Postharvest Biology and Technology 163 (2020) 111139

Contents lists available at ScienceDirect

Postharvest Biology and Technology

journal homepage: www.elsevier.com/locate/postharvbio

The uses of near infra-red spectroscopy in postharvest decision support: A review

K.B. Walsh[a,*], V.A. McGlone[b], D.H. Han[c]

[a] Central Queensland University, Rockhampton, Qld, Australia
[b] The New Zealand Institute for Plant and Food Research Limited, Hamilton, New Zealand
[c] College of Food Science & Nutritional Engineering, China Agricultural University, Beijing, 100083, China

ARTICLE INFO

Keywords:
Near infrared spectroscopy
NIRS
SWNIRS
Herschel region
Quality
Internal defect
Sorting
Statistics

ABSTRACT

This review considers applications of near infrared spectroscopy (NIRS) in context of postharvest decision support, as opposed to an assessment of NIRS technology, *per se*. Two 'generations' of use are discussed, the first involving the direct assessment of chemical or physical attributes related to postharvest quality at the time of assessment, and the second involving the forward prediction of a postharvest attribute of the fruit or vegetable. A review of statistical functions relevant to sorting is also presented, with consideration given to the measurement error inherent in an estimate based on near infrared spectroscopy and the use of Receiver Operating Characteristic and similar parameters. Case studies involving implementation of NIRS into various aspects of postharvest value chains are presented.

1. Introduction

Fleshy fruits and vegetables (hereafter referred to as 'fruit') are perishable commodities, of relatively low unit value. The aim of their postharvest management, from harvest to consumption, is to deliver a product of acceptable eating quality to the customer, despite storage or transport. Managers must make decisions in the orchard on the timing of harvest, in the packhouse on sorting criteria that include internal quality attributes, in ripening and storage rooms on duration, and at receiving centres on consignment acceptance. Other players in the value chain also have need for information on fruit attributes, e.g., to support breeding programme selections. Decision making at all these levels rests on measurement technologies that allow assessment of relevant attributes.

External attributes (colour, shape, weight, external defects) are assessed on packing lines using machine vision and load cell technology. The assessment of internal attributes has traditionally involved destructive sampling; however, a number of non-invasive technologies are relevant, including transmission X-ray, acoustic transmission or resonant frequency, accelerometers and visible - near infrared region spectroscopy (NIRS) (e.g., Magwaza et al., 2013). Of the technologies for non-invasive internal quality assessment, NIRS technology has seen the highest adoption into commercial practice.

The abbreviation NIRS is usually defined in scientific literature as spectroscopy involving the region 750–2500 nm, or a subset of this range, depending of the detector used. Commercial fruit applications of NIRS involve the short-wave NIR (SWNIR or "Herschel') region, 750–1100 nm, and also often include the visible wavelength range, 400–750 nm, as the typical silicon-based spectrometers that are used in commercial applications are sensitive across both ranges. NIRS technology has benefited from a confluence of technological developments in optical hardware, spectrometer miniaturisation, computing speed and chemometric modelling, enabling application on packing lines assessing up to 10 items of singulated fruit per second, and in handheld units suitable for field use, operating in full sunlight and varying ambient temperature. Fruit are thus able to be assessed *in situ*, rapidly and non-invasively. In the current review, we explore the implementation of this capacity within horticultural value chains.

On the packing line, NIRS spectrometers allow assessment of every item in a population rather than assessment of only a sample of fruit, while handheld spectrometers allow lot assessment, involving sampling of a population. Handheld NIRS technology can be applied at any point in the value chain, e.g., on-farm, in the packhouse, at the storage/ripening/repacking centre, at the distribution centre or in retail situations, while in-line implementation is obviously tied to the use of a packing line. The 'user case' varies for each point of application within a given value chain. For example, application on-farm may allow a management decision such as harvest timing, to optimise an attribute

* Corresponding author.
E-mail address: k.walsh@cqu.edu.au (K.B. Walsh).

https://doi.org/10.1016/j.postharvbio.2020.111139
Received 7 August 2019; Received in revised form 13 January 2020; Accepted 29 January 2020

within a population. Lot sampling in a packing shed can provide feedback to suppliers, while in-line grading after harvest can allow maintenance of a specification. In-line grading may be again applied after storage and transport, e.g., in repacking centres of global supply chains. For example, The Netherlands, which has no mango trees, is the world's second largest exporter of mango, after Mexico (http://www.worldsrichestcountries.com/top_mango_exporters.html; doa 15/1/2019), as it acts a trans-shipping point for fruit arriving into Europe. Arriving mango fruit are partly ripened. An assessment is required of the remaining time to ripen and thus shelf life, to inform the management decision on whether to further store the fruit, enhance ripening with ethylene, or to distribute. For use at the 'other end' of the supply chain, Kubota (Naniwa-ku, Osaka, Japan) has marketed a benchtop NIRS from the late 1990s for use in fruit outlets servicing the high-value 'gift market', in which a gloved attendant undertakes the NIRS assessment of fruit for the customer, while, in recent years, SCiO (Consumer Physics, Israel) has produced a handheld unit marketed for use by consumers.

From an overall industry perspective, however, the further 'upstream' in the value chain such measurement technologies are applied, the greater the management options for actions to improve the quality attribute or to segregate the lot into different markets, as opposed to a reject-to-waste decision. To date, the greatest impact of NIRS on commercial postharvest practice has occurred in context of integration into packing line sorting, with sorting to meet a produce specification.

It is notable that although NIRS has been in commercial use for assessment of present internal attributes of fruit since the late 1980s, a surge in scientific publications about this application occurred decades later, representing a catch up in public science understanding of the technology. There are now many published studies and indeed many reviews on the technical detail of NIRS in fruit assessment, from optical geometries to chemometrics (e.g., reviews by Nicolai et al., 2007 and Wang et al., 2015), providing calibration cross validation and (ideally) validation statistics on a series on independent populations to demonstrate model robustness. Papers are only recently beginning to be published where NIRS is 'just' a method, i.e., mentioned in the Materials and Methods section and then used to support a line of enquiry (e.g., on the effect of irrigation scheduling on fruit dry matter (DM) content; Anderson et al., 2017), rather than the focus of the paper as a whole. This situation suggests that the technology is immature, and yet it is in modest but widespread commercial use. An explanation lies in the relative unavailability of the packing line NIRS technology to the scientific community, and it is expected that the trend to publish 'with' the method rather than 'about' the method will accelerate with the relatively widespread availability of handheld NIRS instrumentation targeted at fruit assessment (e.g., the Pocket IR Brix meter PAL-HI-KARi2 from Atago, Japan, and the F750 from Felix Instruments, USA).

The focus of the current review is therefore not on the technology of NIRS *per se*, but rather on the use of NIRS technology in the context of postharvest applications and the value of the technology as a decision support tool to a manager. The intent of the review is to capture a sense of the extent of adoption of NIR technology in postharvest applications. This discussion steps through: (i) postharvest attribute specifications for which NIRS assessment is typically currently used; (ii) consideration of two categories of application, assessment of present and of future attributes.; (iii) currently available technology and restrictions to uptake; and (iv) statistics used in sampling and sorting operations, with attention to the impact of NIRS measurement errors. Examples of postharvest applications of NIRS technology are also explored in case studies.

2. Postharvest specifications

Within a value chain, fruit are subject to a range of specifications on external and internal attributes. The relevant attributes and acceptable level are specific to the commodity and indeed to cultivar. Examples of such specifications can be found in the Codex Alimentarus (http://www.fao.org/fao-who-codexalimentarius/codex-texts/list-standards/en/, doa 19/2/2019) and in specifications set by major retailers (e.g., www.wowlink.com.au, doa 19/2/2019). The attributes of relevance to assessment by NIRS are primarily DM content of fruit tissue and soluble solids content (SSC) of extracted juice, with some claims for assessment of firmness and acidity. A list of specification targets for these internal attributes is presented in Table 1 (see Walsh, 2014, for references to the supporting studies). In fruit that do not have a climacteric change in carbohydrate form, from starch to sugars during ripening, SSC of fruit at harvest is often the attribute of interest in terms of eating quality. For carbohydrate-storing climacteric fruit in which starch is converted to soluble sugars during ripening, DM content in an index of the sum of starch and soluble sugar contents, which in turn is proportional to SSC of ripened fruit. Thus DM content is relevant to a specification for fruit at harvest. In lipid-storing fruit such as avocado, DM content is a useful surrogate for oil content.

