CSIRO PUBLISHING
www.publish.csiro.au/journals/ajar *Australian Journal of Agricultural Research*, 2005, **56**, 405–416

Assessment of internal quality attributes of mandarin fruit. 1. NIR calibration model development

J. A. Guthrie[A,B,C], *K. B. Walsh*[B], *D. J. Reid*[A], *and C. J. Liebenberg*[B]

[A]Delivery, Queensland Department of Primary Industries and Fisheries, PO Box 6014, CQ Mail Centre, Rockhampton, Qld 4702, Australia.
[B]Plant Sciences Group, Central Queensland University, Rockhampton, Qld 4702, Australia.
[C]Corresponding author. Email: john.guthrie@dpi.qld.gov.au

Abstract. The utility of near infrared spectroscopy as a non-invasive technique for the assessment of internal eating quality parameters of mandarin fruit (*Citrus reticulata* cv. Imperial) was assessed. The calibration procedure for the attributes of TSS (total soluble solids) and DM (dry matter) 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 720–950-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 squared standard error of calibration of <0.4 and 0.6%, respectively), whereas that for juiciness and total acidity was unacceptable. The robustness of the TSS and DM calibrations across new populations of fruit is documented in a companion study.

Additional keywords: spectral window, non-invasive.

Introduction

Near infrared 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 1994). Commercial application within pack-houses of Western countries, began in the 2000s (e.g. www.compac.com; www.cvs.com.au).

Various statistical terms and abbreviations have been used by authors working with NIRS technology. In this and the companion manuscript, the following terms and abbreviations have been used: bias (*bias*) is the difference between mean of actual and predicted values; standard deviation of the reference method values (SD); coefficient of determination on calibration dataset (R_c^2); coefficient of determination on validation dataset (R_v^2); root mean squared error of calibration (RMSEC); root mean squared error of cross validation (6 groups; without *bias* correction) (RMSECV); root mean squared error of prediction (without *bias* correction) (RMSEP); and RMSEP corrected for bias [RMSEP(C)]. The 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 that 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*. 1999*b*). 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*. 1999*a*). Near infrared 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

 10.1071/AR04257 0004-9409/05/040405

population with SD of 1.73% 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 0.52–0.74 and RMSEC of 0.41–0.64% 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 $1 - (\text{RMSECV/SD})^2$) 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), Schmilovitch *et al*. (2000), and Sohn *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 0.5–0.8 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 (Guthrie and Walsh 1997; McGlone and Kawano 1998), 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. Although citrus fruit do not store starch, DM content may be a useful index of certain internal quality defects. For example, Peiris *et al*. (1998) 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.

Thus, 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 optical geometry and the chemometric approach used (MLR, cf. PLS, data pre-processing technique, spectral window, etc.). In the current study, we seek to confirm the earlier reports of the utility of the NIRS method to the assessment of TSS, DM, TA, and juiciness in mandarin, using truly independent validation populations. We also seek to optimise each of these features in the development of a calibration model, prior to undertaking 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

Mandarin fruit (Imperial variety) were obtained following commercial harvest from orchards in Munduberra (25.6° S, 151.6° E), Bundaberg (24.9° S, 152.3° E), and Dululu (23.8° S, 150.3° E), Queensland. Fruit were obtained 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 approx. 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, 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). TSS 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).

Reference analysis

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 0.1 M 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 h. 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 306–1130 nm using a NIR-enhanced Zeiss MMS1 spectrometer and a 100-W 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). A Teflon tile (10 mm thick) was used as a white reference. 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 ms. For comparison purposes, 1 population was also assessed using the partial transmittance optical configuration used by Walsh *et al*. (2000) (45° 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 30°C).

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.

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' (Shenk and Westerhaus 1993) 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, λ, 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 (Shenk and Westerhaus 1993). Standard normal variance weights the absorbance at each wavelength by the SD of the calibration set. 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) (Shenk and Westerhaus 1993). Calibration performance was assessed in terms of R_v^2, RMSEP, SDR, slope, and *bias* of the validation sets.

The Matlab PLS and WINISI MPLS techniques gave equivalent model results for a given dataset (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 to 930 nm, with an end position of 800–1020 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'. Fearn (1996) has 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 used, facilitated by a spreadsheet (available from the corresponding author on request) that 'automated' the procedure. A significance level of 95% was used in these tests.

