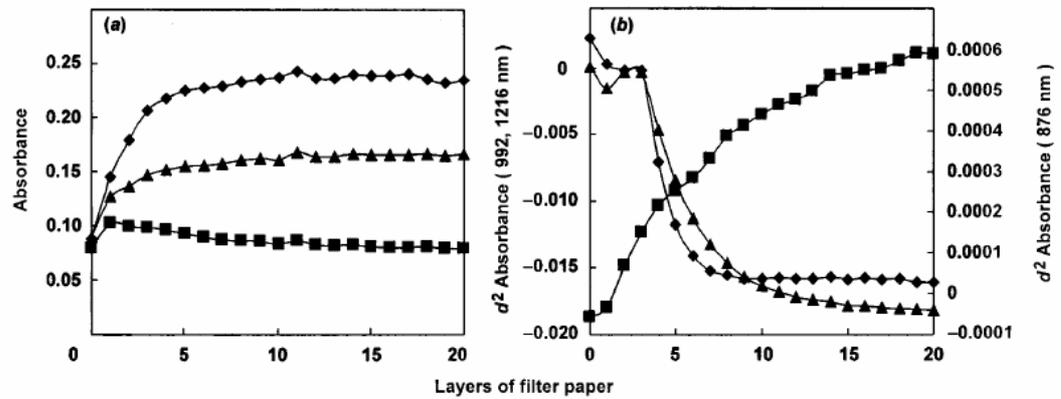


**Figure 2** **(a) Absorbance and (b) second derivative of absorbance at 3 wavelengths ( ■ 876 nm, ▲ 992 nm, ♦ 1216 nm) plotted against the number of layers of filter paper soaked in 10% sucrose solution. Measurements were made using a pre-dispersive system.**

Two procedures were used for the selection of calibration samples from the overall population for the MLR (Table 1). The 'boxcar' selection procedure gave a more uniform weighting across the range of Brix levels than the sequential selection procedure. This resulted in a superior $R^2$ for calibration, in comparison with the 'sequential' method. However, this protocol depleted the number of samples in the prediction set at the extreme ends of the range. In consequence, the $R^2$ and SEP of the prediction set were inferior in the 'boxcar', relative to the 'sequential', procedure.

Brix content was predicted for a population of fruit harvested in 1996 (Tables 2 and 3; Figs. 3-5) using calibration equations developed from a previous growing season (1995, Table 1) for both pre-and post-dispersive systems. The predictions were reasonably precise but inaccurate, with an offset of 1 and 6°Brix for the pre- and post-dispersive systems respectively. This result confirms the need to develop robust calibrations across growing seasons (Guthrie *et al*. 1998).

The signal : noise ratio for $A_{876}$ (as indexed by the ratio of mean to standard error for 5 measurements, Tables 2 and 3) was lower in the pre-dispersive, relative to the post-dispersive, system.  We consider that this result is largely due to differences in the position of the ceramic reference, which was scanned between each sample.  In the pre-dispersive system the reference was placed in the same position as the sample, whereas in the post-dispersive system the reference was built into the direct light head, which was held 7 cm above the sample.
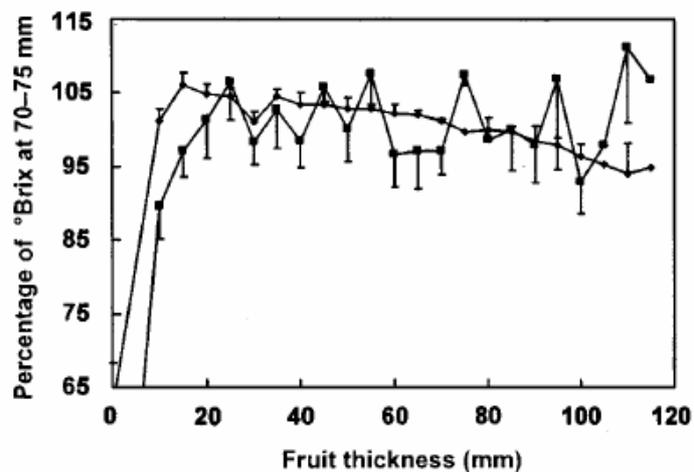


**Figure 3.**     **Predicted Brix of pineapple fruit, as a percentage of that predicted at 70-75 mm.  Slices were cut from each fruit (on the side away from the light source and detector) between each measurement.  Predictions were made using a pre-(♦) and post (■) dispersive system, with calibrations reported in Table 1. Each data point represents the mean of 10 separate fruit with an associated s.e.m.**

**Table 2.** Effect of varying external light conditions on absorbance (876 nm) and predicted Brix of pineapple fruit assessed using a pre- or post-dispersive system. Date presented as mean (*n* = 10) and mean/s.e. as an estimate of the signal : noise ratio. Values in parentheses are expressed as a percentage of the dark value.

| Light conditions | Pre-dispersive | | Post-dispersive | |
|---|---|---|---|---|
| | Mean | Mean/s.e. | Mean | Mean/s.e. |
| *Absorbance (876 nm)* | | | | |
| Dark | 0.1394 | 648 | 0.8832 | 2331 |
| Ambient | 0.1381 | 631 | 0.8820 | 2333 |
| + Fluorescent | 0.1366 | 353 | 0.8791 | 2819 |
| + Floodlight | 0.0413 | 4 | 0.8486 | 2392 |
| *Predicted Brix (°Brix)* | | | | |
| Dark | 19.7 (100) | 263 | 14.8 (100) | 18.7 |
| Ambient | 19.7 (99.8) | 272 | 13.2 (89.1) | 13.2 |
| + Fluorescent | 20.9 (101.2) | 314 | 13.9 (94.7) | 19.1 |
| + Floodlight | 17.7 (89.6) | 37 | 14.7 (98.9) | 14.0 |

**Table 3.** Effect of varying number of scans averaged for each spectrum on absorbance and predicted Brix using a pre- and post-dispersive system. Data presented as mean and mean/s.e. for the absorbance data and as mean and s.e. for predicted Brix date (*n* = 5). Values in parentheses are expressed as a percentage of 50 scans.

| No. of scans | Pre-dispersive | | Post-dispersive | |
|---|---|---|---|---|
| | Mean | Mean/s.e. or s.e. | Mean | Mean/s.e. or s.e. |
| *Absorbance (876 nm)* | | | | |
| 1 | 0.0928 (92.7) | 41 | 0.825 (100) | 1761 |
| 2 | 0.0974 (97.2) | 209 | 0.826 (100.0) | 2611 |
| 4 | 0.0971 (97.0) | 348 | 0.826 (100.1) | 5449 |
| 8 | 0.097 (96.9) | 321 | 0.826 (100.1) | 5782 |
| 16 | 0.085 (98.3) | 315 | 0.826 (100.1) | 10323 |
| 32 | 0.0997 (99.6) | 448 | 0.826 (100.1) | 5065 |
| 50 | 0.1002 (100) | 537 | 0.826 (100.0) | 5024 |
| *Predicted Brix (°Brix)* | | | | |
| 1 | 20.6 (97.8) | 0.38 | 9.3 (72.2) | 3.9 |

| | | | | |
|---|---|---|---|---|
| **2** | 21.2 | 0.22 | 8.0 | 2.2 |
| | (100.4) | | (64.9) | |
| **4** | 21.0 | 0.20 | 15.7 | 2.9 |
| | (99.7) | | (124.1) | |
| **8** | 21.1 | 0.19 | 14.3 | 1.6 |
| | (99.8) | | (113.3) | |
| **16** | 21.2 | 0.15 | 13.5 | 1.6 |
| | (100.4) | | (105.7) | |
| **32** | 21.0 | 0.05 | 14.1 | 0.8 |
| | (99.6) | | (119.4) | |
| **50** | 21.1 | 0.06 | 12.7 | 0.7 |
| | (100.0) | | (100.0) | |

In consequence, the absorbance measured in the post-dispersive system was much higher than that measured in the pre-dispersive system. Given absorbance is a logarithmic scale, the post-dispersive system is expected to give a higher signal : standard error ratio (but at the expense of signal resolution).

*Penetration*

Spectra were acquired of filter paper soaked in 10% (w/v) sucrose using the pre-dispersive configuration (Fig. 2). Maximum separation of absorbance (actually reflectance) spectra occurred at 992 and 1,216 nm, while absorbance at 876 nm was considered as a wavelength of significance in the pineapple Brix calibration (MLR pre-dispersive, Table 1). Absorbance at 992 and 1,216 nm increased with number of layers, reaching a plateau at about 6 layers (6 mm thickness of paper layers). Absorbance at 876 nm demonstrated little relationship to number of layers, suggesting a greater penetration of the sample by the wavelengths 992 and 1,216 nm. The initial slope of the relationship between absorbance and number of layers is related to the extinction coefficient of the sample.

As absorbance data is prone to spectral baseline shifts due to changes in sample surface reflectance, first or second derivative data are generally used in calibration exercises. The second derivative term is negatively correlated with the concentration

of the analyte absorbing at that wavelength, to the extinction coefficient of the analyte, and to the pathlength (from Beer's Law). As expected, the second derivative of spectra provided more information about the sample. Depending on wavelength, spectral information was obtained from a depth of 10-20 mm of filter paper. The decrease in the second derivative of absorbance at 992 and 1,216 nm, stabilising at about 10 layers (10 mm) can be explained by the absorbance of sucrose, water or cellulose at these wavelengths (i.e. to increasing numbers of paper layers). The increase in the second derivative of $A_{876}$ stabilising at about 20 layers (20 mm) indicates that this term is negatively correlated with the analyte (in this case, number of paper layers). Similar data were obtained using the post-dispersive system (data not shown).

In an alternative approach to the question of the effective depth of the sample from which diffusely reflected light was acquired, Brix was predicted for fruit of varying slice thickness. Using the pre-dispersive configuration, the predicted Brix on a slice of pineapple fruit only 10 mm thick was not different to that of thicker slices of fruit (Fig. 3). Using the post-dispersive configuration, the predicted Brix of pineapple increased with fruit thickness up to about 20 mm. There was more noise on post-dispersive estimation of Brix. The temperature of the fruit will have increased during the prolonged exposure to high light intensity in this experiment. Temperature will affect the wavelength at which maximum absorption occurs for a given chemical bond, and thus the accuracy of a Brix prediction. However, the standard error of the prediction did not consistently increase with the time of exposure (i.e. as fruit was sliced and thickness decreased, Fig. 3) and thus change in temperature does not explain this effect. The consistent decrease on the pre-dispersive estimate with increasing thickness of fruit could be explained by reflection from the teflon

196

background back into the fruit, although this trend was not apparent in the post-dispersive data.

The post-dispersive mode offers the advantage of a higher incident light intensity relative to the pre-dispersive mode, because of the loss of intensity with passage of light through a slit and over a diffraction grating in the pre-dispersive mode. If incident light intensity was doubled, the number of photons at all levels within the fruit will be doubled, improving the signal to noise of the diffusely deflected radiated light and increasing the effective depth of sample for which useful spectral information is acquired. Therefore the diffusely reflected light should contain more relevant information about the internal composition of the fruit (i.e. flesh Brix).

*External light effects*

The pre-dispersive system is expected to be more sensitive to stray light entering the detector than the post-dispersive system. When using the post-dispersive system, $A_{876}$ was not affected by background light levels (even a flood light, Table 2). In contrast, the signal : noise ratio of measurements made using the pre-dispersive system were significantly decreased by increasing external light levels.

The use of second derivatives should remove the effect of (constant) external light on the absorbance spectra with respect to the calibration for fruit Brix, although only if this a external light is spectrally 'neutral' (i.e. its influence is constant over that part of the spectrum of importance to the calibration). The mean and standard error of the predicted Brix of fruit assessed with the post-dispersive technique was not significantly altered by external light levels. The mean and standard error of predicted Brix of fruit assessed with the pre-dispersive technique was affected by the tungsten halogen floodlight, but not by fluorescent lighting. These results were

expected as fluorescent light sources do not produce light of wavelengths relative to the Brix calibration, and the presence of additional white light from a tungsten halogen source should affect the pre-dispersive technique adversely.

*Temperature*

Kawano *et al*. (1995) has emphasised the need to incorporate samples over the range of temperatures expected in an operational setting within a calibration exercise. The calibrations reported in Table 1 were developed over a narrow temperature range (22-24°C). As the temperature of the fruit increased so did the predicted Brix values for the pre-dispersive configuration (Fig. 4). However, predictions for the post-dispersive configuration showed no consistent trend (Fig. 4). The increased light intensity on the sample inherent with the post-dispersive system adds a heat load to the sample. Fruit temperature (as monitored by a thermistor placed 10 mm under the skin surface) increased by about 2°C over the scanning period. However, this effect will be similar in both calibration and prediction samples and therefore does not explain the result.