Table 1
Internal attribute specification recommendations, by commodity (from Walsh, 2014), for attribute/commodities for which in-line NIRS based grading is claimed. Units of soluble solids content (SSC are %w/v; dry matter (DM), % FW; and firmness, N (measured with a 12 mm cylindrical probe). Climacteric fruit are indicated by '+', non-climacteric as '-'.

Fruit	Climacteric	Attribute	Value
Avocado	+	DM	21
Citrus			
(mandarin)	–	SSC:acid	8:1
(orange)	–	SSC:acid	8:1
(navel orange)	–	SSC:acid	navels: 7.5:1-9.0:1
		juice content	50
Grape - table	–	SSC	Ribier, Red Malaga, Emperor 16; other 17
Kiwifruit	+	SSC (at harvest)	6.2
	+	SSC (ripe)	14 or 15
	+	SSC (for storage)	7–9
Mango	+	SSC	15 or 16
		DM (at harvest)	14
		specific gravity	1.01-1.02
Melon	+	SSC	10
Pome fruit			
(apple)	+	SSC	'Jonathan' 11; 'Delicious' and 'Red Delicious' 10 (at harvest) and 12–14 (ripe), 'Starking' and 'Delicious' 10.8–12.2, 'Delicious' and 'Spartan': 9-11
	+	SSC (for storage)	'Delicious' 10; 'Bonza' 13; 'Golden Delicious' 12; 'Gala' 12.5; 'Granny Smith' 12; 'Fuji' 13; 'Cripps Pink'/Pink Lady® 15; 'Sundowner' 14.5; 'Lady Williams' 14.5
(pear)	+	SSC	12 (at optimal firmness, TSS:acid ratio) or 13
Stonefruit			
(apricot)	+	SSC	10
(cherry)	–	SSC	14-16 depending on cultivar
(nectarine)	+	SSC	10
(peach)	+	SSC	10
		SSC	11
		Firmness	0.9-1.4
(plum)	+	SSC	11
			12

3. Available technology and adoption issues

3.1. Technology

The first commercial applications of NIRS to postharvest fruit quality assessment involved implementation on fruit grading lines in Japan in the late 1980s (Norris, 1992). In this implementation the complete NIRS component (spectrometer, lighting, data processing, etc.) for the grading line was delivered by a second party (e.g., by Mitsui Mining and Smelting Co., Ltd and Sumitomo Metal Mining Co., Ltd, Japan), This uptake was driven by the availability of premium prices for premium produce, with initial focus on the attribute of SSC in apples, stonefruit and melons. Adoption by packing-line manufacturers in Australia and New Zealand followed in 2000, and in Europe and the USA by the mid-2000s (Walsh, 2005). Development of handheld, field-portable variants of the technology occurred by the mid-2000s, driven by a need to make assessments elsewhere in the value chain, from in-field to guide timing of harvest, to at-retail to enforce a specification. These implementations use original equipment manufacturer (OEM) spectrometers (e.g., from Hammamatsu, Japan; Zeiss, Germany; Avantes, The Netherlands) within a custom design (e.g., of optical geometry, lighting and user interface).

Subsequent years have seen both consolidation of the initial field of instrument providers and the entry of new providers. There are now a number of providers of on-packing line NIRS equipment and a smaller number of providers of handheld NIRS equipment targeted at the assessment of macro-constituents in fruit.

The first generations of in-line equipment in Japan used a reflectance geometry, for ease of implementation. This was superseded by partial (i.e., interactance) or full transmission based units which provided superior prediction robustness across lots of fruit. In the authors experience, all current instrumentation targeted to fruit application is based on a silicon photodiode array or charge coupled device (CCD) detector (i.e., wavelength range to 1100 nm), with use of multiple linear regression (MLR) or partial least squares regression (PLSR) models for assessment of continuous variates (e.g., SSC). A discriminant analysis (DA), with categorisation of samples to two or more categories (e.g., pass/fail), can be appropriate for assessment of internal defects such as internal translucency or browning (e.g., Khatiwadi et al., 2016).

3.2. Adoption issues

A limiting factor to NIRS technology adoption is the relatively small market size for such equipment (Walsh, 2005). Effectively this is a niche application. There are probably fewer than 200,000 electronic fruit grading lanes globally, of which a fraction would be candidates for NIRS technology (authors' estimate from interactions with grading line manufacturers). Furthermore, the optical geometry required in NIR instrumentation can differ by commodity and attribute, fragmenting an already small market. Of course, the case for NIRS technology adoption can be re-opened by market changes that bring tighter specifications and/or by improvements in technology. For example, 'sugar-end' is a defect in potato tubers in which one end of the tuber accumulates a high concentration of soluble sugars. One manufacturer of packing line equipment developed equipment capable of sorting for this defect based on a series of point NIRS assessments across the tuber, but discontinued development in the mid-2000s owing to lack of demand. Hyperspectral imaging has recently been incorporated into commercial sorting equipment to enable sorting of the potato sugar-end defect in cut products (e.g., Key Technology, Walla Walla, WA, USA).

Another limitation to NIRS technology adoption is its status as a secondary method, requiring calibration against a primary reference method. Quality control procedures must be in place to ensure performance through correction or updating of calibration models as necessary. This requirement does not suit a 'set and forget' or casual use, but rather is suited to implementation in high throughput situations with the ability to maintain a quality control of instrumentation. This calibration requirement has also underpinned a range of business models from straight equipment sale with a maintenance contract to lease and throughput models, with some manufacturers allowing full customer control over calibration development and others requiring customers to have calibration data processed as a provided service. Low-cost handheld equipment (e.g., the SCiO from Consumer Physics, Israel, and the Hikari from Atago, Tokyo, Japan) provided with 'set and forget' calibration models offer ease of use, but presumably sacrifice accuracy and precision.

Yet another consideration in the adoption of NIRS technology in postharvest operations is the fit to the whole operation. For example, consider that in-line grading for an attribute complicates packhouse operation by increasing the number of categories of sorted fruit, e.g., colour × weight × NIR-attribute, and potentially creates a lower value category of under-specification fruit.

4. Assessing 'present' attributes

The majority of current postharvest NIRS applications involve assessment of a 'present' attribute, i.e. an attribute existing at the time of measurement, in the context of a specification value for that attribute (as per Table 1). For example, the very first report of NIRS in postharvest biology was by the founder of the NIRS discipline, Karl Norris of the USDA, who demonstrated the use for apple internal rot detection in the early 1960s (Norris, 1992). This work was followed by that of Birth et al. (1985) for the assessment of DM content of onions, another present attribute. Vis-NIRS has been subsequently reported in commercial use primarily in the context of measurement of (i) pigmentation, (ii) DM content or its inverse, water content, (iii) extracted juice soluble sugar content (SSC), (iv) titratable acidity (TA) and (iv) detection of internal defects such as watercore or internal browning (e.g., the review by Wang et al., 2015). A bias-corrected RMSEP of > 0.5 % is commonly reported for macro-attributes such DM (% w/w) and SSC and TA (% w/v) (e.g., Walsh, 2015).

There are published claims for assessment of (i) attributes at less than macro-constituent thresholds, such as TA in stonefruit, (ii) attributes that do not have a chemical basis, such as firmness, and (iii) for direct assessment of attributes or indices that are based on a combination of constituents (such as SSC and TA, as used in the Brix:acid ratio, or ripeness) (e.g., McGlone and Kawano., 1998, Subedi et al., 2012). Documentation of the spectroscopy science behind each application case (commodity and attribute) should be provided, to better anticipate the detection limits of the technique and the conditions under which a predictive model will fail.