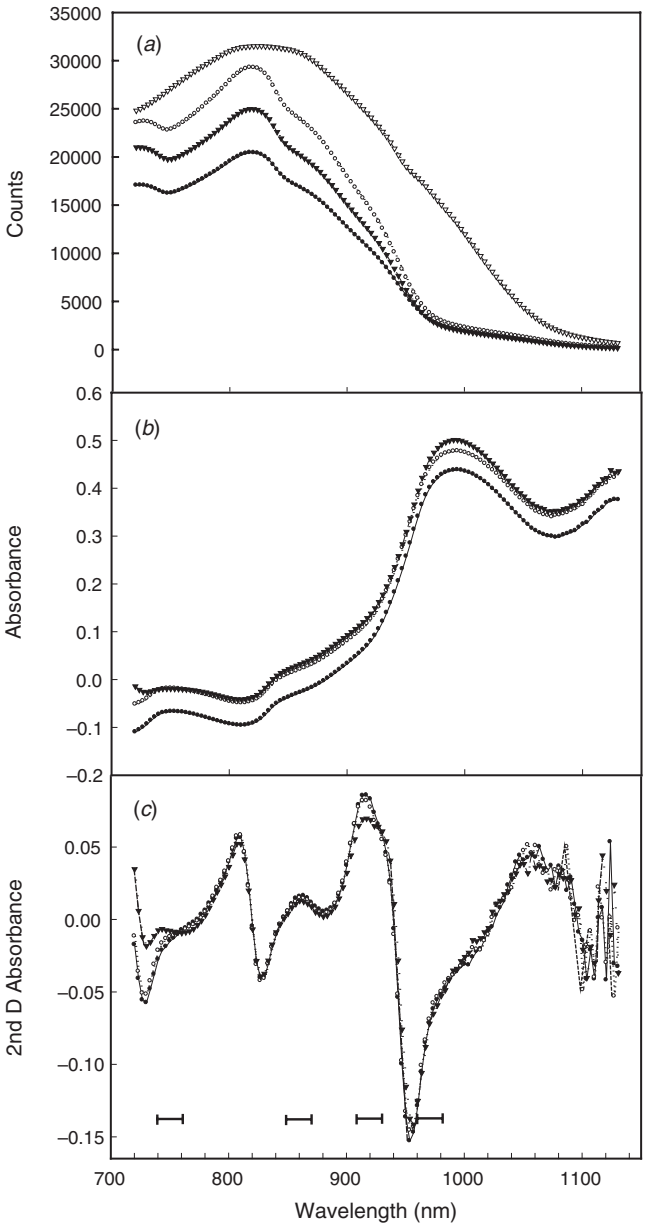

Fig. 1. Near infrared spectral data of 3 intact mandarins displayed as (*a*) analogue to digital counts including the white reference (open triangle), (*b*) absorbance, and (*c*) second derivative (2nd D) absorbance data. The mandarin fruit were chosen at random, representing fruit with high (12.7%, closed circle), medium (10.2%, open circle), and low (7.6%, closed triangle) total soluble solids. Horizontal bars show important spectral areas.

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

Table 1. Spatial distribution of TSS, DM, 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 followed by a common letter are not significantly different ($P = 0.05$)

	Pop. 1	Pop. 2			Pop. 3	
		Prox.	Equat.	Distal	Prox.	Distal
			TSS (%)			
External	8.9a	8.4a	8.4ab	8.3a		
Internal	8.3b	8.3b	8.4ab	8.5c		
Combined					9.2a	9.9b
			Juiciness (%)			
External	48.2a					
Internal	38.8b					
Combined					55.0a	55.6a
			DM (%)			
External		9.7b	9.6b	9.7b		
Internal		9.1a	9.7b	9.5b		

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%, whereas 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, whereas for internal tissue, DM at the proximal position was less than at 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, $n = 379$).

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 45° geometry was not significantly better in terms of RMSECV than any of the 0° geometries (Table 2) and therefore the 0° 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 0° 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 was similar for equatorial and distal positions (Table 2).

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.