Incorporation of a range of sample temperatures into a calibration equation, as suggested by Kawano *et al.* (1995), will reduce the accuracy and precision of the prediction. Alternatively, calibrations could be developed for a series of temperature ranges, to cover the expected sample temperatures. Temperature of the fruit sample was predicted reasonably accurately using either pre- or post-dispersive configurations (Fig. 4, $R^2 = 0.97$ and SEC = 2.3° Brix; $R^2 = 0.97$ and SEC = 2.7° respectively). Such a temperature prediction could be used to select a relevant calibration equation.
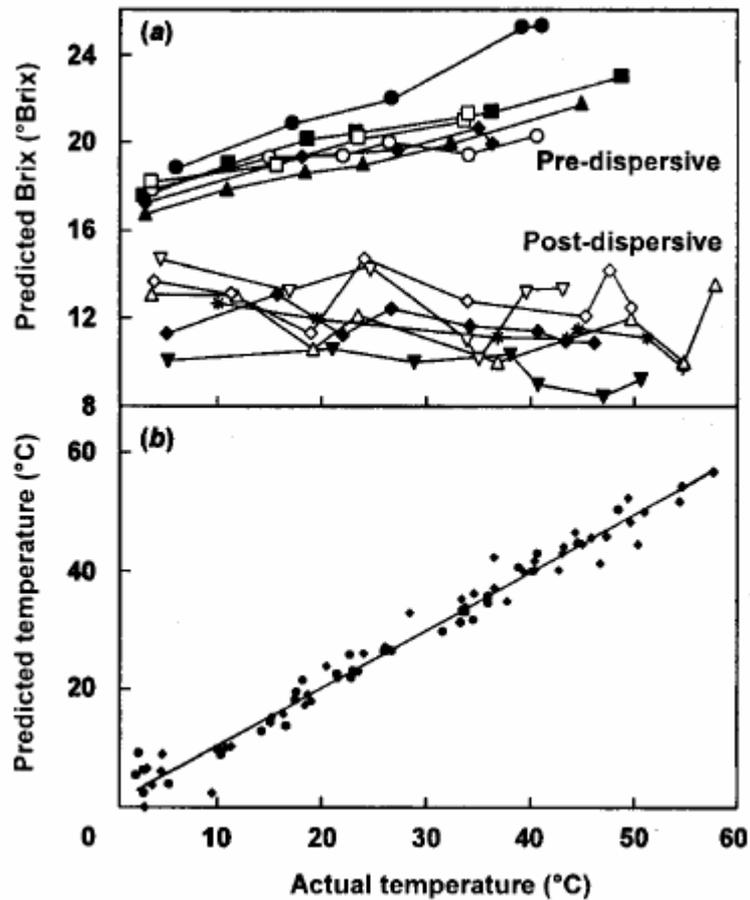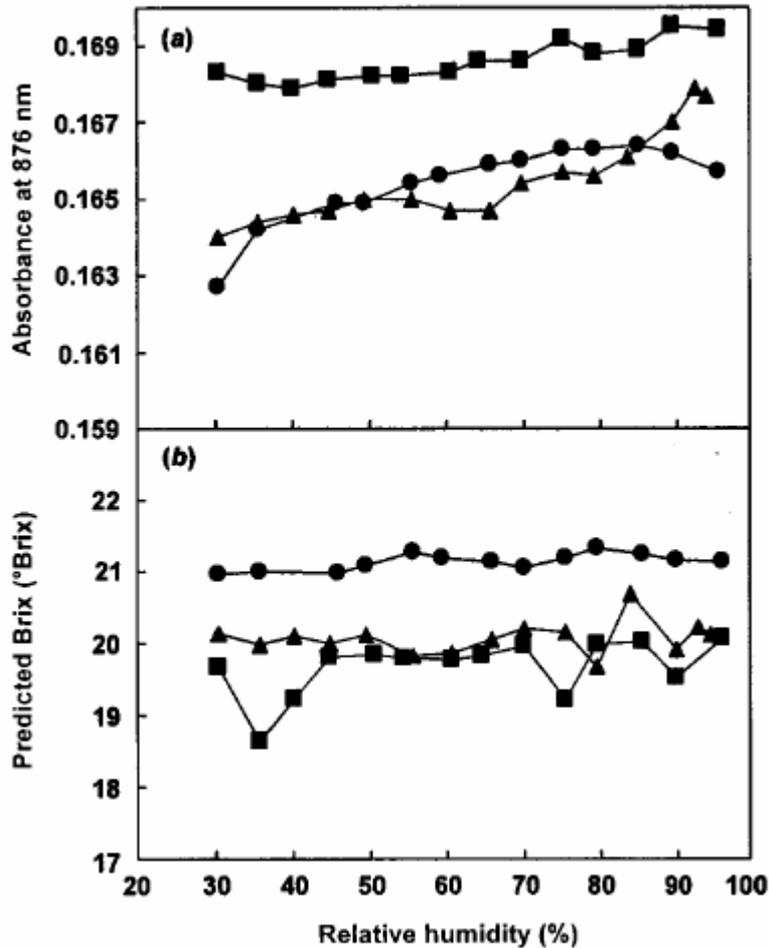
**Figure 4.** **(a) Predicted Brix of 5 separate fruit over a range of temperatures using a pre- and post-dispersive system. (b) Fruit temperature was predicted from spectra. For the pre-dispersive system (◆) a calibration $R^2$ = 0.97, SEC = 2.3 and SECV = 3.0°C was achieved. For the post-dispersive system (●) a calibration $R^2$ = 0.97, SEC = 2.7 and SECV = 3.4°C was achieved.**

**Figure 5.** (a) $A_{876}$ and (b) predicted Brix of pineapple fruit over a range of humidities. Humidity was varied in the vicinity of the sample and remote reflectance probe, with a reference value taken at 57& relative humidity only (●). The sample and remote reflectance probe were maintained under constant conditions while the spectrometer was subjected to varying humidities and either reference taken at one humidity level only (■) or taken paired with each sample (▲).

### Relative Humidity

Increasing the humidity of the sample environment, while temperature was held constant, had minimal effect on the predicted fruit sample Brix value using the pre-dispersive system (Fig. 5). However, when the instrument (i.e. monochromator) was subjected to a range of humidity values, the predicted Brix values were more erratic than in the former situation. This result could be caused by the light scattering effect

of water droplets on the incident radiation.  Re-referencing between samples, at the

prevailing humidity, served to improve the precision of the Brix prediction.

### *Number of scans per spectrum*

As noted with external light effects, predicted Brix was inaccurate, and the

difference in the $A_{876}$ mean/s.e. estimate between pre- and post-dispersive systems is

interpreted as largely due to the position of the reference (Table 3).  The mean/s.e.

ratio of both $A_{876}$ and predicted Brix improved with the number of scans averaged for

each spectrum in both systems.  A standard error of less than 1° predicted Brix was

achieved with only 1 scan in the pre-dispersive system, while 32 or more scans were

required with the post-dispersive system to achieve the same result.  The time

required to acquire each spectrum is a limitation to the adoption of NIRS technology

into pack house settings.  Fruit packing lines currently sort fruit at rates in excess of 1

item per second (typical belt speed 1 m/s).

## CONCLUSION

The use of a NIR-based technology for the sorting of fruit in a packing shed

requires a robust and rapid technology, 'tolerant' of changes in temperature and

humidity, and capable of assessment of at least 2 pineapple fruit per second.  A post-

dispersive optical configuration is recommended over a pre-dispersive system to

increase the incident light intensity on the sample (i.e. to assess a greater depth of

fruit) and to decrease sensitivity to stray white light.  A detector array (CCD or

photodiode) rather than a scanner grating and a single detector is required to decrease

analysis time.  These 2 features are seen in prototype 'bench top' NIRS units for Brix

assessment of kiwifruit (Osborne *et al*. 1998), peach (Jaenisch *et al*. 1990), melon (Matumato *et al*. 1996) and apple (Bellon *et al*. 1993).

The influence of relative humidity on prediction can be addressed by re-referencing as humidity changes, or by enclosing the spectrometer. The influence of sample temperature on prediction accuracy is suggested to be best addressed by predicting temperature by using NIRS, and applying a calibration developed for this temperature. Consideration should also be given to the effect of temperature changes on the spectrometer. We are currently undertaking further work to integrate a post-dispersive array system into a commercial fruit packing shed.

## ACKNOWLEDGEMENTS

## REFERENCES

Bellon, V., Vigneau, J.L., and Leclerq, M. (1993). Feasibility and performance of a new, multiplexed, fast and low cost fiber-optic NIR spectrometer on-line measurement of sugar in fruits. *Applied Spectroscopy* **47**, 1079-83.

Guthrie, J. and Walsh, K. (1997) Non-invasive assessment of pineapple and mango fruit quality using near infra-red spectroscopy. *Australian Journal of Experimental Agriculture* **37**, 253-263.

Guthrie, J., Wedding, B. and Walsh, K. (1998) Robustness of NIR calibrations for soluble solids in intact melon and pineapple. *Journal of Near Infrared Spectroscopy* **6**, 259-265.

Jaenisch, H.M., Niedzwiecki, A.J., Cernosek, J.D., Johnson, R.B., Seeley, J.S., Dull, G. G., and Leffler, R.G. (1990). Instrumentation to measure the near-IR spectrum of small fruits. Society of Photo-optical Instrumentation Engineers; Bellingham, WA,USA. *Optics in Agriculture* **1379**, 162-7.

Kawano, S.(1994) Present condition of nondestructive quality evaluation of fruits and vegetables in Japan. *Japan Agricultural Research Quarterly* **28**, 212-216

Kawano, S., Abe, H., and Iwamoto, M. (1995). Development of a calibration equation with temperature compensation for determining the Brix values in intact peaches. *Journal of Near Infrared Spectroscopy* **3**, 211-18.

Kawano, S. Watanabe, H., and Iwamoto, M. (1992) Determination of sugar content in intact peaches by near infra-red spectroscopy with fibre optics in interactance mode. *Journal of the Japanese Society of Horticultural Science* **61**, 445-451.

Matumoto, K., Aoki, H., Mizuno, K., and Maeda, H. (1996). Development of a new near infra-red spectrophotometer. *In*: 'Proceedings of the International Conference on Tropical Fruits'. Kuala Lumpur, Malayasia. (Ed. S. Vijaysegaran.) pp. 207-12. (Malayasian Agricultural Research and Development Institute: Kuala Lumpur.)

Osborne, S.D., Jordan, R.B., and Kummemeyer, R. (1998). Using near infra-red (NIR) light to estimate the soluble solids and dry matter of kiwifruit. *Acta Horticulturae* **464**, 109-14.

Shenk, J.S., and Westerhaus, MO (1993)  NIRS Handbook  Analysis of agriculture
and food products by near infra-red reflectance spectroscopy.  Infrasoft
International, marketed by NIRSystems, Inc. Silver Springs, MD, USA.

Shiina, T., Ijiri, T., Matsuda, I., Sato, T., Kawano, S., and Ohoshiro, N. (1993)
Determination of Brix value and acidity in pineapple fruits by near infra-red
spectroscopy. *Acta Horticulturae* **334**, 261-272

# ROBUSTNESS OF NIR CALIBRATIONS FOR SOLUBLE SOLIDS IN INTACT MELON AND PINEAPPLE[7]

## ABSTRACT

The soluble solids content of intact fruit can be measured non-invasively by near infrared spectroscopy, allowing 'sweetness' grading of individual fruit. However, little information is available in the literature with respect to the robustness of such calibrations. We developed calibrations based on a restricted wavelength range (700-1100 nm), suitable for use with low-cost silicon detector systems, using a stepwise multiple linear regression routine. Calibrations for total soluble solids (°Brix) in intact pineapple fruit were not transferable between summer and winter growing seasons. A combined calibration (data of three harvest dates) validated reasonably well against a population set drawn from all harvest dates ($\sim = 0.72$, SEP = 1.84 °Brix). Calibrations for Brix in melon were transferable between two of the three varieties examined. However, a lack of robustness of calibration was indicated by poor validation within populations of fruit harvested at different times. Further work is planned to investigate the robustness of calibration across varieties, growing districts and seasons.

---

# INTRODUCTION

Fruit of the pineapple accumulates soluble sugars and acids during fruit maturation, with little change in internal chemical composition occurring after harvest [1]. As with pineapple fruit, the mesocarp of melon fruit does not contain starch, and thus the sugar content of the edible flesh does not change following harvest[2]. Soluble sugars represent the major component of the total soluble solids content (° Brix) in these fruit, and thus refractometry of flesh juice represents a reasonable measure of sugar content.

For pineapple, fruit eating quality is related to Brix, acidity, pH, Brix/acid ratio and ester concentrations.[1] The variable best correlated with eating quality was Brix (linear relationship, coefficient of determination, $R^2 = 0.70$). Fruit of less than 14° Brix was considered to be unacceptable for the fresh market[1].

For melon fruit, consumer acceptability is correlated with sugar content at harvest,[2] although it has been noted that melon eating quality depends not only on sweetness, but on various volatile compounds[3] However, while a high Brix alone does not adequately define melon eating quality,[2,4] the absence of high Brix makes good quality very unlikely[5]. Research has confirmed that flesh firmness and Brix are more useful indicators of rockmelon maturity and acceptability than titratable acidity or juice pH.[6] Similarly, in a neural net analysis using three physical measurements of honeydew quality, only Brix had consistent links to fruit flavour and sweetness as rated by a consumer preference study[7]. This link was of a linear form, allowing the use of regression techniques. In sensory panel work,[2] overall preference for melon was strongly correlated ($r = 0.97$) with perceived sweetness of the fruit, although the

correlation (*r*) of panel sweetness preference with Brix was only 0.52°. However, descriptive panels alone may not predict consumer response very well[7].

In the USA, legal minimum soluble solids (SS) levels has been set for the marketing of netted muskmelon (9 and 11 ° Brix for US No. 1 and Fancy Grades, respectively),[8] and similar standards exist for honeydew melons[4]. Some melon types, such as netted muskmelon and cantaloupe, develop a visible maturity index, such as an abscission layer or exocarp colour change, after the fruit has reached maximum sugar concentration. Other melon types, such as honeydew type melons, have no such qualitative index, and thus immature fruit are likely to be harvested along with mature fruit[2].