For example, attempts have been made to relate NIRS-based estimates of intact fruit firmness to pectin contents or water mobility (assessed by Proton nuclear magnetic resonance, 1H NMR), or to changes in cell structure and thus light scattering and apparent absorption (e.g., Uwadaira et al., 2018). These claims deserve re-examination given that other reports suggest that NIRS firmness models are primarily based on secondary correlation with other fruit macro-attributes amenable to spectroscopic assessment (e.g., chlorophyll or soluble sugar contents) (McGlone and Kawano, 1998; Subedi and Walsh, 2008). Similarly, NIRS assessment of acidity in intact fruit may rely on secondary correlations with attributes such as SSC or chlorophyll, except in fruit containing high amounts of organic acids and little sugar (e.g., lemons, limes) (McGlone et al., 2003; Subedi et al., 2012). For postharvest management use of a NIRS-estimated value, it does not matter if the measurement is based on a direct correlation with the NIR spectra (e.g., a DM prediction predicated on sugar and water molecule absorption) or a secondary correlation (e.g., TA and firmness prediction predicated on correlated maturity features such as fruit chlorophyll content), as long the relationship is robust in the face of variations presented by fruit populations differing because of environmental and/or seasonal changes. However, the onus is on the proponent of such a relationship

Table 2
List of attributes claimed to be measurable using near infra-red spectroscopy (NIRS) on commercially available on-packing-line equipment, by commodity, e.g., http://www.shibuya-sss.co.jp/sss_e/product/miq.html (doa 15/2/2019). Attribute key is: 1, dry matter (DM); 2, soluble solids content (SSC); 3, titratable acidity (TA), 4, internal browning; 5, internal rot; 6, degree of ripeness; 7, astringency; 8, sorbitol; 9, juiciness; 10, water soaked; 11, internal cavity; 12, skin separation. The symbol ‘x’ indicates ‘in use’, and ‘o’ under consideration/development.

	Attribute											
Commodity	1	2	3	4	5	6	7	8	9	10	11	12
peach		o		x		x	x					
Asian pear		o	o	o		o				o		
apple		o	o	o		o		o				
persimmon		o				x						
mandarin		o	o									
orange		o	o						o			x
cantaloupe		o				x				o		
watermelon		o				x				x		
tomato		o	o									
kiwifruit		x	x									
European pear		o				x						
potato	o										x	
onion					x							
mango	o											

to demonstrate robustness. Such secondary correlations may not be robust for fruit of populations from different growing conditions.

Instrument providers currently claim use of NIRS in assessment of a range of fruit internal quality parameters, principally in the context of thin-skinned fruit (Table 2).

Case studies are presented below of NIRS technology in use for assessment of present attributes in postharvest value chains (e.g., DM and internal defects).

It is expected that quantitative applications will continue to expand in scope and approach. For example, although pesticide residues on fruit occur at rates unlikely to be detectable by NIRS of whole fruit, Acharya et al. (2012) demonstrated detection of contact pesticides through use of a dry extract concentration method. It was proposed that the method be used as a preliminary screen, with positive samples targeted for follow up gas chromatograph mass spectrometers (GC−MS) analysis. Such a protocol allows for wider screening effort on the same budget. Other vibrational spectroscopies such as FTIR and Raman also hold promise for this type of application (see reviews by Bureau et al., 2019 and Qin et al., 2019). For practical adoption, such applications need not be non-destructive nor able to be used on-line, if convenience is provided at appropriate accuracy and cost relative to existing assessment techniques.

There are also an increasing number of literature reports of NIRS in fruit categorisation beyond defect assessment, although commercial use is yet to be reported. Claims include assessment of analysis of geographic origin (e.g., Li et al., 2018), and production system (e.g., organic compared with conventional, Amodio et al., 2017). The robustness of such models requires further testing and documentation to establish reliability.

5. Assessing future attributes

Another category of NIRS application involves prediction of a future attribute. Examples of such applications include storage life or future eating quality. The assessment can involve prediction of a present attribute with a known relationship to the future attribute. Alternatively, the model can be directly based on the future attribute.

For example (see case studies 1 and 3 below), the future eating quality of a kiwifruit or mango fruit is related to the SSC of the ripened fruit, which is determined by the carbohydrate content (starch and sugar) of the unripe fruit, with starch converted to sugar during ripening (Fig. 4). As total carbohydrate content is indexed by DM, NIRS-based estimates of DM of fruit at harvest can be used to estimate future (ripened) SSC (e.g., for mango, Subedi et al., 2007; Subedi and Walsh, 2011).

In an alternative approach, the NIRS model can use fruit spectra for direct prediction of a future attribute value, e.g., the time to harvest for apple fruit, as in case study 4. As for assessment of a current attribute, it is advisable to understand the spectroscopic basis of the NIRS model, and the model should be tested in terms of robustness and interpreted in terms of its mechanism.

There have been scientific reports of applications involving less interpretable relationships between current spectra and a future attribute. For example, spectra have been related to to ripening and storability time of harvested fruit, and to estimation of predisposition of fruit to expression of a disease or a disorder. For example, Jacobs et al. (2016) reported estimation of the prior storage period of lettuce using NIRS. Magwaza et al. (2012) reported on use of NIRS to estimate citrus fruit rind susceptibility to breakdown. Proof of reliability of such applications lies in demonstration of robust performance across multiple independent populations, varying in growing and storage conditions. No commercial implementation is known to the authors.

6. Decision support statistics

Four areas are explored in this section: (i) the need to provide NIRS method validation statistics; (ii) sampling statistics relevant to lot sampling using NIRS; (iii) statistics relevant to time series analyses using NIRS measurements; and (iv) sorting statistics involving a categorical (pass-fail) decision.

6.1. Operating a NIRS-based fruit assessment system

As for any method, NIRS method performance should be characterised, especially because NIRS is a secondary measurement technology involving the modelling of a relationship between spectra to an attribute for which a reference measurement is made. Many literature reports provide performance statistics only in context of a single harvest population (allocated to calibration and validation sets). Variation in NIRS prediction performance across fruit populations varying in growing condition, region or season has been documented, e.g., in apple (Peirs et al., 2003) and mango (Subedi et al., 2007), with a generalisation that a robust model should contain populations from at least three production seasons. In terms of practical performance of the technology, prediction results should be presented for fruit populations from harvests separate from that supplying the calibration set (e.g., different harvest date, orchard, season, cultivar).

The measurement error associated with NIRS technology is often reported as root mean square error of cross validation (RMSECV). As noted above, practical performance of the technology should involve assessment of an independent population, in which case the relevant metrics are root mean square of error of prediction (RMSEP) and bias. Bias is the difference between the mean of predicted and actual values for a population, which is zero for the NIRS calibration set, but can be quite large for predictions of independent populations. The RMSEP encompasses prediction bias (accuracy) as well as prediction precision, as:

$$RMSEP = \sqrt{bias^2 + SEP^2} \tag{1}$$

where SEP is bias-corrected RMSEP, i.e.

$$SEP = \sqrt{\frac{\sum (predicted - actual - bias)^2}{n}} \tag{2}$$

Note that SEP is related to R^2 and population standard deviation (SD) as:

$$R^2 = 1 - (\frac{SEP}{SD})^2 \quad (3)$$

Thus R^2 alone is not a useful description, as it will increase with increasing population SD. Rather, model performance should be, at a minimum, documented in terms of SEP or R^2, bias and SD.

In practice, SEP is reasonably stable for established (multi-year) NIRS-based PLSR models of attributes that are commonly assessed in commercial practice (e.g., > 0.5 % FW for DM content, and 0.5 %w/v for SSC), but unacceptable biases (> 1 %FW on DM) on prediction can occur on incoming populations. This bias will be associated with instrument or sample change. Change due to sample or instrument temperature can be accommodated in the modelling process, having a predictable impact on water absorption peaks and detector sensitivity, respectively. Other changes are less predictable. For example, growing conditions may affect fruit cell size, skin thickness and composition, with impact on a NIR-based prediction, primarily as bias. The impact on bias (accuracy) rather than SEP or R^2 (precision) occurs because the primary measurement, absorbance, is determined as the negative logarithm of the ratio of the intensity of detected light to the intensity of the source light. The additive rule with logarithms means that both any change in the source light component and any multiplicative gain factors on the detected light that are constant with wavelength can be collated and segregated from the intrinsic spectral pattern of the detected light.