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

Table 2. The influence of scan averaging, optical geometry, and fruit positioning on calibration statistics (R_c^2, RMSECV, and SDR) for 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° 1 scan treatment

Optics	Fruit position	No. of scans	R_c^2	RMSECV	Terms	SDR
0°	Equatorial	1	0.87	0.32a	7	2.4
0°	Equatorial	2	0.88	0.32a	7	2.4
0°	Equatorial	4	0.87	0.34a	7	2.4
0°	Equatorial	32	0.92	0.26a	7	2.9
45°	Equatorial	4	0.91	0.39a	9	2.0
0°	Proximal	4	0.68	0.63b	5	1.2
0°	Distal	4	0.88	0.30a	6	2.6

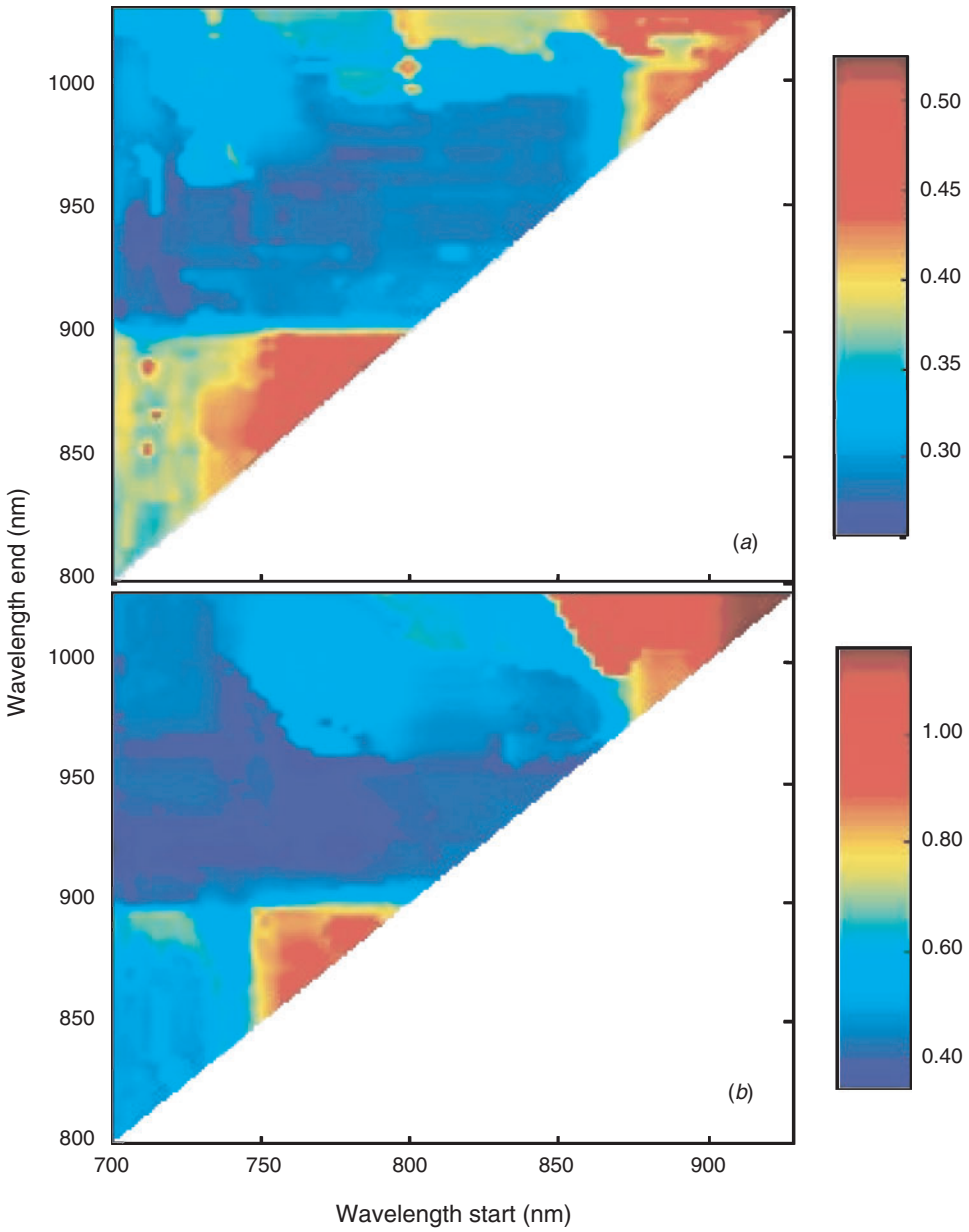

Fig. 2. Calibration model performance as assessed by RMSEC for varying spectral windows (varying start and end wavelengths). Partial least squares calibration models for (*a*) total soluble solids for populations J and K combined and (*b*) DM for population T. The colour code to the RMSEC values (% TSS) is shown in the bar scale to the right.

terms, and SDR) and prediction statistics (variance, RMSEP, and *bias*). The calibration models were used to predict TSS of an independent set drawn from the same harvest as the calibration group (data not shown), TSS for 5 independent populations (Populations A, C, J, L, and M; Table 3), and DM for 5 independent populations (Populations T, S, X, W, and V; Table 4).