We conclude that a non-invasive measure of fruit Brix would be useful in the assessment of both fruit maturity and eating quality. The Brix content of the fruit can be measured non-invasively by near infrared (NIR) spectroscopy, thus allowing the grading of individual fruit. Regression techniques are typically employed in such applications to correlate second derivative spectral data to fruit Brix content. NIR spectroscopy has been used to assess the Brix content of intact pineapple (e.g. $R^2$ of 0.75, *SEC* of 1.21 ° Brix[9]), sliced melon tissue and intact melon fruit (e.g. $R^2$ of 0.94, *SEC* of 0.56% and $R^2$ of 0.36, *SEC* of 1.67%, respectively)[10]. The decrease in predictive ability with the intact fruit was ascribed to a lack of substances in the rind that were related to Brix in the edible portion. Subsequent work employing stronger light sources, improved detectors and multiple regression data processing has improved the calibration for melon[11,12,13].

However, these calibrations and those developed for other intact fruit (e.g. peach,[14] kiwifruit[15]), involve only one cultivar, from one growing district and from one growing season. To our knowledge, only one paper has been published on the

robustness of such calibrations. Peiris *et al.*[16] reported on the calibration of NIR spectra with the SS of (one cultivar of) peach collected over three years. Individual season calibrations were successful in predicting SS within that season, but not for other seasons, as indicated by higher *SEP* (to 1.52%) and lower *r* (to 0.48). In contrast, a calibration based on data from all three seasons was superior in predicting SS in any season (highest *SEP*, 1.06%; lowest *r*, 0.66). This improvement was attributed to the inclusion of a broader range of SS in the combined calibration data set.

In the current study we attempt to describe the robustness of a calibration developed for pineapple flesh Brix[9] across growing seasons, and a calibration developed for melon flesh Brix across varieties.

## MATERIALS AND METHODS

Pineapple fruit [*Ananas comosus* (L.) Merrill cv. "Smooth Cayenne"] of the one cultivar and growing district (Yeppoon, Central Queensland) were collected during December 1995, August and September 1997. Melon fruit [*Cucumis melo* (L.) varieties "El Dorado", "Eastern Star" and "Hammersley"], were collected over August and September, 1997 from the one growing district (Burdekin, North Queensland). Approximately 100 fruit were assessed per population (Table 1).

A Linbrook–NIRSystems 6500 spectrophotometer (Silver Springs, MD, USA) equipped with a remote reflectance fibre optic probe was used to collect NIR reflectance spectra (400–2500 nm, predispersive configuration) from an area of 16 cm2 of the surface of intact rockmelon and pineapple fruit[9]. Two spectra were collected per fruit, from opposite sides of the equatorial region. The part of the fruit that touched the ground during growth was avoided. Spectral attributes were

correlated against juice Brix of the fruit flesh (60 mm diameter core taken to approximately 20 mm depth). Juice Brix was measured using an Erma (Tokyo, Japan) digital refractometer.

Fruit were assigned to calibration and validation sets following ranking of the fruit by average (of the two measurements per fruit) analyte concentration. Fruit of similar Brix content were paired and one fruit from each pair randomly allocated to the calibration set and the other to the validation set. Samples within each fruit were then treated as separate spectral sets. The combined variety / growing season calibration and validation sets were developed similarly with approximately equal numbers of each variety in a set.

Data treatment followed that established in a previous study,[9] using ISI software (version 3.0; Infrasoft International, PA, USA). Analysis involved stepwise multiple linear regression (MLR) of second derivative data from the 700 to 1100 nm region of the spectrum. Gap and smoothing of four data points, respectively, was undertaken (i.e. ISI math treatment of 2,4,4,1). Standard normal variance (SNV) and detrend was used as the option in scatter correction.

**Table 1.** **Calibration population statistics and regression parameter for three melon varieties and three harvest dates of pineapple.**

| Melon | $n$ | Min | Max | Mean | SD | $R^2$ | SEC (°Brix) | Wavelengths (λ, nm) |
|---|---|---|---|---|---|---|---|---|
| El Dorado a | 86 | 6.7 | 13.2 | 10.6 | 1.49 | 0.82 | 0.63 | 948, 844, 876, 900 |
| Eastern Star | 92 | 9.8 | 13.6 | 11.7 | 0.80 | 0.44 | 0.60 | 892, 844, 948 |
| Hammersley | 117 | 5.5 | 13.4 | 8.1 | 1.55 | 0.70 | 0.84 | 1020, 900, 780, 884, 1036 |
| Combined (EDa,ES,H) | 91 | 5.5 | 13.6 | 10.1 | 2.02 | 0.81 | 0.89 | 1012, 884, 980, 780, 908 |
| Combined (EDa,H) | 102 | 5.5 | 13.4 | 9.24 | 2.00 | 0.81 | 0.87 | 860, 1004, 900, 876 |
| El Dorado b | 96 | 7.1 | 12.7 | 9.9 | 1.17 | 0.40 | 0.91 | 820, 1004, 908 |
| Pineapple | $n$ | Min | Max | Mean | SD | $R^2$ | SEC (°Brix) | Wavelengths (λ, nm) |
| December 1995 | 103 | 11.5 | 21.0 | 15.4 | 1.90 | 0.65 | 1.12 | 1052, 1028, 756, 740 |
| August 1997 | 101 | 6.3 | 18.5 | 12.0 | 3.28 | 0.85 | 1.27 | 1052, 876, 860, 988 |
| Combined (12/95,8/97,9/97) | 90 | 7.2 | 18.9 | 12.2 | 3.23 | 0.79 | 1.47 | 988, 900, 876 |
| September 1997 | 93 | 6.6 | 15.3 | 10.0 | 1.94 | 0.72 | 1.022 | 844, 900, 876, 828, 868 |

Table 2. Validation regression parameters. Calibrations described in Table 1 were validated against population sets (*n* as for Table 1) as indicated.

**Calibration groups**

| Validation Groups — Melons | El Dorado a | | Eastern Star | | Hammersley | | Combined (EDa/ES/H) | | Combined (EDa/H) | | El Dorado b | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $r^2$ | SEP (°Brix) | $r^2$ | SEP (°Brix) | $r^2$ | SEP (°Brix) | $r^2$ | SEP (°Brix) | $r^2$ | SEP (°Brix) | $r^2$ | SEP (°Brix) |
| El Dorado a | 0.77 | 0.74 | | | 0.56 | 1.32 | 0.69 | 0.92 | 0.65 | 0.96 | | |
| Eastern Star | 0.12 | 2.31 | 0.10 | 0.86 | 0.12 | 2.41 | 0.12 | 1.69 | 0.14 | 2.02 | | |
| Hammersley | 0.61 | 1.32 | | | 0.65 | 0.95 | 0.59 | 1.13 | 0.66 | 0.93 | | |
| Combined (EDa/ES/H) | 0.45 | 1.53 | | | 0.58 | 1.64 | 0.55 | 1.43 | | | | |
| Combined (EDa/H) | 0.69 | 1.14 | | | 0.73 | 1.15 | | | 0.75 | 0.91 | | |
| El Dorado b | 0.12 | 1.32 | | | 0.12 | 1.44 | 0.15 | 1.50 | 0.13 | 1.35 | 0.17 | 1.13 |

**Calibration groups**

| Validation Groups — Pineapple | December '95 | | August '97 | | Combined (12/95, 8/97, 9/97) | | September '97 | |
|---|---|---|---|---|---|---|---|---|
| | $r^2$ | SEP (°Brix) | $r^2$ | SEP (°Brix) | $r^2$ | SEP (°Brix) | $r^2$ | SEP (°Brix) |
| December '95 | 0.28 | 1.64 | 0.08 | 3.17 | 0.20 | 1.87 | 0.19 | 3.60 |
| August '97 | 0.42 | 6.93 | 0.66 | 1.94 | 0.76 | 1.84 | 0.63 | 2.20 |
| Combined (12/95, 8/97, 9/97) | | | | | 0.72 | 1.84 | | |
| September '97 | 0.25 | 8.57 | 0.46 | 2.54 | 0.44 | 1.76 | 0.34 | 1.63 |

# RESULTS

## Fruit Brix characteristics

The Brix of melon flesh varied from 5.5 to 13.6°, with little difference between varieties (Table 1), although the variety "Eastern Star" tended to have a higher and narrower range of Brix (9.8–13.6°). The Brix of pineapple flesh varied from 6.3 to 21°, with summer (December) fruit having higher values than winter (August/September) fruit. Fruit below 10° and 14°Brix for melon and pineapple, respectively, are generally considered unacceptable for the fresh market.

In pineapple fruit, skin and flesh Brix is related[9]. However, for melon fruit, flesh Brix (8–28 mm) was poorly related to both skin (outside 2 mm) and green (c. 2–8 mm) layers (Figure 1).

## Melon calibration

For melon, reasonable calibrations were obtained for one population of the variety "El Dorado" (EDa) and for "Hammersley" (H) (coefficient of determination $R^2$ = 0.82 and 0.70, standard error of calibration $SEC$ = 0.63° and 0.84°Brix, respectively) (Table 1). A poor calibration was established for the variety "Eastern Star" (ES) and for a second population of "El Dorado" (EDb) (Table 1). A reasonable calibration was obtained using a combined calibration population of all varieties [EDa, ES and H (Figure 2)] or ED and H only ($R^2$ = 0.73 and 0.81, $SEC$ = 1.02° and 0.86°Brix, respectively).

Predictions of Brix in fruit of EDa and H populations validated well against calibrations developed from the fruit of their respective populations ($r^2$ = 0.77 and 0.61, $SEP$ = 0.74° and 1.32°Brix, respectively) and also against each other (Table 2).

Indeed, the validation of the combined EDa and H calibration was similar to that of the variety specific calibrations of EDa and H (e.g. validation for H population, $r^2 = 0.73$, *SEP* = 1.15°Brix). In contrast, the calibrations derived for ES and for a second ED population (EDb) validated poorly against themselves, and other populations. The second population of ED (EDb) validated poorly against the Eda specific calibration ($R^2 = 0.12$, *SEP* = 1.32° Brix), and also against all other calibrations. The calibration derived from the combined population set (EDa, ES and H) had decreased validation accuracy, as expected with the inclusion of the ES population (Table 2, Figure 2).
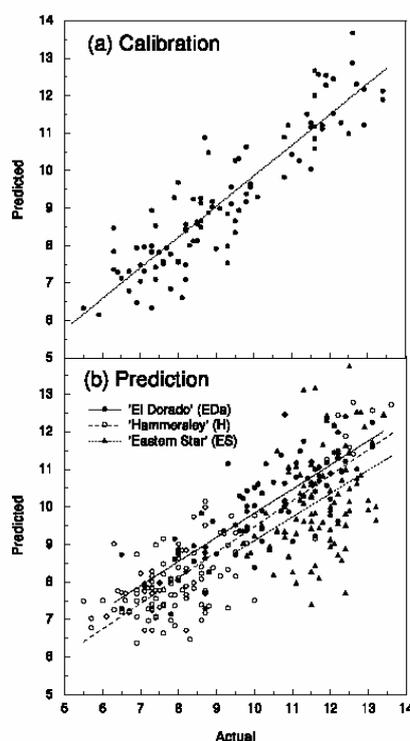


**Figure 2.     Calibration (a) and prediction (b) relationships between NIR predicted and actual flesh Brix of melon fruit. Calibration (a) is of combined data set of three melon varieties (EDa, H and ES). Validation (b) is of calibration relationships shown in (a) tested against population sets of the three varieties (EDa, H and ES).**

## Pineapple calibration

Both summer fruit (December 1995) and winter pineapple fruit (August 1997 and September 1997) supported reasonable calibrations ($R^2 = 0.65$, 0.85, 0.72; *SEC* =

1.12 °, 1.27 °, 1.02° Brix, respectively) (Table 1). A combined calibration using independent samples from all populations also gave a good explanation of the population variance, although with a larger *SEC* than for the individual calibrations.
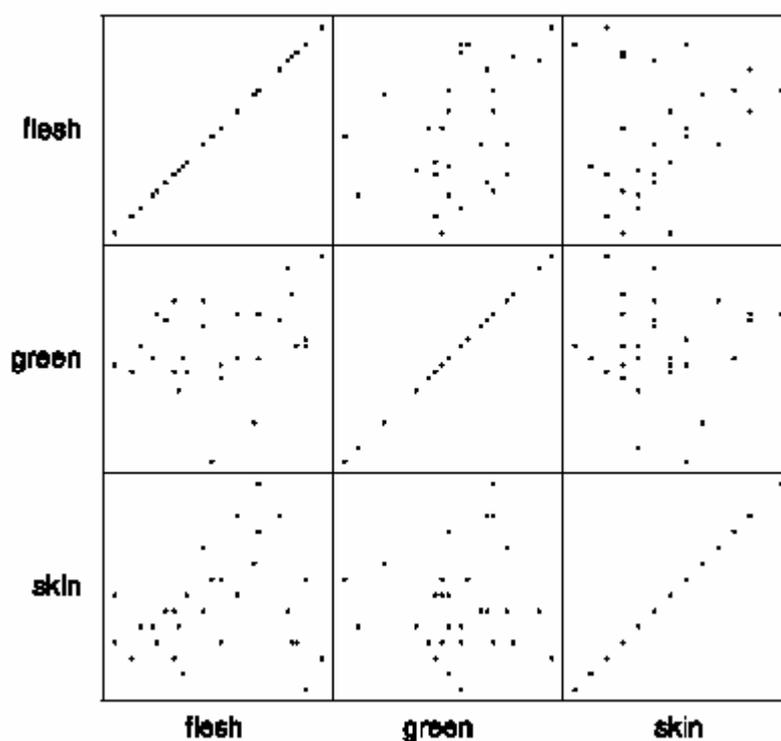


**Figure 1.** **Relationships of the Brix content between skin (outside 2 mm), green (2–5 mm) and flesh (8–20 mm depth) of rockmelon (*Cucumis melo* (L.) variety "El Dorado").**

The winter fruit population harvested in August 1997 validated well against an independent set from the same population ($r2 = 0.66$, *SEP* = 1.94° Brix) (Table 2). In contrast, December 1995 fruit and September 1997 fruit validated poorly against independent sets of the same populations ($r2 = 0.28$ and 0.34, respectively). Despite this, the combined calibration achieved a *r2* of 0.72 and a *SEP* of 1.84° Brix against an independent validation set drawn from all harvest dates (Table 2).