Bias can be quantified by measuring the average difference between predicted and actual values on a sample set. Thus, bias can be 'simply' accommodated in further predictions. This correction is central to commercial viability for NIRS applications, avoiding the need for expensive re-development of multivariate models. However, there is an art involved in judging how often to check for the need for adjustment.

A statistic of practical importance is that related to 'outlier detection' – a measure of the spectral fit of a sample to be predicted to the calibration set. Commercial NIR-based assessment systems should be equipped with some form of 'outlier detection' to warn the user when current samples depart from 'space' covered by the calibration set. Common methods involve estimation of Mahalanobois distance or Hotelling's T-squared statistic using the PLS factors of samples (e.g., Garrido-Varo and Garcia-Olmo, 2019).

6.2. Lot sampling

Every fruit of a consignment can be assessed on a packing line using NIRS technology. However, in many situations, NIRS is used in assessment of a sample of a population, rather than the whole population. The central limit theorem (CLM) of statistics can be used to determine the number of samples (n) required to accurately estimate the mean value of a continuous attribute (Fornasini, 2008) (Eq. 4):

$$n = \left(\frac{t.\ SD}{e}\right)^2 \quad (4)$$

where t is the t-statistic and e is the appropriate measurement error, such as the (bias-corrected) SEP value from the NIRS model. A preliminary sampling is required to estimate the population SD.

For example, internal browning of apple was non-invasively assessed using NIRS using a 1–5 reference scale on severity of the disorder (Khatiwadi et al., 2016). As a prior sampling estimated that a SD of 1.5 existed at the attribute level in a population, the minimum number of samples required to estimate the mean severity is estimated at n = 162, using t = 2.04 (t statistic for a 95 % confidence interval at n = 30) and a measurement error of e = 0.24 (the RMSE of operator repeatability on the five-point visual score scale). If < 162 samples were used in the estimate of SD, a second sampling event using > 162 samples should be undertaken, and n recalculated. If acceptable error was relaxed to e = 0.5, a minimum of only 37 samples would be required.

The sample size n required to detect a difference d between the two distributions can be estimated using power analysis (Faul et al., 2007). This approach attempts to balance out the Type I and II errors, choosing suitable values for them, and then working back to obtain an estimated sample size. For example, considering a one-sided test and assuming normal distributions with the same variance σ^2 for simplicity, sample size can be estimated as:

$$n = \frac{(Z_\alpha + Z_\beta)^2 .\ 2.\ \sigma^2}{d^2} \quad (5)$$

where Z_α and Z_β are the appropriate Z-scores to take for chosen Type I (α) and Type II (β) error rates respectively.

Further consideration of sampling statistics is warranted, with the aim of producing decision support aids to postharvest management, to assist in design of sampling regimes. At the least, future publications involving detection of a character should justify the sample size used, as described above, consider the presence of bias and slope issues when estimating sample statistics.

Another consideration is the measurement error introduced by the use of NIRS estimated values to estimates of population SD and mean, requiring consideration of both the bias and slope of the relationship between the actual (reference) values and the NIRS-predicted values. For calibration training data, the slope and bias are pre-determined to be one and zero, respectively, but for prediction data sets the (predicted to actual) slope is often less than unity. In consequence, the standard deviation of predicted values σ^2_{nir} can be less than that of the actual values σ^2_{lab}:

$$\sigma^2_{lab} = \sigma^2_{nir} + SEP^2 \quad (6)$$

where SEP is the NIRS modelling error. The slope of the relationship must be considered in computing the likely laboratory error estimate from a subsample:

$$\sigma^2_{lab} = a + b^*\sigma^2_{nir} + s^2 \quad (7)$$

where a is the intercept, b is the slope and s is the error on the linear regression of laboratory values on NIRS-predicted values. For instance, if the slope b is less than one, then $\sigma_{lab} < \sigma_{nir}$ can occur. Without a full validation set to allow accurate estimation of slope, the presumption has to be that the slope is near unity, and that $\sigma_{lab} > \sigma_{nir}$.

In some cases, however, the fresh fruit value chain specification is one of discrimination (accept/reject), not estimation of degree of severity of defect. Diffuse browning in apple is described as a major defect by major retailers, with a specification of no more than 2 % of fruit in a consignment to be affected (e.g., Woolworths, 2019). The typical test imposed by retailers is the cutting and assessment of 30 fruit on receipt of a consignment of fruit at the distribution centre. This is a binomial sampling problem, in which the probability (P) of encountering a defect fruit when n fruit are sampled is:

$$P = 1 - (1 - p)^n \quad (8)$$

where p is the actual proportion of defect fruit in the population. The expected mean value of observed defects is $E_{mean} = np$, with an expected standard deviation:

$$Estd = \sqrt{np(1 - p)} \quad (9)$$

For a consignment with a defect incidence of 2 % (p = 0.02), the expected number of defects in a random subsample of 30 fruit is $E_{mean} = 30 \times 0.02 = 0.6$, with $E_{std} = 0.77$. The probability of selecting a defect fruit is 0.45 (Eq. 8). Obviously, a sampling strategy using only 30 fruit would more often be in error than not, failing to detect presence of defects in most samplings of a population with a 2 % defect incidence rate. Indeed, a supplier with perfect knowledge of product defect rates could choose to provide the retailer with fruit of a higher degree of defect incidence than the specification, given the low probability of detection. To be more than 95 % certain of selecting a defective fruit, the sample size would need to increase to n = 150 and would yield an expected number of defects $E_{mean} = 3$ with $E_{std} = 1.7$.

Further consideration of sampling statistics is warranted, with the aim of producing decision support aids to postharvest management, to assist in design of sampling regimes. At the least, future publications involving detection of a character should justify the sample size used, as described above, and consider the presence of bias and slope issues when estimating sample statistics.

6.3. Population time series modelling

Many reports involving NIRS monitoring of fruit properties over time use the basic statistics of means and standard deviation of sample lots taken from a population across time. However, the non-invasive nature of NIRS assessment allows for repeated measures of individual fruit. Repeated measures of a set of samples avoids the variation inherent in random lot sampling and thus allows clearer observation of time series events.

Jordan and Loeffen (2013) developed a modelling approach using quantile functions, which do not assume normal distribution for an attribute, to describe fruit population variation across time, using statistical descriptors of biological age and aging rate, measurement uncertainty and other sources of variability. While this report did not involve use of NIRS in assessment of attributes, the concept suits such assessments. Walsh et al. (2015) used repeated NIRS assessments of DM content and internal flesh colour in mango fruit. Each fruit was demonstrated to follow a similar maturation trajectory, with a time offset. The offset was presumed to represent differences in date of pollination or environmental conditions around the fruit (e.g., inner or outer canopy). The time series data were described with a non-linear indexed regression model and a biological shift factor. The shift factors were greater for dry matter than for flesh colour, consistent with an earlier but lower rate change in dry matter.

These modelling approaches provide an estimate of the time at which a specified fraction of the fruit should meet a desired specification. These techniques could also be used to trace the source of variation within a population (e.g., to time of fruit set, location in canopy or plant water status), towards the goal of reducing this variation, leading to crops of greater uniformity.

6.4. Sorting statistics

The landmark paper in the field of sorting statistics in fruit postharvest applications is that of Bollen and Prussia (2009), but the topic remains rather weakly developed, with few published applications specific to produce sorting. For example, Bollen and Prussia chose to focus on an analysis approach termed Signal Detection Theory (SDT). However, while the SDT theory has valid application, especially for detecting one signal distribution in the presence of another distribution, it does not necessarily apply to all produce-sorting circumstances. One common circumstance for NIRS-based sorting involves attribute values (e.g., DM) of a population, with sorting required to apportion off a fraction of that distribution with high or low attribute levels. The theoretical modelling required in that circumstance involves truncated distributions, split around the chosen cut-point by convolution with an appropriate NIR sorting (i.e., error) function, and probably can only be modelled numerically. Future studies should consider modelling use of Monte Carlo methods (e.g., Harrison, 2010).