For a model developed on 65 fruit from population L, and validated on the remaining 10 fruit, 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. The lowest RMSEP value (10.42% TSS) was recorded with the use of first derivative absorbance data with a gap of 4 nm and a 3 nm smoothing interval (data not shown). 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 approx. 10–100 nm, either side of the data point)

Table 3. Optimisation of data pre-treatment in terms of derivative treatment (none, first, or second order) and 4 scatter correction routines (none, SNV, detrend, or SNV and detrend combined) (a total of 12 treatments) for TSS calibration

Model performance is reported in terms of prediction of 5 independent sets 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, 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. The RMSEP (or *bias*) in bold was then compared with each underlined RMSEP (or *bias*) in the population using Fearn's significance test. 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.

Scatter correction	Variance (RMSEP)			*bias*		
	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	0.386	0.409	0.357	**0.092**	−0.096	−0.045
Detrend	0.492	0.402	0.360	−0.176	−0.087	−0.020
SNV & detrend	0.424	0.385	**0.332**	0.082	−0.011	**−0.048**
			Population C			
None	0.762	0.888	0.770	−0.485	−0.725	−0.561
SNV	0.973	0.867	0.760	−0.638	−0.674	−0.491
Detrend	0.960	0.849	0.648	−0.728	−0.605	−0.295
SNV & detrend	**0.707**	**0.611**	0.686	**−0.432**	**−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	**0.579**	0.476	0.532	**0.352**	0.193	0.304
SNV & detrend	0.640	**0.415**	0.450	0.448	**0.001**	**0.128**
			Population L			
None	0.635	0.693	0.660	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	**0.170**	0.448	0.579
SNV & detrend	0.631	0.526	**0.606**	0.474	**0.355**	**0.424**
			Population M			
None	**0.655**	0.732	0.498	**0.413**	0.503	−0.102
SNV	0.786	0.781	0.475	0.593	0.618	−0.039
Detrend	0.938	0.542	0.494	0.753	0.196	0.119
SNV & detrend	0.688	**0.462**	0.500	0.540	**0.007**	−0.248

(RMSECV was degraded from 0.43 to 0.49% TSS for first derivative data). In this respect, second derivatives appeared more sensitive than first order, with a decrease in calibration and validation performance at gap sizes greater that 15 data points (RMSECV of 0.40, 0.39, and 0.48% TSS at gap sizes of 3, 15, and 30 nm, respectively). Model performance was also relatively insensitive to smoothing interval.

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 SNV and detrend scatter correction, although there was little difference between first and second derivative treatments (Table 3). There was no clear requirement for derivative or scatter correction for DM calibrations (Table 4). In all further chemometric analysis reported in this study, unsmoothed first derivative absorbance data calculated using a gap of 4 data points, with SNV and detrend scatter correction, were used.

The RMSEP and *bias* for TSS and DM (averaged for the 5 populations of Table 3 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 5). Generally,

Table 4. Optimisation of data pre-treatment in terms of derivative treatment (none, first, or second order) and 4 scatter correction routines (none, SNV, detrend, or SNV and detrend combined) (a total of 12 treatments) for DM calibration

Model performance is reported in terms of prediction of 5 independent sets 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, SD = 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. The RMSEP (or *bias*) in bold was then compared with each underlined RMSEP (or *bias*) in the population using Fearn's significance test. 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

Scatter correction	Variance (RMSEP)			*bias*		
	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	0.749	0.761	0.792	**0.224**	**−0.188**	−0.202
SNV	0.777	0.786	**0.744**	0.396	−0.340	**−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	0.675	**0.614**	0.642	**0.208**	**0.042**	**−0.038**
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	0.806	**0.805**	0.904	**0.038**	**−0.043**	−0.417
SNV	0.947	0.837	**0.897**	−0.464	−0.153	**−0.402**
Detrend	0.832	0.858	0.909	−0.048	−0.310	−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**	**0.960**	0.867	**0.081**	**0.334**	0.163
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	**3.805**	**3.532**	3.643	**3.552**	**3.253**
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

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 5).