## Discussion

Using the NIRSystems 6500 remote reflectance system, spectral data are obtained from tissue within 5 mm of the fruit surface. 9 As melon skin Brix was not correlated to flesh Brix, the calibration of NIR spectra with melon flesh Brix must represent an indirect correlation with another constituent of the skin/green layers (e.g. it is possible that starch content of the rind is related to Brix content of the flesh). It is reasonable to expect an improved accuracy of calibration if spectral information on the fruit flesh through the thick skin layers could be obtained. The use of an optical arrangement (e.g. increased intensity of incident light, decreased detection of specular radiation) which allows collection of spectral information from a greater depth within the fruit is desirable in this respect[10]. To be commercially useful in fruit grading, NIR spectroscopy must be capable of sorting fruit into at least two grades (i.e. above and below an acceptable Brix value) with approximately 80% accuracy. This requirement involves attainment of a validation correlation coefficient of at least 0.65[17]. This criterium was generally obtained within populations. The exceptions (e.g. melon populations of Eastern Star and El Dorado b) were attributed to the nature of the skin of the fruit (irregular and thicker epidermal layers than other varieties used in this work), preventing collection of spectral information from internal tissues, or by a change in the chemical matrix of the fruit, resulting in spectral "interference". Differences in epidermal thickness or chemical constituency is to be expected between varieties, and can be accommodated by developing variety specific calibrations within a grading system. However, the differences between the two populations of El Dorado melons, harvested within weeks of each other, and from the same farm, was not expected. As expected,16 the application of these calibrations to populations involving other growing seasons (pineapples) or varieties (melons) was

not as successful (lower $r2$, higher *SEP*) as the use of calibrations developed across growing seasons or varieties, with the compromise of a decreased *SEP* in the combined calibration in prediction of its "own" population. With pineapple, calibration equations were not transferable between the summer and winter fruit populations, however, a combined population was sufficiently robust to allow grading of fruit into two grades of sweetness. The application of NIR spectroscopy to the sorting of intact fruit on the basis of Brix value is thus dependent on the development of robust calibrations, developed with respect to variety, growing season (and possibly growing district). We anticipate that such calibrations will need continual updating for changes due to the time of harvest (within and between growing seasons), as practiced in the grains industry[18], and updating or replacement with the introduction of new varieties. Further work is required to address the issue of the stability of the variety specific calibrations across growing districts, growing season and time of harvest within a season.

## Acknowledgements

## REFERENCES

1. L.G. Smith, *Qld J. Agric. and Animal Sci*. **45,** 219 (1988).

2. G.E. Lester and K.C. Shellie, *HortScience* **27,** 1012 (1992).

3. M. Yamaguchi, D.L. Hughes, K. Yabmoto and W.G. Jennings, *Scientia Hort.* **6,** 59 (1977).

4. B.B. Aulenbach and J.T.Worthington, *HortScience* **9,** 136 (1974).

5. V.V. Bianco and H.K. Pratt, *J. Am. Soc. Hort. Sci.* **102,** 127 (1977).

6. L.L. Mutton, B.R. Cullis and A.B. Blakeney, *J.Sci. Food Agric.* **32,** 385 (1981).

7. C.N. Thai, A.V.A. Resurreccion, G.G. Dull and D.A. Smittle, *International Winter Meeting of The American Society of Agricultural Engineers,* St Joseph, Michigan, Paper no. 907550 (1990).

8. USDA, *United States Standards for grades of cantaloupes*. United States Department of Agriculture, Agricultural Marketing Service, Washington, DC (1968).

9. J.A. Guthrie and K.Walsh, *Aust. J. Exp. Agr.* **37,** 253 (1997).

10. G.G. Dull, G. Birth, D. Smittle and R. Leffler, *J. Food Sci.* **54,** 393 (1989).

11. G.S. Birth, G.G. Dull and R.G. Leffler, *Optics in Agric.* **1379,** 10 (1990).

12. G.G. Dull, G.S. Birth, D.A. Smittle and R. Leffler*, Transactions of the ASAE (American Society of Agricultural Engineers)* **35(2),** 735 (1992).

13. H. Aoki, Y. Matumoto, K. Mizuno and H. Maeda, in *Proceedings: International Conference on Tropical Fruits*, Kuala Lumpur, Malaysia. Malaysian Agricultural Research and Development Institute, p. 207 (1996).

14. S. Kawano, H.Watanabe and M. Iwamoto, *J. Japan. Soc. Hort. Sci*. **61(2),** 445 (1992).

15. S.D. Osborne, R.B. Jordan and R. Kümmemeyer, *in Proceedings of the PH'96 International Postharvest Science Conference*, Taupo, New Zealand (1996).

16. K.H.S. Peiris, G.G. Dull, R.G. Leffler and S.J. Kays, in *Proceedings from the Sensors for Nondestructive Testing: Measuring the Quality of Fresh Fruit and Vegetables,* Florida, USA. Northeast Regional Agricultural Engineering Service Cooperative Extension, New York, p. 77 (1997).

17. J.S. Shenk and M.O.Westerhaus, in *NIRS Handbook*. Infrasoft International, marketed by NIRSystems, Inc., Silver Springs, MD, USA, p. 14 (1993).

18. B.G. Osborne, T. Fearn and P.H. Hindle, *Practical Spectroscopy with Application in Food and Beverage Analysis (second edition).* Longman Scientific and Technical, UK, p. 147 (1993).

# DEVELOPMENT AND USE OF AN 'AT-LINE' NIR INSTRUMENT TO EVALUATE ROBUSTNESS OF MELON BRIX CALIBRATIONS[8]

## ABSTRACT

Melon eating quality is largely dependent on soluble sugar content, which can be non-invasively assessed using near infrared spectroscopy. Cost effective application of this technology to fruit sorting requires optimisation of (i) instrumentation, (ii) the optical interface between sample and detector, and (iii) the calibration population and data treatment. In terms of instrumentation requirements, a resolution of 20 nm and signal to noise (standard deviation) ratio of 4,600 was indicated to be adequate for a Si array spectrometer based system for the prediction of sucrose content. An optical system using the Zeiss MMS1 photodiode array spectrometer was optimised in terms of optical geometry, with reference to fruit and lamp(s). This system was used to collect spectra of a range of melon varieties, growing localities and growing times. Calibration sets were trimmed on the basis of global and neighbourhood Mahalanobis distances. For example, one calibration set containing 1991 spectra was reduced to 449 with use of a NH of 1.0, with no loss of prediction precision (SECV). A calibration developed on a single population

---

(variety-time-location) gave poor predictions of populations of the same variety harvested at other times or locations. A calibration developed over five time-locations of the one variety predicted sugar content of other populations of the same variety well, and populations of one of four other varieties. A calibration developed over five varieties proved to be acceptably precise (SECV ca. 0.8 %TSS) and robust.

# INTRODUCTION

Melon eating quality is indexed by total soluble solids (TSS)[1,2]. Other attributes (e.g. volatiles, texture) contribute to eating quality, but TSS is often positively correlated with these attributes, and high TSS is a prerequisite for good eating quality[1]. Therefore, the ability to grade every fruit for TSS (eating quality), as well as external appearance (shape, size, colour, etc) is desired. As TSS can vary between 4 and 16% w/v, and as 80% of TSS is simple sugars (predominately sucrose)[3], a method of measurement of sucrose within intact melons to a resolution of approximately 1% is required for a fruit eating quality assurance program.

Near infrared spectroscopy (NIRS) was first applied in a reflectance mode to the measurement of TSS in melons by Dull *et al.*[3]. A correlation standard error of prediction (SEP) of 1.6% for sliced fruit, and 2.2% for intact fruit, was reported in this work. Subsequent reports of the use of NIRS to assess the TSS of intact fruit show a progressive decrease in the SEP, from 2.2%[4] and 1.9%[3] to 0.4%[5]. This improvement reflects change in the instrumentation used, and in the optical geometry (light-sample-detector) employed. Systems employing reflectance mode suffer from a background of specular light. However, the optical density of melon fruit makes transmission mode difficult to employ. Aoki *et al.*[5] employed a multiple lamp system, with lamps

mounted at 90° to the detector, with respect to the centre of the fruit. NIRS technology is now in commercial use in Japan for melon sorting (e.g. Fantech, Mitsui), with a reported SEP of 0.5%.

The studies mentioned above report the development of a calibration on one population of fruit only, and it is not clear if this calibration is variety-locality-season specific, or is robust across such variations. In a previous study, we[6] employed a NIRSystems 6500 reflectance mode spectrometer to consider the robustness of calibrations across melon varieties, growing seasons and growing locations. A combined calibration was useful (ie. SEP below 1% TSS) across time-locality, and across several, but not all varieties.

In the present study we document the selection and optimisation of a spectrometer system suitable, in terms of cost and speed, for the grading of fruit in an at-line setting, with a view to incorporating these results into an in-line setting, and report on the robustness of calibrations across varieties, time and locality.

## MATERIALS AND METHODS

Cost effective application of this technology to fruit sorting requires optimisation of (i) instrumentation, (ii) the optical interface between sample and detector, and (iii) the calibration population and data treatment.

### Detector attributes.

An optical table based spectrometer was constructed using a Hammamatsu photodiode array, in order to consider the effect of wavelength resolution on calibration. To change resolution, slit width of the system was altered, with a corresponding change in intensity of illumination of sample to maintain a constant amount of light reaching the detector. The spectrum of a Hg-Ar lamp (Ocean Optics,

Dunedin, Florida) was used in the characterisation of resolution. In a parallel experiment, the effect of detector signal to noise on calibration performance was considered. In this exercise, the signal to noise of a Zeiss MMS1 spectrometer unit (Jena, Germany) was altered by changing signal strength or the number of spectra averaged.

In both experiments, a bifurcated fibre optic interactance probe, consisting of eight 400 μm illuminating fibres concentrically arranged about a single 400 μm read fibre (Ocean Optics), was used in conjunction with a tungsten halogen lamp (Ocean Optics) to gather spectra of sucrose solution soaked cellulose filter papers (0-20% w/v sucrose) for calibration characterisation.

Two low cost (<A$5,000), miniaturised spectrometers were chosen for comparison, based on the use of different detector technologies (charge coupled device and photodiode array in the Ocean Optics S2000 and Zeiss MMS1, respectively). Spectra of sucrose soaked filter paper were collected using the two instruments.

### *Optimisation of an optical configuration for melon calibration.*

Melons were obtained from commercial farms (varieties Doubloon, Eastern Star, Hammersley, Highline, and Malibu), with spectral collection and juice extraction and TSS determination made on the same day, and not more than 5 days after harvest. A series of trials were undertaken in terms of calibration performance for a range of angles between the incident light on the fruit surface and the area of fruit detected by the spectrometer, with reference to the centre of the fruit. The protocol of sampling for wet chemistry was also considered with respect to calibration performance. Light distribution within a melon fruit was assessed by sequentially cutting the fruit on the

axis perpendicular to the lamp – fruit centre, and measuring light output over 1 cm$^2$ areas of the fruit surface using the MMS1.

### *Calibration development and robustness.*

Calibrations were developed using WINISI software, using first derivative data (derivative calculated over four data points), without smoothing or scatter correction. Outlier spectra were removed from calibration population sets using the 3.0 Global H criterion of WinISI software.

## RESULTS AND DISCUSSION

## Wavelength Resolution

Wavelength resolution of the MMS1 and S2000 instruments is illustrated by the FWHM of the 912 nm peak of the Hg-Ar lamp spectra.  The MMS1 achieved a 13 nm resolution, while the S2000 achieved a 2 nm resolution (Fig. 1).   Given that the second and third overtone bands assessed using NIRS are typically broad spectral features, ca. 50 nm, a resolution of less than 20 nm should not be necessary.  This view is reinforced by the spectral averaging option typically employed in chemometric procedures[6].  However, a typical MPLS correlation developed on absorbance data has coefficients which can vary widely between spectral data points. This variability hints at a requirement for better resolution.
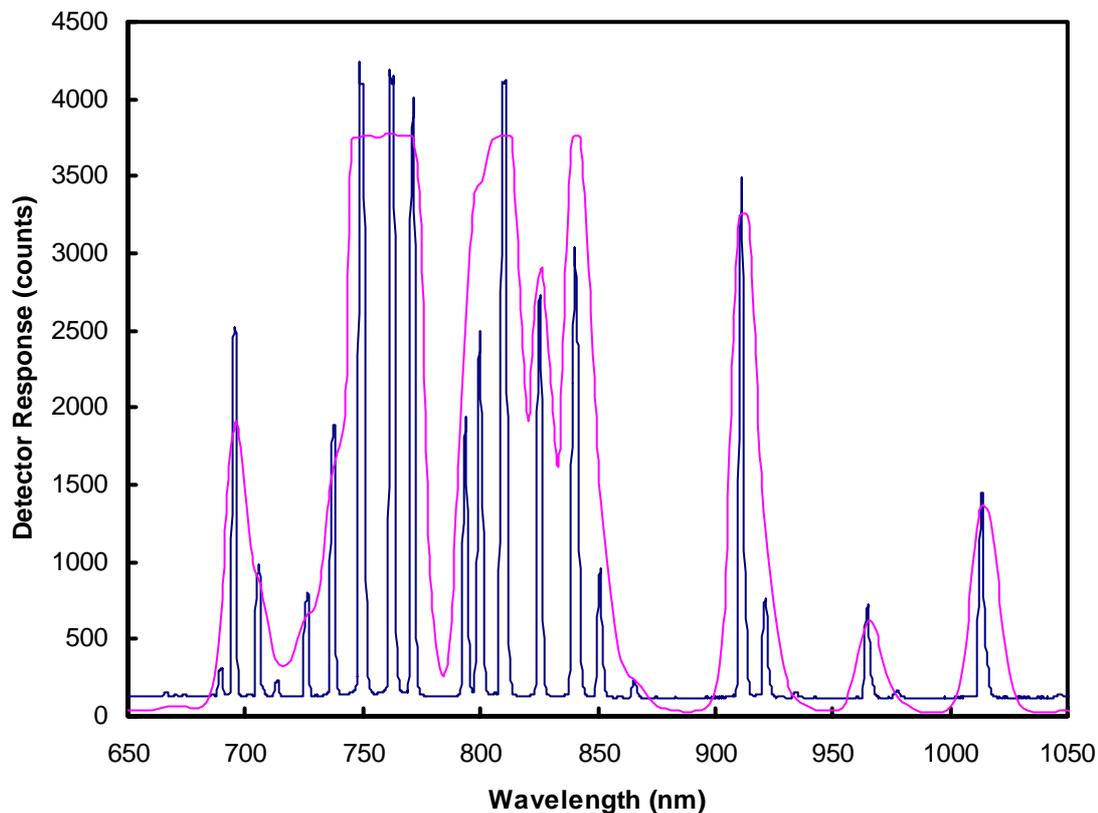
**Figure 1.** **Spectra of a mercury argon lamp acquired with the Zeiss MMS1 (dotted line) and the Ocean Optics S2000 (solid line) spectrometers. Inset illustrates the resolution of the 912 nm peak by the two devices (with detector response normalised to output at this wavelength).**