If the estimate of a continuous variable is being used to judge acceptability of a population with reference to a specification on that attribute, the issue of Type I and II errors must be considered. A Type I error involves the false rejection of an acceptable or good fruit, while a Type II error involves the false acceptance of a non-acceptable or poor fruit. The proportion of Type I and II errors will change with the population mean and SD, and with 'threshold value' in a sorting action, e.g., Fig. 1. In the following text, positive, P, refers to the total count of below-specification fruit, while TP and FP refer to the number of true positive and false positive assignments, respectively. Similarly, N refers to the total count of above-specification fruit, while TN and FN refer to the number of true negative and false negative assignments, respectively.

A receiver operating characteristic (ROC) curve is a plot of true positive rate (TP/P, or Sensitivity) against false positive (FP/N, or 1 - Specificity) rate (Ooms et al., 2010), and is typically generated by varying the threshold value used in the sorting operation (Fig. 2). This plot illustrates the trade-off between the success of the sorting operation in removing actual defective or poor fruit (TP/P) in terms of the cost of also removing good fruit (FP/N). An alternative, the detection error trade off (DET) graph plots detections of poor as good fruit (false negative rate, FNR, FN/P) against detections of good as poor fruit (false positive rate, FPR), using ordinate and abscissa scales transformed by the quantile function of the Normal distribution (the inverse of the cumulative Normal distribution) (Fig. 3). Both types of graphs report on the same underlying data, but provide different perspectives, and which is to be preferred depends on the intended message and the intended audience. As in much of statistics, there is seldom one single best summary statistic or graph to represent all aspects of a sorting operation. Indeed, there are many summary statistical metrics even for simple two-way classifiers, with no one metric being definitive. Common examples include Accuracy (ACC), Area Under the Curve (AUC), F1 score and the Matthews Correlation Coefficient (MCC).

In fruit sorting applications, decisions are often made based on assessment of very small population subsample sets, without accurate knowledge of the incoming population in terms of the number of good (G) and poor (P) fruit. What is generally known with good accuracy is the recovery rate of the sorting operation, the volume fraction of outgoing fruit to incoming fruit (P/(P + G). With that parameter and good estimates of the false positive and negative rates in the sorted bins, the Sensitivity and FPR can be calculated as:

$$Sensitivity = \frac{1}{\left(1 + \frac{PnG}{1 - GnP}\right)\left(\frac{R}{1 - R}\right)} \tag{10}$$

$$False\ Positive\ Rate = \frac{1}{\left(1 + \frac{GnP}{1 - PnG}\right)\left(\frac{1 - R}{R}\right)} \tag{11}$$

where PnG and NnP are sample estimates of the fractional number of fruit incorrectly sorted into the good (G) and poor (P) bins, respectively, and R is the total recovery rate in terms of the number of good fruit.

Given knowledge of the measurement error (i.e., SEP) for an NIRS-based prediction of an attribute and knowledge of attribute distribution (mean and SD) in a population, it is possible to model ROC and/or DET curves, as demonstrated in Fig. 3 for a simulated DM segregation. The simulation exercise involved calculation of a recovery rate and resampling of 50 fruit from each of the good and poor bins, to calculate the PnN and NnP parameters for a sorting threshold of 16.1 % FW. The data were generated from a simulated fruit DM population (of normal distribution with mean 17 % FW and variance 2 % FW) with added NIR prediction noise (Normal with mean 0 % FW and variance 1 % FW). The exercise reveals considerable spread in the reported data which must be due to sampling variation (Fig. 2). Further simulation studies showed clear benefits of increasing sample sizes for the good and poor bins (Fig. 2, bottom panel). In this example, if the sampling bin size increased from 50 to 200 fruit per sample, the average spread around the true or mean ROC or DET curve in the simulations decreased from near 10 % to below 5 %.

A sorting optimization curve (SOC) utilises additional information to optimise the selection of the threshold value (Ooms et al., 2010). For example, pricing data on the value of the categories of fruit after grading can be incorporated. For example, apple fruit with internal browning defects may have some value for juicing, and while fresh fruit command a higher value, there is a penalty cost if a consignment is rejected.

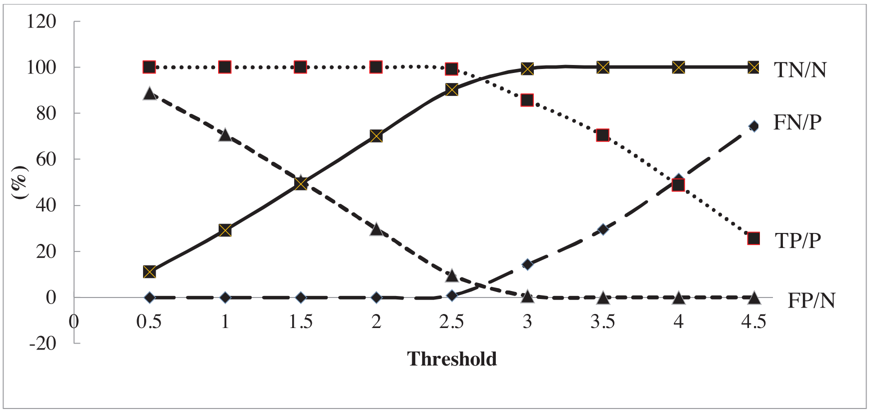

Fig. 1. Example of impact of change in threshold value in a sorting operation on apple internal browning in terms of % of good fruit correctly accepted (TN/N) and rejected (FP/N), and of defect fruit accepted (FN/P) and rejected (TP/P), for a population of mean score 3.1 and SD 1.4 (data from Khatiwadi et al., 2016). A 1–5 severity scale is used, in which scores of 1 and 2 are considered acceptable, and 3–5 as unacceptable).

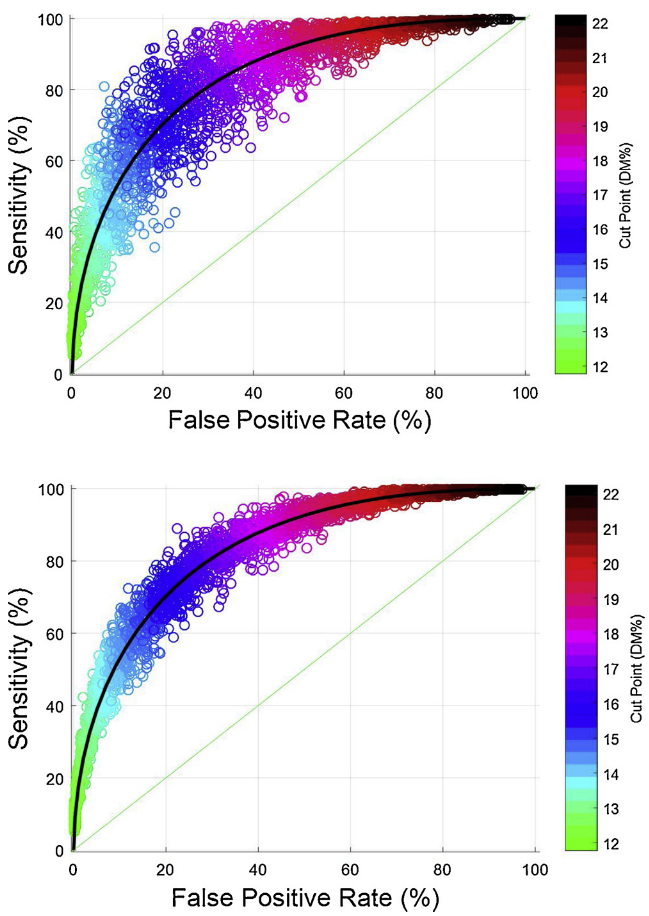

Fig. 2. ROC (Receiver Operating Characteristic curve: Sensitivity plotted against false positive rate) plots for a simulation of near infra-red spectroscopy (NIRS) sorting to a dry matter (DM) specification 16.1 % FW. Statistics were generated by random subsampling of 50 (top panel) or 200 (bottom panel) fruit from both the high and low sorted output bins. The colour scale represents the variation with change in the selected sorting threshold value (colour bar).