Calibration models developed on the combined (A–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 6) for the individual calibration populations (A and E) and the combined population (A–N). Multiple linear regression model validation performance (R_v^2) was generally improved (Table 6) 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 5), and for DM were 0.9 and

Table 5. Calibration model statistics for TSS in each of 17 populations (A–N) of mandarin fruit harvested over 2001, populations A–N combined, and populations from 3 other years

Calibration models were developed using MPLS and stepwise MLR, with a data pre-treatment of first derivative, SNV, 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	0.53	0.41	0.41	5	1.2	0.48	0.17	1	1.1
E	78	9.6	1.00	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	0.60	0.86	0.22	8	2.2	0.30	0.74	5	2.0
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	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	0.74	0.87	0.27	7	2.2	0.74	0.00	1	1.0
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	0.80	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	2.0
A–N	770	9.4	0.87	0.84	0.35	10	2.4	0.50	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	0.40	0.91	7	3.3

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.

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

Table 6. Prediction statistics for the validation of MPLS, MLR, and 'forced' MLR (FMLR) models on mandarin TSS, developed on populations A–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 sets	Regression analysis	Calibration sets					
		R_v^2			*bias*		
		A–N	A	E	A–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	0.20
	FMLR	0.68	0.73	0.72	0.14	1.09	0.64
Q	MPLS	0.73	0.73	0.68	0.23	0.00	0.76
	MLR	0.57	0.25	0.10	0.84	−0.33	0.16
	FMLR	0.60	0.55	0.64	0.00	0.86	0.55