**Table 1.** **Calibration performance of (sucrose soaked cellulose) in terms of standard error of cross validation with respect to spectrometer wavelength resolution and signal to noise ratio. Means followed by the same letter within the two experiments are not significantly different at a 95% confidence level.**

| FWHM (912 nm) | Maximum Count | Spectra averaged | S/N | SECV (°Brix) |
|---|---|---|---|---|
| 7.7 | 30,300 | 1 | | 1.04 a |
| 10.6 | 30,300 | 1 | | 0.97 a |
| 13.8 | 30,300 | 1 | | 0.93 a |
| 16.7 | 30,300 | 1 | | 0.93 a |
| 20.0 | 30,300 | 1 | | 0.98 a |
| 13.2 | 2,000 | 1 | 1,400 | 2.02 a |
| 13.2 | 8,000 | 1 | 4,600 | 1.29 b |
| 13.2 | 30,300 | 1 | 9,700 | 1.22 b |
| 13.2 | 30,300 | 2 | 15,900 | 1.29 b |
| 13.2 | 30,300 | 16 | 30,300 | 1.46 b |

In practice, decreasing wavelength resolution (characterised at the 912 Hg-Ar line peak) to 20 nm did not significantly decrease the performance of a calibration of sugar solutions on cellulose (Table 1).  We conclude that spectral resolution below 20 nm is not a priority characteristic for instrumentation in this application.

## Signal to Noise Ratio

The signal to noise ratio of the MMS1 and S2000 instruments was characterised by collecting 200 spectra (raw A/D output, 12 bit A/D) of a teflon tile using the interactance probe and light source, and calculating a value for mean / standard error of measurement for every spectral data.  Light intensity was first adjusted to achieve a signal close to saturation for the two instruments.  The MMS1 demonstrated a relative enhancement in the 750-950 nm spectral region, relative to the S2000 (Fig. 2A).  The signal to standard error ratio broadly paralleled the mean signal for both instruments, reflecting the importance of signal shot noise (square root of number of photons received per pixel).  However, the ratio of signal to standard error of signal of the MMS1 reached a maximum of 40,000, in contrast to only 1,000 for the S2000 (Fig. 2A).  This result was expected, insomuch as photodiodes deliver a higher signal to noise ratio than CCD's at high signal levels.

A similar exercise was undertaken at a low light level, held constant for the two instruments.  As expected for a CCD detector relative to a PDA detector, the recorded count from the S2000 unit was greater than that of the MMS1 detector, although only by a factor of two (Fig. 2B).  This result reflects the wider slit width, greater pixel size and lesser pixel dispersion of the Zeiss MMS1 unit.  The signal to standard error ratio of the MMS1 was again higher than that of the S2000 (achieving a maximum of 7,000 and 150, respectively; Fig. 2B).  This result is contrary to that expected on the basis of PDA and CCD detectors type, and presumably reflects

differences in electronics between the two systems. Indeed, after initial powering up, detector output decreased slightly for the MMS1 (maximum at 750 nm, with 30 counts decrease on a signal of 30,000, or 0.1% change), stabilizing after 1.5 h (data not shown). However, the S2000 unit demonstrated greater fluctuations (ca. 1% change), with continuing fluctuation after 1.5 h (data not shown). Frequent referencing would be required for the latter unit in a fruit sorting application.
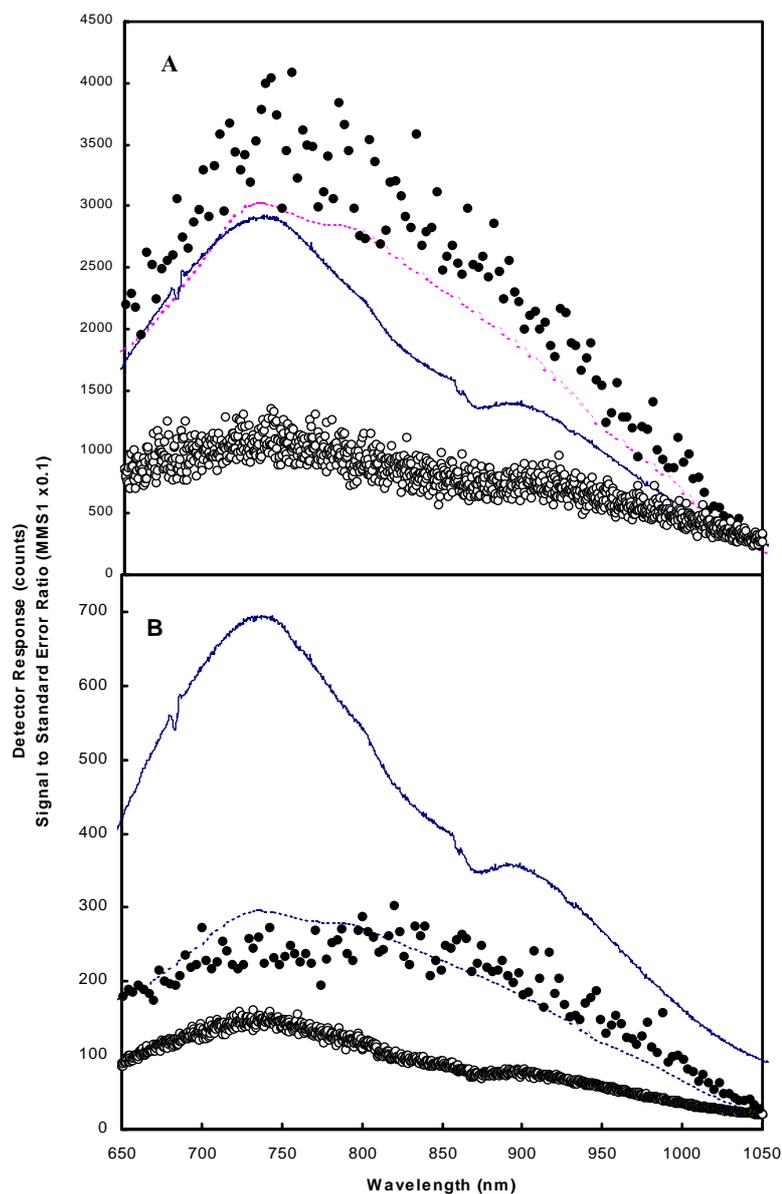
**Figure 2.** Relative spectral sensitivity (lines) and signal to standard error ratio (circles) of spectra collected using the MMS1 (dashed line, solid circle) and S2000 (solid line, open circle) spectrometers. Spectra were acquired using the same integration time (100 ms), light source, fibre optic guides and sample (reference material) for the two devices. Mean signal and mean signal divided by standard error of measurement at each wavelength (n= 50) are displayed. Note the scale change for the Zeiss MMS1 signal to noise ratio. (A) Light intensity was adjusted such that the output of each detector was near saturation, and normalised to output at 720 nm. (B) Spectra were acquired on both instruments at the same, relatively low, light intensity.

Another source of signal noise is variation in lamp intensity or spectral output.

After initial powering up, lamp output changed, with a decrease across most of the

227

spectrum (maximum of 200 counts on a signal of 30,000, or 0.67%), but an increase around 833 nm. Lamp output stabilised after ca. 1.5 h. These changes are ascribed to changes in lamp chemistry during lamp 'warm-up'. For all other experiments reported in this study, instrument and lamp were allowed to stabilise for at least 2 h before use.

The importance of signal to noise ratio on calibration was investigated by undertaking calibration of spectra with a range of signal to noise conditions, collected of cellulose soaked with a range of sucrose solutions. Signal to noise ratio was varied by changing signal level by altering light level and number of scans averaged per spectrum. The SECV of the resulting calibration was significantly affected only below a signal to standard deviation ratio of 4,600 (Table 1). We conclude that a single scan, with a count level at 25% or greater of saturation, is adequate for the task of sucrose calibration (0-20% w/v on cellulose matrix) using the MMS1.

On the basis of signal to noise ratio it is expected that the S2000 spectrometer would support a poorer calibration than the MMS1. Indeed, the SECV of a calibration of cellulose soaked with a range of TSS solutions was three times higher when developed with the S2000, in contrast to the MMS1 (5.4 and 1.8, respectively). The MMS1 was therefore adopted for the fruit assessment work reported below.

## Optimisation of an Optical Configuration

Light was diffusely scattered through the melon flesh (mesocarp), but assumed some directionality through the seed cavity of the fruit (Fig. 2). The contact angle of the light beam with the fruit surface was essentially irrelevant because of diffuse scattering of light within the fruit, with the detected light level determined by the distance from illuminated area to detected area. For convenience, however, lamp and detector were aligned with the centre of the fruit. There is a compromise position

between long path-length of light within the fruit and a high signal. A small angle between detector and lamp allows for a high signal (low noise) but gives a short pathlength in the fruit, with measurement of proportionally more non-edible parts of the fruit, i.e. 'skin'. A larger angle between detector and lamp provides a longer pathlength, representing more of the edible flesh of the fruit, but results in a low (noisy) signal. However, full transmission mode (lamp-detector angle of $180^{o}$) is undesirable on both counts (low signal, and measurement of seed cavity attributes as well as flesh attributes). Reasonable signal levels (ie >25% of detector saturation with a 200 ms integration time) were measured at up to a $60^{o}$ lamp-fruit-detector angle). With four lamps positioned at $90^{o}$ increments around the fruit, and at an angle of $45^{o}$ to the detector, a near saturation signal was achieved with an integration time of 200 ms. The choice of this angle was confirmed by the performance of calibrations developed of spectra collected at a range of lamp-fruit-detector calibrations (Table 2).
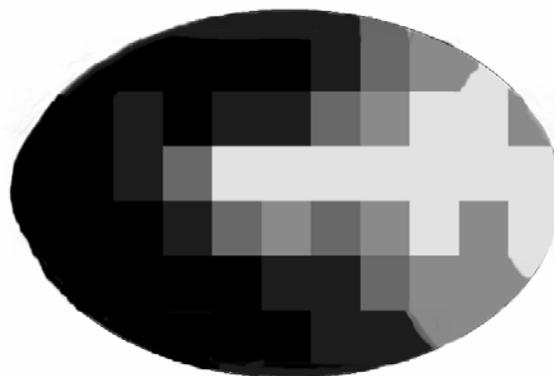


**Figure 3.      Two-dimensional diagrammatic representation (10 mm squares) of light penetration through a rockmelon from an incident light spot (on right side of 'fruit'). Data presented as absorbance units, within the following grey scales: (lightest to darkest) 0, 0.61, 1.22, 1.83, 2.44, 3.05 absorbance units.**

## Sampling and Soluble Sugar Content of Fruit

Soluble sugar content was variable within the melon fruit, varying with longitudinal position within the fruit by ca. 4% TSS, with circumferential position by

ca. 1% TSS, and with depth (skin to seed cavity) by ca. 4% TSS. This observation is consistent with the report of Peiris *et al.*[7]. Thus the four lamp configuration involves detection of light that has passed through volumes of fruit tissue which can be expected to vary in TSS. A variety of sampling procedures were assessed in terms of calibration performance (data not shown). The optimal method involved removal of 0.5 mm cores at a point between the centre of each of the four illuminated areas and the detected area and trimming of skin and seed cavity material from these cores, before pressing to extract juice for the refractometer measurement.

## Calibration Development and Robustness

A calibration developed on a single Doubloon population (n=200) failed to predict the TSS of other Doubloon populations, and of other varieties, as assessed by the standard error of cross validation (SECV) (Table 4). To develop a robust calibration, a data set of five Doubloon populations, varying in time and locality of harvest, and an extension of this data set involving a further five populations of four other melon varieties were created. Spectral 'redundancy' was reduced by assessing the influence of the Neighbourhood H criterion on SECV. A Neighbourhood H of 1.0 decreased the number of spectra from 1991 to 449 for the five Doubloon and five variety calibration sets, respectively (Table 3).

The calibration developed across five Doubloon populations predicted another Doubloon population well, and predicted population of one other variety well (Table 4). However, the performance of this calibration on two further melon varieties was less convincing (Table 4). The calibration developed across 10 populations of five varieties performed acceptably across all conditions. These results indicate that a calibration can be relatively robust across varieties, growing region and time.