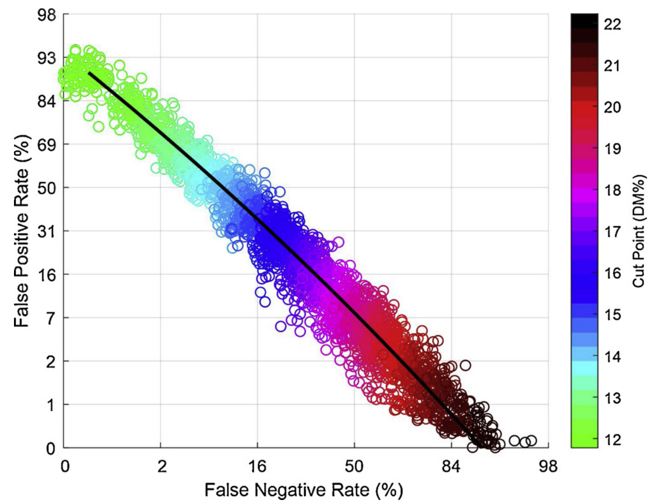

Fig. 3. DET (Detector Error Trade-Off curve: False negative rate plotted against false positive rate) (bottom panel) plot for a simulation of near infra-red spectroscopy (NIRS) sorting to a dry matter (DM) specification 16.1 % FW (data of Fig. 2). Statistics were generated by random subsampling of 200 fruit from both the high and low sorted output bins. The colour scale represents the variation with change in the selected sorting threshold value (‘cut point’, as shown in colour bar).

Greater use of ROC and SOC in the context of the use of NIRS in assessment of fruit populations is anticipated in future commercial sorting operations, e.g., in the setting of the threshold for sorting in the context of variation in population mean and SD, and measurement RMSEP.

7. Case studies

For a technology to be adopted, there has to be a “pull factor”, generally an increase in product quality and thus value for postharvest technologies. The following case studies explore examples in which NIRS use has created value in horticultural supply chains, leading to adoption of the technology in routine postharvest use. We believe these cases offer insight to postharvest researchers on the context of NIR technology adoption, and some of the drivers of, and restrictions to, adoption. This should inform the focus of future research effort. The cases draw on the personal experiences of the authors in the New Zealand, Australian, Chinese and Japanese markets, and a pers. comm. from Prof. Bart Nicolai (KU Leuven) on a European case.

7.1. Case study 1 – in-line NIRS sorting in Japan

While NIRS assessment of fruit was first developed in the USA, it was first effectively commercially implemented in Japan, with relatively widespread use on packing lines by the 1990s (Norris, 1992). Development of portable units followed in the early 2000s both by private companies (e.g., Kubota Co. Ltd) and public-private partnerships (e.g., Fantec Co., Ltd). This rapid uptake was made possible by several conditions unique to the Japanese industry at that time. With an ageing farmer base, labour constraints had driven the development of well-mechanised packhouses servicing co-operatives of growers. Large packhouses serviced producers who were essentially artisanal, focused on production of high quality, high value fruit. These packhouses were

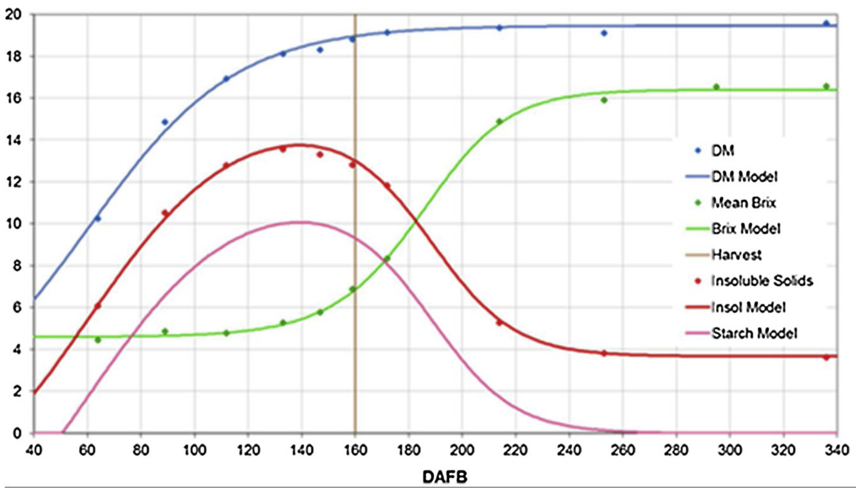

Fig. 4. Time course of actual dry matter (DM) and soluble solids content (SSC) (points) and modelled DM, SSC (Brix) values, insoluble solids and starch contents (lines) for kiwifruit fruit cv. Hayward on the vine. Commercial harvest occurred at 160 days after full bloom (DAFB) (top panel).

receptive to new technology, and generous government grants encouraged both the development of appropriate technology and the adoption by packhouses. NIR technology was therefore rapidly adopted to serve an existing market need, allowing branding of fruit with stickers carrying messages such as 'light sorted' and 'guaranteed sweet' to a consumer base conditioned to pay a premium price for a premium product.

The packing line technology was exported into Taiwan and Korea by the early 2000s, but uptake by other industrialised nations was deterred by the lack of a premium price for premium quality, and the lack of government support as existed in the Japanese market. These circumstances led to the development of lower cost packing line NIR technology, first in Australasia and then in Europe.

7.2. Case study 2 – in-line NIRS sorting for kiwifruit DM in New Zealand

NIRS fruit sorting has been commercially used in kiwifruit packing lines in New Zealand since 2015. The New Zealand kiwifruit exporter, Zespri Group Limited, enforces a minimum DM as a minimum taste standard (MTS) on all grower-supplied fruit populations. For a population failing the MTS, NIRS sorting is used to segregate out a high-DM subpopulation of fruit that exceed the MTS and so are able to be exported. This recovery of good high-DM fruit from otherwise poor populations has become increasingly attractive to the New Zealand industry as the more recently available and increasingly popular golden-fleshed cultivar, 'Zesy002' (Marketed as Zespri™®SunGold Kiwifruit), produces high yielding vines that can, unless carefully managed, produce a high proportion of low DM fruit. Packing facilities offering NIRS sorters have increased from two in 2015 to more than ten separate facilities in 2018, with many of millions of individual fruit examined and graded for export.

The market place for kiwifruit grading in New Zealand is competitive, with three major manufacturers having installed systems (Tomra-Compac, Auckland, New Zealand; MAF-Roda, Montabaun, France; and Greefa, The Netherlands). In addition, the market place for kiwifruit supply itself, involving independent packing facilities and/or grower co-operatives, is also very competitive. Commercial sensitivities prevent fully objective performance examinations of the NIRS grading systems; however, anecdotal evidence suggests the following:

1 NIRS grading systems are not as precise in prediction of individual kiwifruit DM as the science literature might suggest is possible. The relevant literature (e.g., McGlone and Kawano, 1998) suggests precision (SEP) for predicting DM at less than 0.5 % FW, whereas the anecdotal evidence from the private studies are of higher SEP, at least above 0.6 % and more often larger, at around 1 % FW.
2 The accuracy of the NIRS grading systems can be variable, with biases sometimes in excess of 1 % FW, and with slopes on the scatterplots of predictions vs actual values also shifting strongly away from the desired unity slope.
3 The NIRS grading systems require active management, in terms of regular bias adjustment and/or calibration updating, to maintain accuracy over multiple days and certainly weeks.

McGlone and Wohlers (2016) reported on auditing data from the 2016 season in terms of ROC. Sensitivity (true positive rate) and false positive rate refer to the respective conditional probabilities of correctly predicting a fruit is of low DM or incorrectly predicting a fruit is of high DM, respectively. The average sensitivity of the NIRS systems in operation at the time was a moderate 75 %, with a false positive rate of about 15 %. While those average figures were good, there was a great deal of spread in performance across the different submitted grower lines on the one grader (Fig. 5). This result is consistent with variation in NIRS prediction performance across the grower lines, i.e. the model

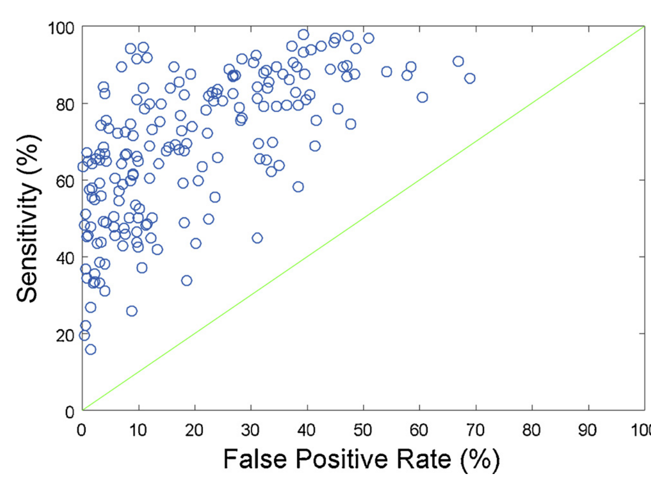

Fig. 5. Sensitivity and False Positive Rates generated from one near infra-red spectroscopy (NIRS) sorting facility in New Zealand during the 2016 kiwifruit harvest season. Each point represents statistics based on destructive sampling for oven-DM content assessment of n = 50 fruit from each of the high and low dry matter (DM) bins resulting from segregation of 92 separate grower lines on a DM cut-point of 16.1 % FW.