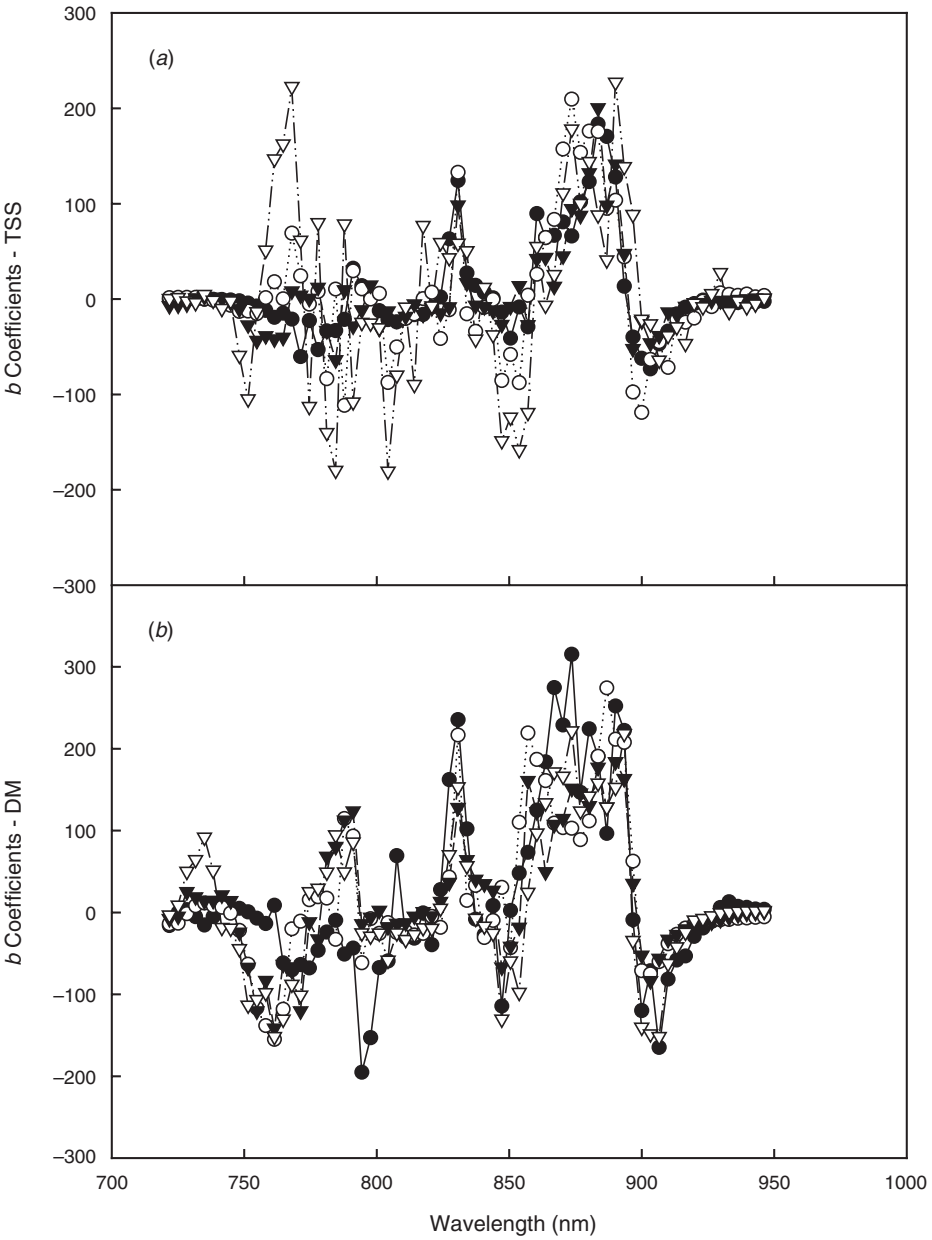

Fig. 3. Modified partial least squares calibration *b* (regression) coefficients for (*a*) TSS and (*b*) DM models using second derivative of absorbance with a gap size of 4 points. Model coefficients for populations E, G, L, and P are shown for TSS, and those for populations R, S, T, and X are shown for DM.

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 at proximal or distal positions. The 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*. (1999*b*) for a single orange and grapefruit. However, Peiris *et al*. (1999*b*) 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 700–950 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 1180 nm, respectively (Golic *et al*. 2003). Spectral features relevant to sugar CH groups include second order stretching bands between 1100 and 1200 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, whereas absorption at 840–855 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 4–9 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 750–1100 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 703–950 nm for both TSS and DM (Fig. 3).

Calibrations for TSS based on second derivative absorbance contained points of inflection for the *b* calibration coefficients at around 760, 860, and 905 nm, whereas for DM, points of inflection were around 760, 810, 850, and 910 nm (Fig. 2). 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 set. 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 (TA) are present in intact fruit at relatively low levels (*c*. 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 0° (interactance) and 45° (transmittance) geometries (e.g. R_c^2 of 0.91, RMSECV of 0.4% TSS) were not significantly different. The 0° (interactance) geometry was expected to produce a poorer calibration model than the 45° (transmittance) geometry due to increased detection of specularly reflected radiation and/or the shorter path length of light through the fruit in the 0° 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 0° geometry.

The 45° geometry was applied with the detector probe in contact with the fruit to exclude specular reflection. The 0° 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 0° 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 (1999).

Most literature reports use 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). On a commercial pack-line, operating at a belt speed of 1 m/s, there is sufficient time for only 1 scan.

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

Calibration data treatment

Models in which the coefficients give more weight to 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 4–6 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 6). This result was not related to outlier detection and removal routines as no outlier removal was undertaken in these validation exercises. Thus, although 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 sets (Tables 5 and 6). 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 infrared region.

For TSS, the optimal derivative and scatter correction condition differed among validation groups, but in general, a first derivative with SNV and detrend routines supported superior model performance. For DM, no method was consistently superior to other methods. Hence, 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 720–950 nm window, pre-treated as first derivative absorbance data (gap size of 4 data points) with SNV 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.

Acknowledgments

Funding support was received from Horticulture Australia (Citrus Marketing and Development Group) and Central Queensland University-Research Training Scheme. Fruit was supplied by Steve Benham of Joey Citrus, Munduberra, and by Jim and Deslea Yeldham, Citrus Farm, Dululu, Qld. This manuscript represents an extension of work reported at the 10th International Conference of Near Infrared Spectroscopy, Konju, Korea, June 2001.

References

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

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. doi: 10.1366/000370203321535033

Greensill CV, Walsh KB (1999) 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. doi: 10.1088/0957-0233/11/12/304

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.

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. doi: 10.1071/EA96026

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

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.

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. doi: 10.1016/S0925-5214(98)00007-6

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.

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 and Technology* **41**, 908–912.

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 (1998) Non-destructive detection of section drying, an internal disorder in Tangerine. *HortScience* **33**, 310–312.

Peiris KHS, Dull GG, Leffler RG, Kays SJ (1999*a*) 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 (1999*b*) 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.

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. doi: 10.1016/S0925-5214(00)00102-2

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

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.

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

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, 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.

Manuscript received 3 November 2004, accepted 23 February 2005