**Table 2.**      Calibration performance (melon sugar content) with respect to the angle between the illuminated and detected areas of the fruit, with reference to the centre of the fruit.  Calibration population n=40, range 7.0-11.9, mean 9.5° Brix.

| Lamp angle (°) | $R^2$ | SECV | SEC |
|---|---|---|---|
| | | (°Brix) | |
| 20 | 0.19 | 1.11 | 0.97 |
| 40 | 0.64 | 1.24 | 0.65 |
| 60 | 0.82 | 0.84 | 0.43 |
| 80 | 0.38 | 1.03 | 0.84 |

**Table 3.**      Effect of population size reduction using Neighbourhood 'H' (NH) criterion on calibration performance (melon sugar content).

| NH | Population numbers | SEP(C) (°Brix) |
|---|---|---|
| 0.2 | 1991 | 0.85 |
| 0.4 | 1458 | 0.90 |
| 0.6 | 984 | 0.82 |
| 0.8 | 647 | 0.87 |
| 1.0 | 449 | 0.69 |
| 1.2 | 303 | 0.77 |
| 1.4 | 232 | 0.84 |

**Table 4.** Performance of a calibration developed on (a) one population (200 spectra) of fruit of variety Doubloon (DubA), (b) five populations (1000 spectra) of fruit of variety Doubloon (DubA-E) and (c) ten populations (2,000 spectra) of five varieties of melon (5var) on the prediction of melon sweetness. Calibration groups for (b) and (c) were selected using a criterion of 1.0 NH. Results marked with an * represent a SECV (population data included in the calibration set), while unmarked results represent a true standard error of prediction (SEP).

| Validation group | SECV/SEP (°Brix) | | |
|---|---|---|---|
| | Calibration groups | | |
| | DubA | DubA-E | 5 var. |
| Doubloon A | 0.53* | 0.62* | 0.72* |
| Doubloon B | 1.33 | 0.86* | 0.93* |
| Doubloon C | 1.28 | 0.66* | 0.75* |
| Doubloon D | 1.17 | 0.74* | 0.79* |
| Doubloon E | 1.13 | 0.92* | 1.03* |
| Doubloon F | 0.93 | 0.66 | 0.67 |
| Doubloon A-E | 1.42 | 0.76* | 0.75* |
| Eastern star | 1.13 | 1.11 | 0.84* |
| Hammersley | 1.13 | 0.92 | 1.03* |
| Highline | 0.93 | 1.16 | 0.70* |
| Malibu | 0.89 | 0.66 | 0.61* |

## ACKNOWLEDGEMENTS

## REFERENCES

1.      L.L. Mutton, B.R. Cullis and A.B. Blakeney, *J. Sci. Food Agric.* **32**, 385 (1981).

2.      G.G. Dull, G.S. Birth, D. Smittle and R.G. Leffler, *J. Food Sci.* **54**, 393 (1989).

3.      G.G. Dull, R.G. Leffler, G.S. Birth, and D. Smittle, *Transaction ASAE* **35**, 735 (1992).

4.      G.S. Birth, G.G. Dull and R.G. Leffler, *Optics in Agric.* **1379**, 10 (1990).

5.    H. Aoki, Y. Matumoto, K. Mizuno and H. Maeda, *Proceedings: Conference on Tropical Fruits*, Kuala Lumpa, Malayasia. Malaysian Agricultural Research and Development Institute, 207 (1996).

6.    J. Guthrie, B. Wedding and K. Walsh, *J. Near Infrared Spectrosc.* **6**, 259 (1998).

7.    K.H.S. Peiris, G.G. Dull, R.G. Leffler, and S.J. Kays, *HortScience* **34**, 114 (1999)

# DETECTION OF INSECT DAMAGE IN INTACT SWEET CORN COBS USING NIR REFLECTANCE SPECTROSCOPY[9]

## ABSTRACT

Discrimination of the defects of sweet corn cobs (poor grain tip fill, insect damage and presence of the heliothis grub) from clean cobs can be achieved non–invasively by near infra-red spectroscopy. The NIRSystems 6500 scanning spectrometer and the Zeiss MMSI miniature spectrometer (coupled with a high intensity quartz halogen lamp -100 watts) can achieve this discrimination. However, in a practical packing shed operation the relatively rapid and inexpensive Zeiss MMSI spectrometer offers the most promise.

## INTRODUCTION

Near infra-red reflectance spectroscopy (NIRS) is a non-destructive procedure that uses optical data rather than wet chemistry methods to analyse both liquid and solid products for chemical composition. Near infra-red reflectance spectroscopy has been used for over twenty years to analyse grain products for protein, oil and moisture (Shenk and Westerhaus 1993). The use of the technique however has been limited to low moisture materials, as water absorbs strongly in the near infra-red (NIR) region of the electromagnetic spectrum of radiation. The advent of powerful personal computers, fibre optics, improved sensor technology and chemometric software

---

[9] This appendix has been published in the Final Report for the project titled 'Insect Management in Sweet Corn' – VG97036 authored by Peter Deuter *et al*. (2002) and funded by Horticultural Australia Limited.

packages has allowed this technology to be applied to high moisture materials (such as fresh whole fruit) in an in-line situation, in the last decade.

Near infra-red is a small part of the electromagnetic spectrum of radiation (700 – 2,500 nm). At one end of this spectrum are the high energy waves such as x-rays and gamma rays, while at the other end of the spectrum are the low energy waves such as micro waves and radio waves. Near infra-red is between the visible and the infra-red regions of the spectrum. The area of the electro-magnetic spectrum of interest is between 700 and 2500 nm and concerns the bending and stretching of electronic bonds (C-H, N-H and O-H). These bonds are involved in most organic compounds, such as sugars, protein, lipids and water. Near infra-red spectroscopy is a secondary method of measurement and so must be calibrated against a primary or reference method, such as a refractometer reading (°Brix). Because of this requirement, the technique is only cost effective with large sample numbers. The calibration may be established in either a quantitative (e.g. linear regression of Brix content) or qualitative (e.g. discrimination between groups) basis.

Ridgway and Chambers (1996) and Chambers and Ridgway (1995) used NIR reflectance spectroscopy to detect external and internal insect (grain weevil) infestation of intact wheat kernels. Large spectral differences were observed between non-infested kernels and kernels infested internally with *Sitophilus granarius* (L) (grain weevil) larvae or pupae, arising from both a changed chemical composition and physical structure. Single non-infested and infested kernels were distinguished by their second derivative ($d^2$) spectra. For both external and internal infestation there was substantial evidence that insect protein and/or chitin and moisture were being detected. Near infra-red spectroscopy should be useful as a rapid method of detection. Further work by Chambers and Ridgway (1998) with single wheat grain kernels

internally infested with pupal stages of the grain weevil showed the possibility of detecting such infestation by measuring just two NIR wavelengths (1202 and 1300 nm). These workers used the 1100-2500 nm region of the spectrum and discriminated infested from non-infested samples simply by the increased reflectance (decrease in absorbance) of infested kernels due to the increase in specular radiation from the internal cavities (as a result of insect feeding ) and from the insect itself.

Near infra-red spectroscopy (NIRS) (700 to 2500 nm), concerns the bending and stretching of molecular bonds (C-H, N-H and O-H) involved in most organic compounds. NIRS is a secondary method of measurement and must be calibrated against a primary or reference method, such as a refractometer for Brix.

Ridgway and Chambers (1995, 1996) used NIR reflectance spectroscopy (second derivative absorbance) to detect external and internal insect *Sitophilus granarius* (L) (grain weevil) infestation of intact wheat kernels arising from changes in both chemical composition and physical structure. Infested samples demonstrated increased reflectance (decrease in absorbance) due to the increase in specular radiation from the internal cavities (as a result of insect feeding). There was substantial evidence that insect protein and/or chitin and moisture were being detected. Further work by Chambers and Ridgway (1998) with single wheat grain kernels demonstrated detection of infested grains using just two NIR wavelengths (1202 and 1300 nm).

In this study the potential of NIRS for the detection of grub damage, grub damage with grub present and poor grain tip fill in corn was assessed

## MATERIALS AND METHODS

Sweet corn cobs fruit were harvested from two locations in Queensland, namely the Lockyer Valley and the Burdekin irrigation areas.  Cobs were selected to include clean cobs, cobs with poor grain tip fill, cobs with grub damage and cobs with grubs present.  Sweet corn cobs were transported under dry ice (8-10° C on arrival) and assessed within three days of harvest.  Near infra-red spectra were collected from the cob tassel end (first 7 cm of tip) of individual cobs.  Spectra were collected from two commercially available research instruments, the NIRSystems 6500 (700 – 2500 nm, remote reflectance probe), and the Perten DA 7000 (700 – 1700 nm, interactance probe), and a purpose built unit based on the Zeiss MMS1 miniature spectrometer (700 – 1050 nm).  Various optical configurations were used to gather spectra from the sweet corn cobs (Figs. 1).  Integration time for spectral acquisition varied from 100 to 160 milliseconds per cob (four scans per spectra), to maximise the signal to noise for the Zeiss unit.  Discriminant equations were developed using partial least squares regression analysis within the WinISI II (vers. 1.02a) chemometric package.

Four experimental runs were undertaken over the season using various spectrophotometers and optical configurations (Table 1 and Fig. 1).  In run four, the tip (tassel) of the corn cob was halved longitudinally and presented to the instrument sheath uppermost.  In this experiment, spectra were acquired of 80 cobs (20 of each category), of which five were randomly selected from each of the four groups for validation of the discriminant equation (developed on the remaining set).  The spectral data was analysed using the discriminate function of the WinISI II vers. 1.02a chemometric software package (Table 2 and 3).  In the discriminant analysis, spectral data were pre-treated with regard to derivatives, smoothing and scatter correction.

# RESULTS AND DISCUSSION

According to Shenk and Westerhaus (1993), discriminant analysis is best undertaken with no scatter correction (particle size and scattering of light may assist in sample discrimination) and first derivative to eliminate base line error. A mathematically pre-treatment of first derivative derived over four data points with no smoothing, was found to give the optimum results and subsequently used.

The Perten DA 7000 spectrophotometer was unable to discriminate between groups although it operates with a high intensity tungsten halogen lamp (42 watt) and covers an area of the electromagnetic spectrum from 700-1700 nm, using both silicon (Si) and (InGaAs) photodiodes. This unit was operated in the interactance mode (Fig. 1(a)) and as such, the area viewed by the probe was relatively small (e.g. less than 10% of the area viewed by the remote reflectance fibre optic probe of the NIRSystems 6500). The bifurcated fibre optic bundle carries incident light down the outside bundle and the reflected light from the sample back to the detectors through the centre fibres. However, the high intensity light and interactance mode should have resulted in good light penetration of the sample. The poor result cannot be explained except by the fact that the instrument was on loan for a short time period and was possibly not set up optimally with regard to integration time. Possibly, the interactance mode was not the ideal configuration, or gathering information over a very small area of sample with the potential to miss the localised defect.

The experimental run utilising the Zeiss MMSI spectrophotometer in a side and tip presentation of the sweet corn cobs (Table 1 and Fig. 1(b &c)), again showed an inability to discriminate between the various groups of sweet corn cobs (clean, damaged, poor grain tip fill and grub presence). Both these optical configurations utilised the transmission mode. Because of the difficulty in sealing the shroud

containing the fibre optic bundle (carrying transmitted light to the detector array), the poor result could be best explained by the intrusion of excessive specular radiation. Also most of the light reaching the detector would have arisen from the sheath with little useful information regarding the actual defect.

The third experimental run utilised a high intensity quartz halogen car lamp (100 watt) with a parabolic reflector to deliver light to the object, operated in the transmission mode (Fig. 1e).  Again poor discrimination occurred and this could be attributed to the transmission mode reducing the proportion of the signal containing defect information reaching the detectors.  In the fourth experiment, discrimination was achieved using the same configuration but reduced sample thickness.

In the fourth experimental run, good discrimination also was achieved (Table 2 and 3, Fig.1 (d & f)) with the NIRSystems 6500.  In this run, the transmission mode was utilised but only half of the tip of the cob was viewed.  The slightly better results of the 6500 could be attributed to the operation of the remote reflectance fibre optic probe in a light proof box (no ambient light adding to the signal) and the general precision of the instrument compared to the photodiode array of the Zeiss. Discriminant analysis for the 6500 as reported in Table 2 and 3, was carried out on the full spectrum (700-2,300 nm).  However, separate analysis also was undertaken on 700-1,100 nm (as with the Zeiss) and 1,100-2,300 nm areas of the spectrum.  These results demonstrated the region of 700-1100 nm gave better results than the 1,100-2,300 nm region alone.

In a plot of the first derivative data obtained form both the Zeiss and 6500 instruments (experiment 4) the areas of the spectrum showing most divergence, occurred around 960 and 1,030 nm.  These areas could be attributed to water and protein, respectively.

Clean cobs were always distinguished (by both instruments) from damaged ones. However, there was some confusion with poor tip fill cobs sometimes misdiagnosed as insect damaged. Also, the discriminant equation could not always differentiate between damaged and damaged with grub. With both the Zeiss and 6500 (experiment 4), in distinguishing the groups and various combinations of the groups (Table 2 and 3) on no occasion was a defect cob (i.e. damaged, poor tip fill or grub present) included in the clean group.