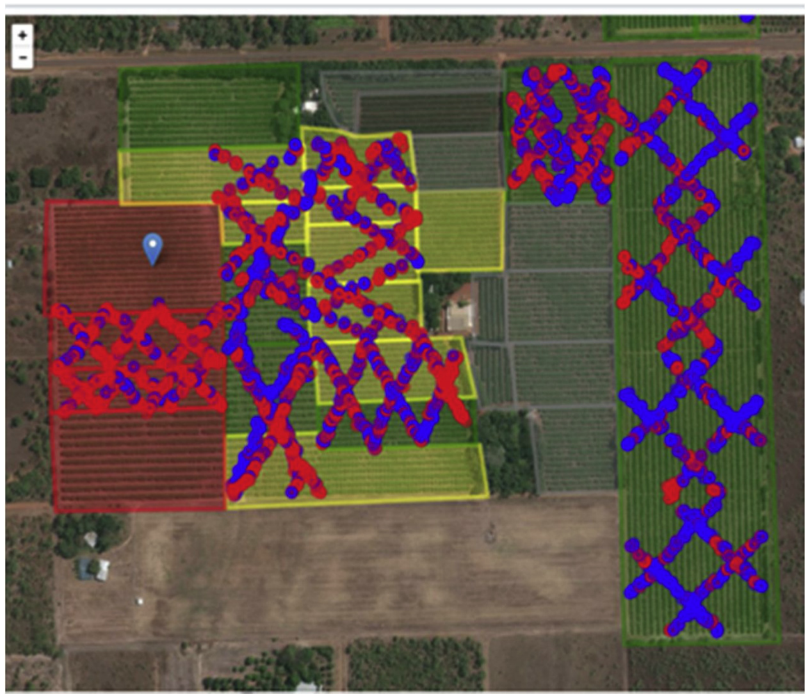

Fig. 6. Example of farm map tool for assisting in the decision to harvest mango fruit, with individual fruit measurements (dots) and the block average colour coded with respect to a user set specification on dry matter (DM).

was not robust, and contained impacts of population distribution and threshold setting on sorting errors.

7.3. Case study 3 – Australian mango maturity and eating quality

As noted earlier, mango fruit are climacteric, with conversion of starch to soluble sugars during ripening. Dry matter content acts as an index of starch and sugar contents, and so DM at the hard-green stage (i.e., at harvest) is strongly correlated to SSC and eating quality at the fully ripe stage (e.g., Subedi et al., 2007). The Australian Mango Industry Association (AMIA) has established DM specifications by cultivar required to deliver an acceptable eating quality assessment of ripe fruit (15 % FW for the dominant Australian cultivars, Henriod et al., 2015). The specifications have been adopted by major retailers (e.g. Woolworths, 2019).

In addition to use to ensure a minimum eating quality specification, fruit DM can be used to guide the decision on when to harvest. Harvest maturity is a user-defined concept, with latitude between physiological maturity (after which fruit are capable of ripening) and fruit drop (fruit ripened on tree, with no shelf life). The farm manager is required to make a decision on time to harvest to suit market need, with compromise between eating quality (increased carbohydrate reserves) and shelf life. Indices on fruit maturity are useful to assist in this decision.

Fruit DM increases reasonably linearly in the last 6 weeks of mango fruit maturation, except if large changes in soil moisture occur (Anderson et al., 2017). In contrast, kiwifruit DM reaches a plateau rather strongly towards harvest and so the attributes of SSC and flesh colour have more utility in harvest maturity estimation (Fig. 1, case study 1). Flesh colour is a good index of harvest maturity for mango, but while prediction of flesh colour can be achieved using Visible-NIRS, the prediction result is not as robust across new populations as that of DM (Subedi and Walsh, 2011).

The rate of increase in fruit DM under a given set of growing conditions is established through estimation of NIR-DM of fruit on the tree on at least two occasions. The fruit DM content associated with harvest maturity can be established for a given set of growing conditions, with maturity established by destructive assessment of flesh colour.

The net result is that specifications on fruit DM content can be useful for gauging both harvest readiness and potential eating quality. It is possible to have growing conditions that result in a fruit DM content at harvest maturity that is less than eating quality, in which case effort is required to change growing conditions to result in higher DM.

The Australian mango industry has adopted use of handheld NIRS for assessment of fruit DM content in the field to guide the decision to pick, rather than in packing lines to sort fruit after harvest. The handheld technology is also used in some packhouses that consolidate harvests from several growers, in central markets and in the receipt centres of large retailers. Large growers possess their own handheld NIRS units. AMIA provides the technology within an on-farm service to smaller growers, and for several years undertook assessments of incoming consignments in the major wholesale markets across Australia, with results published in a weekly industry newsletter to encourage compliance with the specification (https://www.industry.mangoes.net.au/; doa 15/1/2019). Major retailers also employ units to assist in specification checks.

The instrumentation and multi-season NIRS models in use have been demonstrated to be relatively robust to changes in ambient light and sample temperatures, and across growing locations and seasons, with a typical model prediction performance of bias-corrected RMSEP < 0.8 % FW and bias generally < 0.8 % FW. Multi-cultivar models result in decrease in predictive performance, but as the decrease is small for some combinations of cultivars, the compromise of ease of use over performance is accepted. Model performance is assessed relative to destructive oven-DM, both at the start of each season on an assembly of all available instruments, and during the season on an *ad hoc* basis. In the 2018-19 season a single model was used across most cultivars, with a multi-season model maintained on a master unit and transferred across instruments.

The on-line decision support tool (www.fruitmaps.com) is used in parts of the Australian industry. This tool displays geo-located NIR-DM measurements on a farm map, with values associated with farm management blocks, and attribute statistics and rate of increase displayed (Walsh and Wang, 2018; Fig. 6). The rate of increase (typically between 0.07–1.2% w/w per day) is established by time series measurements and used in a prediction of when the user-set specification on harvest maturity will be reached. This allows harvest planning, with ranking of orchard blocks by recommended harvest date.

7.4. Case study 4 – removing apple internal defects in Australia

Given consumer sensitivities, value chains often have strict criteria on the acceptable incidence rates for internal defects in fruit. For example, major retailers in Australia typically set a tolerance for internal browning in apple at an incidence level of 2 % (Walsh, 2014). Given the high financial penalty incurred when a large consignment is rejected by a retailer and must be sold on the central market, there has been a driver for adoption of packing line NIRS sorting technology.

A typical inspection procedure at the distribution centres of one retailer involves cutting of 30 fruit, followed by cutting of a second set of 30 fruit if a defect is found in the first sample. The presence of defective fruit in both samples results in rejection of the entire consignment. The statistical validity of use of two lots of 30 fruit is dependent on the proportion of defective fruit in the population (Eq. 7), with the sampled number constrained by the logistics of destructive sampling. For the acceptable limit of 2 % incidence, a sample of 30 fruit will contain a defect fruit in 0.45 of assessments, and in each of two lots in $(0.45 \times 0.45) = 0.20$ of assessments. Thus, the specification is effectively relaxed over the stated 2 % incidence level.