**Table 1.** **Discriminant analysis of the spectral data obtained from the various instruments and configurations, using the chemometric package WinISI vers. 1.02a. Inability to distinguish between groups occurred when < 20% correctly identified.**

| Experiment | | No. of spectra | Instrument | Optical configuration | Groups distinguished |
|---|---|---|---|---|---|
| No. | Date | | | | |
| 1 | 18-09-98 | 198 | Perten DA7000 | Interactance (42 watt) | No |
| 2 | 24-11-98 | 410 | Zeiss MMSI | side –transmission (50 watt) | No |
| | | 410 | Zeiss MMSI | tip –transmission (50 watt) | No |
| 3 | 25-11-98 | 106 | Zeiss MMSI (700-1100 nm) | large lamp – transmission (100 watt) | No |
| 4 | 25-02-99 | 80 | 6500 (700-2300 nm) | Reflectance (75 watt) | Yes |
| | | 80 | Zeiss MMSI | Half cob tip – transmission (100 watt) | Yes |

**Table 2.** **Discriminant analysis of sweet corn cobs into 4 groups using partial least squares regression (WinISI vers. 1.02a).**

| Groups | Instrument | Diagnosis (out of 5) | Correct (of diagnosed) | Incorrect diagnosis |
|---|---|---|---|---|
| **1 Tip fill (T)** | 6500 | 5 | **3** | 1D, 1G |
| | Zeiss | 5 | **4** | 1G |
| **2 Clean (C)** | 6500 | 5 | **5** | |
| | Zeiss | 6 | **5** | 1T |
| **3 Damaged (D)** | 6500 | 5 | **3** | 2T |
| | Zeiss | 3 | **2** | 1G |
| **4 Damaged + grub (G)** | 6500 | 5 | **4** | 1D |
| | Zeiss | 6 | **3** | 3D |

The discrimination was undertaken on four groups of sweet corn cobs with 15 samples in each calibration set and 5 samples in each validation set.  The number of samples from the validation set (Diagnosis column) correctly assessed is given under the heading 'Correct out of diagnosis' and the misdiagnosed in the final column.

**Table 3.** Discriminant analysis of sweet corn cobs into 2 groups, using partial least squares regression (WinISI II vers. 1.02a) analysis. Data of Table 2 (re-categorised).

| Groups | Instrument | No. calibration set | No. validation set | Diagnosis | Correct | Incorrect |
|---|---|---|---|---|---|---|
| **1. Clean vs Damaged** | 6500 | 30 | 5C,5D | 10 | **10** (5C,5D) | |
| | Zeiss | 30 | 5C,5D | 10 | **10** (5C,5D) | |
| **2. Clean vs Damaged with grub** | 6500 | 30 | 5C,5G | 10 | **10** (5C,5G) | |
| | Zeiss | 30 | 5C,5G | 10 | **10** (5C,5G) | |
| **3. Clean vs Tip Fill** | 6500 | 30 | 5C,5T | 10 | **10** (5C,5T) | |
| | Zeiss | 30 | 5C,5T | 10 | **9** (4C,5T) | 1 T |
| **4. Clean vs Tip fill, damaged, grub** | 6500 | 45 | 15TDG,5C | 20 | **20** (15TDG,5C) | |
| | Zeiss | 45 | 15TDG,5C | 20 | **18** (15TDG,3C) | 2 TDG |
| **5. Tip Fill & Clean vs Damage & damage + grub** | 6500 | 60 | 10DG,10CT | 20 | **17** (10DG,7CT) | 3 DG |
| | Zeiss | 60 | 10DG,10CT | 20 | **19** (10DG,9CT) | 1 DG |
| **6. Clean vs Damaged & Damaged + Grub** | 6500 | 45 | 5C,10DG | 15 | **15** (10DG,5C) | |
| | Zeiss | 45 | 5C,10DG | 15 | **14** (11DG,4C) | 1 DG |
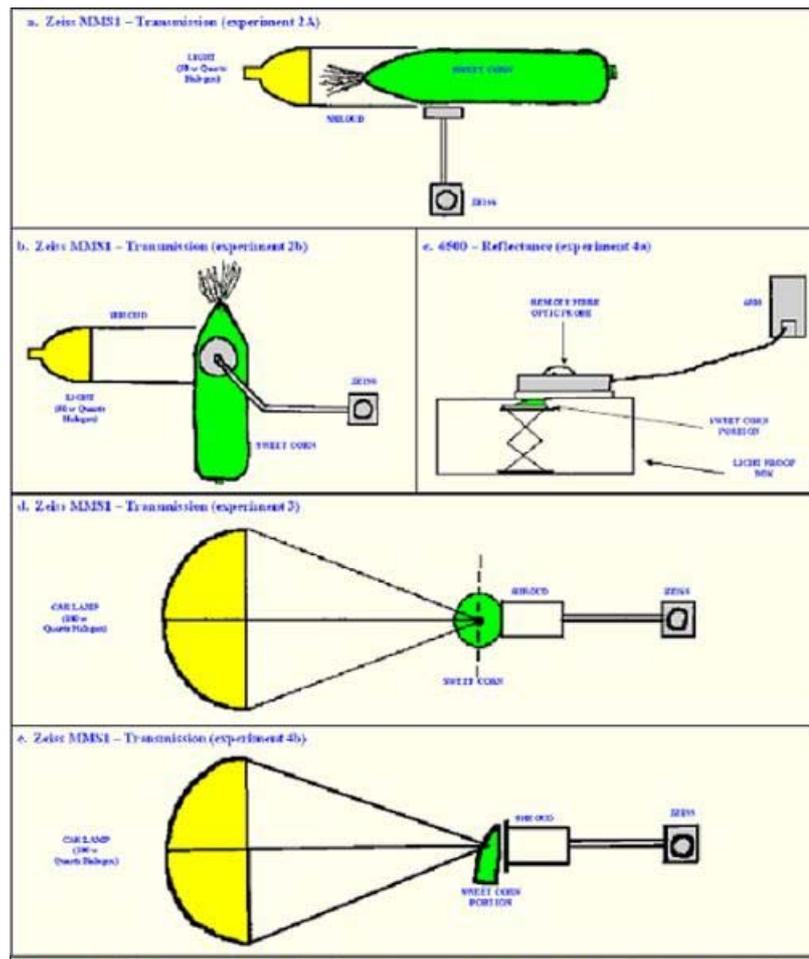
**Figure 1.** Schematics of optical configurations used in near infra-red spectroscopy of sweet corn groups.

## CONCLUSION

Discrimination of the defects of sweet corn cobs (poor grain tip fill, insect damage and presence of the heliothis grub) from clean cobs can be achieved non–invasively by near infra-red spectroscopy. The NIRSystems 6500 scanning spectrometer and the Zeiss MMSI miniature spectrometer (coupled with a high intensity quartz halogen lamp -100 watt) can achieve this discrimination. However, in a practical packing shed operation the relatively rapid and inexpensive Zeiss MMSI spectrometer offers the most promise. Further work needs to be undertaken to improve the signal to noise ratio of the optical configurations. This will involve

further experimenting with optical configurations, integration times for spectral acquisition and light intensities.

## ACKNOWLEDGEMENTS

## REFERENCES

Chambers, J. and Ridgway, C. (1995) *Near Infrared Spectroscopy: The Future Waves*. Eds. A.M.C. Davies and Phil Williams, NIR Publications, Chichester, UK, 484-489.

Ridgway, C. and Chambers, J. (1996) Detection of External and Internal Insect Infestation in Wheat by Near-Infrared Reflectance Spectroscopy. *J.Sci. Food Agric.* **71**, 251-264.

Ridgway, C. and Chambers, J. (1998) Detection of insects inside wheat kernels by NIR imaging. *J. Near Infrared Spectrosc*. **6**, 115-119.

Shenk, J.S. and Westerhaus, M.O. (1993) *In:NIRS Handbook, Infrasoft International*, marketed by NIRSystems, Inc. Silver Springs, MD, USA, p. 14).

Shenk, J.S. and Westerhaus, M.O. (1993) *In:NIRS Handbook, Infrasoft International*, marketed by NIRSystems, Inc. Silver Springs, MD, USA, p. 14).

## ASSESSING AND ENHANCING NEAR INFRARED CALIBRATION ROBUSTNESS FOR SOLUBLE SOLIDS CONTENT IN MANDARIN FRUIT [10]

### ABSTRACT

Near infra-red (NIR) spectroscopy has been used for the non-invasive assessment of intact fruit for eating quality attributes such as total soluble solids (TSS) content. However, little information is available in the literature with respect to the robustness of such calibration models validated against independent populations (however, see Peiris *et al.* 1998 and Guthrie *et al.* 1998). Many studies report 'prediction' statistics in which the calibration and prediction sets are subsets of the same population (e.g. a three year calibration validated against a set from the same population, Peiris *et al.* 1998; calibration and validation subsets of the same initial population, Guthrie and Walsh 1997 and McGlone and Kawano 1998). In this study, a calibration was developed across 84 melon fruit ($R^2 = 0.86°$ Brix, $SECV = 0.38°$ Brix), which predicted well on fruit excluded from the calibration set but taken from the same population (n = 24, $SEP = 0.38°$ Brix with 0.1° Brix *bias*), relative to an independent group (same variety and farm but different harvest date)

---

(n = 24, *SEP*= 0.66° Brix with 0.1° Brix *bias*).  Prediction on a different variety,

different growing district and time was worse (n = 24, *SEP* = 1.2° Brix with 0.9° Brix

*bias)*.

Using an 'in-line' unit based on a silicon diode array spectrometer, as described

in Walsh *et al.* (2000), we collected spectra from fruit populations covering different

varieties, growing districts and time.  The calibration procedure was optimised in

terms of spectral window, derivative function and scatter correction.  Performance of

a calibration across new populations of fruit (different varieties, growing districts and

harvest date) is reported.  Various calibration sample selection techniques (primarily

based on Mahalanobis distances), were trialled to structure the calibration population

to improve robustness of prediction on independent sets.  Optimisation of calibration

population structure (using the ISI protocols of neighbourhood and global distances)

resulted in the elimination of over 50% of the initial data set.  The use of the ISI Local

Calibration routine was also investigated.

**Additional keywords**: acidity, Brix, citrus, dry matter content, non-destructive,

section dryness


# INTRODUCTION

Near infrared (NIR) spectroscopy has been applied to the sorting of intact fruit

with a high moisture content for constituents such as soluble solids content (SSC) in

cantaloupe fruit,[1] sugar content in intact peaches,[2] sugar content, acidity and hardness

of intact plum fruit[3] and SCC of intact citrus (mandarin fruit).  Commercial

application to pack-house fruit sorting lines commenced in Japan in the mid 1990s,

for the sorting of sweetness, ripeness and acidity of citrus fruit, apples, pears and

peaches at three pieces per second per lane.[4, 5] Commercial application within pack-houses of Western countries is nascent.

The application of NIR technology requires an appreciation of the distribution of the character of interest within the fruit and the absorption and scattering of light through the fruit, in order to design an appropriate optical configuration of light source, detector and fruit (for example References 6 and 7). The robustness of the NIR calibration model must be assessed across populations of fruit differing in, for example, temperature, variety and growing district. Unfortunately, these parameters are not well reported in the literature, with many NIR studies reporting the use of a standard optical design for spectral acquisition and the use of a single harvest population, divided into a calibration set and a validation set. Few studies have explored the issue of validation across populations varying in the locality of harvest, the time of harvest with a given season, or across years. A notable exception is that of Peiris *et al.*[8] who reported calibration validation across three seasons for peaches. A calibration developed in one year predicted poorly on other years, but a combined calibration performed well for validation groups drawn from those years. In the current study we report on issues related to calibration robustness for intact mandarin fruit assessed for SSC.

## MATERIALS AND METHODS

### Plant material and SSC analysis

Imperial variety of mandarin were sourced from commercial orchards in Munduberra, Queensland. Fruit were sourced from three separate farms on one day, from three separate harvests over a five-day period from one tree and from one

packhouse over three seasons. Fruit were halved, juiced and SSC determined by refractometry (Bellingham and Stanley RMF 320).

### *Spectroscopy*

Spectra were collected using an NIR enhanced Zeiss MMS1 spectrometer and a tungsten halogen light in the optical configuration reported by Greensill and Walsh.[7] Spectra were collected from one side of each fruit, on the equator of the fruit, equidistant from pedicel and stylar ends.

### *Chemometrics*

The software package WinISI (ver.1.04a) was used for all chemometric analysis. Calibration performance was assessed in terms of coefficient of determination ($R^2$) standard error of prediction (*SEP*), variance ratio (*1-VR*), standard deviation ratio (*SDR*), slope and bias of the validation sets. Further, the criteria of Wortel *et al*.,[9] based on the Taguchi concepts as used in process control, were applied to evaluate model robustness. This approach involved calculation of an average *SEP* and a signal to-noise statistic (s/n = 20 log10 [mean *SEP* / *SD SEP*]) for the performance of a given model across a range of validation sets.

**Table 1.** Calibration and validation statistics for a calibration on one population of mandarin SSC, used in prediction of three populations varying in (a) days of harvest, (b) location of harvest and (c) season of harvest.

| Fruit population | SD | $R^2$ | SECV/SEP | BIAS |
|---|---|---|---|---|
| Time | | | | |
| Cal | 0.95 | 0.90 | 0.35 | |
| Val | | | | |
| Day 1 | 0.73 | 0.68 | 0.48 | 0.173 |
| Day 3 | 0.72 | 0.71 | 0.52 | −0.352 |
| Day 5 | 0.68 | 0.55 | 0.52 | 0.209 |
| s/n | | | 26.8 | |
| Av | | | 0.51 | |
| Location | | | | |
| Cal | 0.85 | 0.87 | 0.353 | |
| Val | | | | |
| A | 0.51 | 0.55 | 0.37 | −0.12 |
| B | 0.57 | 0.69 | 0.41 | 0.25 |
| C | 0.50 | 0.55 | 0.52 | 0.40 |
| s/n | | | 14.93 | |
| Av | | | 0.43 | |
| Seasons | | | | |
| Cal | 0.95 | 0.84 | 0.42 | |
| Val | | | | |
| Year 1 | 0.96 | 0.83 | 0.49 | 0.24 |
| Year 2 | 1.05 | 0.31 | 2.45 | 0.40 |
| Year 3 | 1.05 | 0.82 | 3.76 | 3.73 |
| S/N SEP | | | 2.65 | |
| Av SEP | | | 2.23 | |

# RESULTS AND DISCUSSION

## Calibration statistics and B coefficients

Typical MPLS calibration statistics for intact mandarin SSC were: R2 0.87, SECV 0.35, using six principal components, on a population SD 0.85, n = 100 (Table 1). The MPLS B coefficients for the mandarin SSC calibrations contain negative weightings on second derivative spectra around 910 and 850 nm and positive weightings around 880 nm (data not shown). Absorbance at ca 910 nm is ascribed to a third overtone stretching of CH bonds (Golic and Walsh, this volume). Absorbance at 880 nm may convey pathlength information. A calibration that does not contain spectroscopically 'relevant' information is likely to be over-fitted to the data and, thus, can be expected to perform poorly when applied to new validation populations.

A calibration developed from a single population of fruit (100 spectra) was applied to validation sets harvested on different days, different locations and different growing seasons (Table 1). The cause of the decrease in performance of a calibration when applied to a 'new' group presumably reflects change in the physical (optical) properties or the chemical properties (acid, water content) of the fruit. Temperature of the fruit was constant at scanning. Calibration performance across harvest day and location was comparable, as indicated by the mean *SEP* and s/n statistic, while performance was dramatically degraded across seasons. There was no clear trend for performance to degrade with increasing time (days) or distance/soil type (location, data not shown) of harvest relative to the calibration group. Performance was apparently degraded with increasing time (seasons). The cause of the dramatic decrease in performance of a calibration when applied to a new season of fruit is not clear and could reflect changes in the instrument used as well as change in the sample (fruit).

To improve calibration performance on a new validation set, a typical strategy involved addition of samples from the new set to the calibration group. The validation sets were divided into two equal groups. One group was retained as a validation set and the other group used for selection of samples for addition to the calibration set. Any validation sample with a GH > 3.0 (calculated on calibration set scores and loadings) was excluded from this process. Several approaches were used in the selection of samples from the validation group for addition to the calibration group, (1) random, (2) selection, on the basis of ascending GH (validation set ordered in ascending order of GH calculated on calibration set scores and loadings, and samples selected at equal GH intervals), (3) selection of the basis of spaced GH (calculated as

per 2) and (4) selection on the basis of NH (increasing NH values calculated on calibration set scores and loadings to select increasing numbers of validation set samples, using the ISI 'Expand a Product File with New Spectra' feature). The performance of a calibration developed in one growing season and applied to fruit of a subsequent season was improved in terms of *SEP* and bias as increasing numbers, up to *ca* 10, of 'validation set' samples were added to the calibration set, using any of the three selection approaches (Figure 1).
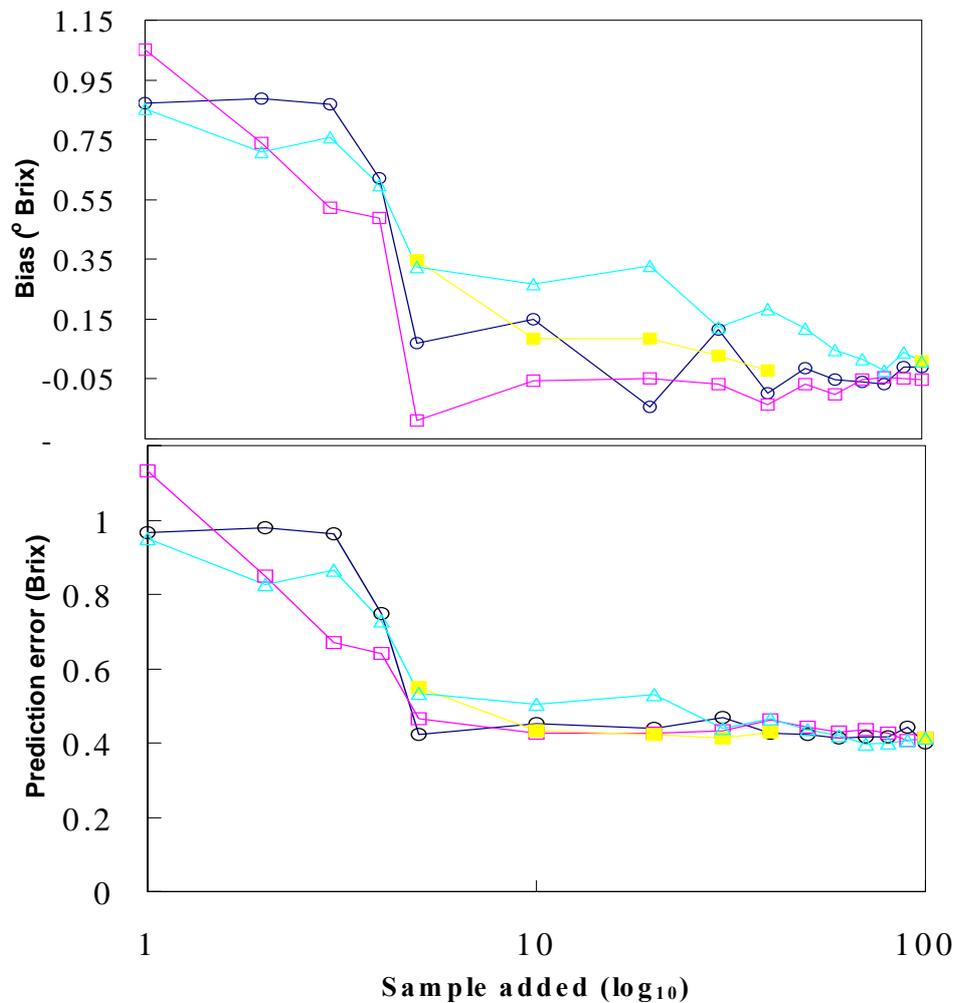
**Figure 1.** Prediction statistics for SSC of a mandarin validation population (different growing season to calibration population) using three treatments for sample selection from the new season group for addition to the calibration group. Open circle, random selection; open squares, central GH selection; closed squares, spaced GH selection; open triangle, NH selection.

It is surprising that so few fruit were representative of any physical or chemical change in the validation, relative to the calibration, set. In practical terms, we recommend it is sufficient to add data of *ca* 15 fruit to a calibration to update it for use across growing seasons.

## ACKNOWLEDGEMENTS

## REFERENCES

1. G. Dull, G. Birth, D. Smittle and R. Leffler, *J. Food Sci.* **54,** 393 (1989).

2. S. Kawano, H. Watanabe and M. Iwamoto, *J. Jpn Soc. Hort. Sci.* **61,** 445 (1991).

3. T. Onda,M. Tsuji and Y. Komiyama, *Journal of the Japanese Society for Food Science and Technology* **41,** 908 (1994).

4. S. Kawano, T. Fujiwara, and M. Iwamoto, *J. Jpn Soc. Hort. Sci.* **62,** 465 (1993).

5. S. Kawano, *NIR news* **5,** 10 (1994).

6. K.B. Walsh, J.A. Guthrie and J. Burney, *Aust. J. Plant Physiol.* **27,** 1175 (2000).

7. C.V. Greensill and K.B. Walsh, *Meas. Sci. Technol*. **11,** 1674 (2000).

8. K.H.S. Peiris, G.G. Dull, R.G. Leffler and S.J. Kays, *J. Am. Soc. Hort. Sci*. **123,** 898 (1998).

9. V.A.L. Wortel, W.G. Hansen and S.C.C. Wiedemann, in *Near Infrared Spectroscopy: Proceedings of 9th International Conference*, Ed by A.M.C. Davies and R. Giangiacomo. NIR Publications, Chichester, West Sussex, UK, p. 267 (2000).

# Optimising mathematical treatments using significance [11] testing

## J. A. Guthrie[1,2], D.J. Reid[2] and K.B. Walsh[1]

[1] Plant Sciences Group, Central Queensland University, Rockhampton, 4702, Australia
[2] Agency for Food and Fibre Sciences, Food Technology, Queensland Department of Primary Industries, PO Box 6014 CQ Mail Centre, Rockhampton 4702.

A Microsoft® Excel spreadsheet was developed to 'automate' the significance testing of *RMSEP* and bias, following the method reported by Fearn (1996). The method tests the significance of the differences in standard deviation (*RMSEP*) and bias between two models based on the residuals for the reference and predicted values for each sample in the population. For bias, the null hypothesis is that biases from both models are equal; for the standard deviation, the null hypothesis is that the ratio of standard deviations is unity. The user chooses the level of significance required (e.g. 1%, 5% or 10%) with a default confidence level of 5%.

The method requires two sets of matching predicted values and a set of corresponding reference values (Fig. 1). Thus at any one time only two models (e.g. 2nd derivative absorbance data with or without scatter correction) can be compared.

Fig. 1. Data entry worksheet



A summary report (Fig. 2) includes the mean, number of samples (N) and standard deviation (SD) for the reference (Actual) values, predictions (Pred1 and Pred2) and residuals (m1 and m2) from both calibration models and the difference in residuals (d).

Fig. 2. Summary report worksheet



In this example, there was no significant difference between the bias values of the two predictions, but the RMSEP values were significantly different.

A disadvantage of the procedure is that only two calibrations can be compared at a time (similar to a *t* test). We look forward to the development of the equivalent of an ANOVA test (to compare multiple calibrations simultaneously), ideally working with unequal sample numbers! Any takers?

Disk copies of the spreadsheet are available. Alternatively copies can be sourced from john.guthrie@dpi.qld.gov.au.
**References.**

Fearn, T. (1996). Comparing Standard Deviations. NIR News, Vol. 7 No. 5, p. 5-6.

---

[11]This appendix was presented at the 11th Australian Near Infra Red Spectroscopy Group Conference and Short Courses, Fremantle, WA, 18-21 April, 2004. Authors were Guthrie J, Reid D, Walsh K and was titled 'Optimising math treatments using significance testing'.

# COMPARING TWO COMPETING CALIBRATIONS FOR A GIVEN DATA SET[12]

## Comparing Two Competing Calibrations for a Given Data Set
### Using Fearn's Criteria
### Version 2.0

**David J. Reid and John A. Guthrie**
**Queensland Department of Primary Industries and Fisheries**
**PO Box 6014, Rockhampton Mail Centre, 4702**
**Rockhampton, Queensland, Australia.**

We have developed an Excel spreadsheet to optimise the various data treatments used in calibration development (e.g. scatter correction, maths treatment and wavelength window) that follows the method reported by Tom Fearn (NIR News 7(5):5-6, 1996; see also Reid and Guthrie (*The Nir Spectrum 1(3):7, 2003*. This method tests the significance of the differences in standard deviation (RMSEP) and bias between two models based on the residuals for the reference and predicted values for each sample in the population. Terminology used in this spreadsheet is consistent with that used in Fearn's article.

### Instructions

1. Use your chemometric software package (e.g. WinISI or The Unscrambler), to derive the predicted values from your two competing calibration models (calibration 1 and calibration 2, based on two different data treatments). Extract the reference (Actual) and predicted values (Pred1 and Pred2) for each sample in the population and paste into the appropriate columns in the worksheet 'Data'. These correspond to the columns in worksheet 'Data' with headings highlighted in yellow. There is a limit of 1000 samples. At any one time you can compare only two models (e.g. 2nd derivative absorbance data with or without scatter correction). When copying new data to the spreadsheet, either delete prior data or take care that the new data set has the same or larger population size than the original data set.

*Authors*
*Reid and Guthrie*

This will avoid retaining any previous data due to the current data having less samples. It is also recommended that Paste Special / Values be used as it will ensure the current formatting in the worksheet will be retained.

Note: The 'Worksheet' sheet can be used as a temporary storage sheet for data from your own chemometrics package. The data can then be arranged in the correct format ready for copying to the 'Data' sheet.

2. The values m1 and m2 are the differences between the reference value and the predicted values for calibration 1 and 2, respectively (i.e. the residuals between the reference and predicted values). The value d is the difference between the residuals m1 and m2 while m12 and m22 are the squares of the residuals m1 and m2, respectively.

3. A summary of the data is produced in sheet 'Fearn's Comparison'. The summary includes the mean, number of samples (N) and standard deviation (SD) for the reference (Actual) values, predictions (Pred1, Pred2) and residuals (m1 and m2) from both calibration models and the difference in residuals (d).

4. *Fearn's Comparison* worksheet compares the bias and the standard deviations of the prediction errors for the two calibration models being compared. For bias, the null hypothesis is that biases from both models are equal; for the standard deviation, the null hypothesis is that the ratio of standard deviations is unity. The user chooses the level of significance required (e.g. 1%, 5% or 10%) by inserting the required value in the yellow cell as a percentage. If no value is inserted then it defaults to a confidence level of 5%.

5. It is recommended that a record of the comparison be kept. This can be done by copying the results in the 'Fearn's Comparison' sheet to another spreadsheet. Simply using Paste will copy all the formats but will not correctly copy all the data (ie. those that are based on formulae). To transfer the data, Paste Special / Values will need to be used. A two-step procedure of using Paste (to copy formats) and Paste Special / Values(to copy the data) will ensure both formats and data are correctly transferred. A copy of the program is available for the asking:

David.reid@dpi.qld.gov.au
john.guthrie@dpi.qld.gov.au

---

[12] This article was published in the newsletter of the Council for Near Infrared Spectroscopy- *The NIR Spectrum* **2** (3) 7.