Packing line NIRS-based sorting to remove defective fruit can be used for an entire consignment, effectively lowering the incidence rate of defective fruit in the outgoing population. However, NIRS measurement involves an error on the assessment of an individual piece of fruit. For example, Khatiwadi et al. (2016) reported a classification accuracy [(True Positive + True Negative)/(Positive + Negative)] and a false discovery rate [False Positive/(True Positive + False Positive)] for independent validation population at > 95 % and < 2 %, respectively. The operator of the defect sorter has control of the threshold set point, with variation to optimise the number of defective fruit in the sorted accepted class against the number of good fruit in the sorted defect class (see Fig. 1). For example, for a defect scored 1–5 where 1 and 2 are acceptable scores, as the population mean decreases and SD increases, the threshold set point can be adjusted to a lower value to maintain the specification for defective fruit in the sorted accepted class.

7.5. Case study 5 – apple harvest time prediction in Belgium

Apple harvest maturity is related to starch pattern, firmness and SSC. Peirs et al. (2001) reported on the development of a model for estimation of optimum harvest timing (i.e., days to recommended harvest) of apple based on spectra of intact fruit. This work forms the foundation for a service that has been provided for 15 years to the Belgian apple industry for assessment of recommended harvest date, as described below (B. Nicolai, pers. comm.).

The NIRS model is based on absorbance spectra collected using a Zeiss Corona (45° reflectance configuration, using the range 1100–1700 nm. Spectra are used without further processing, preserving scattering information in the spectra. Spectra collected from fruit at a range of maturities are used in a PLSR model on time to harvest. Optimal picking date is estimated from a time course of the maturity attributes of colour, starch, firmness, acidity, SSC and size, with comparison to values from previous seasons. This service is offered to 100–150 orchards across the country, with sampled fruit despatched to a central laboratory for NIR spectra collection, and prediction of optimum picking date for each orchard based on existing models on the attribute of 'time to harvest',

A bias adjustment procedure was used in the initial years of implementation. In this procedure, fruit were collected on about 10 occasions through the season from each of three reference orchards located across the country, with collection of NIR spectra of the intact fruit and destructive assessment of maturity parameters (as above). SSC was predicted from the NIR spectra using a model based on a multi-season dataset. A bias (predicted – actual) on SSC was translated into a time bias using a graph of SSC as a function of time. A prediction of optimum picking date for individual orchards was then made using existing models on the attribute of 'time to harvest', subject to a bias correction for the season as estimated from the reference orchards. However, as the database increased across seasons, the model robustness increased to the point that bias correction was not necessary.

7.6. Case study 6 – kiwifruit disorders in New Zealand

Attempts have been made to predict kiwifruit storage disorders, particularly rots, directly from NIR spectra measured at harvest, for both the gold-fleshed cultivar 'Hort16A' (marketed as Zespri® Gold Kiwifruit) (Clark et al., 2004; Feng et al., 2011) and green-fleshed cultivar 'Hayward' (marketed as Zespri® Green Kiwifruit) (Feng, 2004; Feng et al., 2006, 2010). The general principle has been established that, within a single population, fruit identified as less mature have less risk of developing storage disorders. However, confounding effects related to harvest date and orchard origin differences appear to be problematic for deployment of a single globally applicable model. Attempts to exploit this principle continue, with attempts to use NIRS in estimation of future soft-fruit and rot incidence (Feng et al., 2011) using commercial NIRS graders (Zespri Group Limited, private communication). Research into storage prediction for kiwifruit continues in New Zealand under a mixture of industry and government funding, with a key challenge to solve being reduction of the very large false positive rate. This rate is estimated in one study to be as high as 40 % at a sensitivity of 75 % (Feng et al., 2016). Emphasis has been placed on prediction of a smorgasbord of maturity indices at harvest, such as DM, SSC, firmness and colour, which are combined to provide a storage prediction. However, beyond prediction of the simplest and largest component, DM, the commercial NIRS systems are probably inadequate for the task of maintaining accurate models over the lengthy harvest periods, which can extend many weeks. For example, SSC values are very low at harvest, between 4 and 8 % w/v, and the capability of current NIRS systems to model accurately in that range is limited.

7.7. Case study 7 – New Zealand kiwifruit breeding programme

The use of NIRS technology in a kiwifruit breeding programme provides value in the assessment of sample means by vine, rather than in individual fruit predictions (McGlone, 2009). A key challenge faced by the kiwifruit breeders is determination of when to harvest a new genotype, given that the young vine specimens yield a handful of fruit, perhaps as few as 10 fruit per vine in the first fruiting season. A non-destructive measurement is preferred to maximise the number of fruit that may be made available for other postharvest assessments, such as storage life and sensory quality. Handheld NIRS systems, such as the F-750 (Felix Instruments, Camas, WA, USA), have made non-destructive on-vine measurements possible. Any one individual measurement on a fruit for a new genotype is relatively inaccurate, with an SEP of around 1.2 % FW for DM, but in sampling a number of fruit on a vine, the accuracy can be significantly reduced, for example to about 0.7 % FW for a whole-vine average based on three fruit measurements. In addition, vine averages can be measured at regular intervals during the preharvest period, using the same fruit each time, enabling fine tuning of the harvest date for each particular vine. The method has been used in the New Zealand Institute for Plant and Food Research Limited/Zespri Group Limited kiwifruit breeding programme for over five years now, with five portable units in regular use during the season.

7.8. Case study 8 – apple water core in Japan and China

Watercore in apple is associated with the movement of water into

Fig. 7. Slight watercore in apples will disappear after storage (two panels to left), while serious watercore results in core browning after storage (two panels at right). In each pair of images, the left image is at harvest and the right image is after storage.

apoplastic spaces and is associated with high sugar contents. For this reason, the condition is termed' sweet core' in China. Affected fruit can achieve premium prices in East Asia based on their eating quality. The apoplastic water will be absorbed and the water core symptom will disappear in storage if the condition is slight, but if the condition is extensive, poor gas diffusion results in core browning as storage time increases (Fig. 7). This is a commercially undesirable outcome, and therefore it is important to predict a safe storage time for water-core apples.

Acceptable storage time for water core-affected apples can be predicted based on transmittance spectra of the fruit over the wavelength range 500 − 1100 nm, with absorbance spectra peak area related to effective storage life (Fig. 2) (Wang, 2007). Benchtop systems with a transmittance mode optical geometry are used by fruit distributors in Yuncheng, Shangxi Province, China. In this system, the area under the curve of the apple is related to the maximum acceptable storage time. A correct recognition rate of about 83 % was reported for serious water core-affected apples (Wang, 2007).

7.9. Case study 9 – sultana assessment

The Australian sultana industry remains competitive in the production of a premium sultana product despite the high cost of labour relative to those in competitor production areas. A premium is achieved for a uniform golden colour. Production cost was decreased by the introduction of mechanization of the drying and harvesting process. Previously, human-harvested fresh bunches were spread over wire racks for desiccation. Instead, vines are now defoliated, the bunches dried *in situ* on vine, and harvest is mechanized. Drying to a moisture content specification is critical for maximizing shelf life of harvested fruit, allowing an extended processing period, and for minimizing the energy requirements of further drying. Growers are paid on weight, moisture content and colour of fruit in field bins (approx. 500 kg) arriving at the processing plants.

The use of handheld NIRS to assess moisture content was implemented 2017, based on a temperature-compensated PLS model using four factors. Compared with the previous technology used (electricity resistivity of a minced sample), measurements are rapid and the instrument can be taken to the sample, rather than the reverse. This has enabled a better understand of variation, e.g., within a field bin as due to equilibration with the atmosphere of the surface layer of fruit.

8. Future needs

8.1. Improved instruments and operational protocols

The use of NIRS instrumentation in a packing facility is not a 'turn-key' type operation. While admittedly largely anecdotal, the best evidence from commercial practice is that packing line NIRS applications require a hands-on approach and intelligent attention by operators to maintain performance in terms of SEP and bias. There is a need for optimised calibration procedures, including the procurement and measurement of an appropriate range of calibration samples, the establishment of efficient quality control protocols and the incorporation of straightforward calibration updating procedures.

In the portable or handheld space there are now a moderate number of vendors. It is likely more vendors will enter the market as suitable hardware componentry continues to develop and become more economically viable from a developer/manufacturer perspective. There is a value proposition for use of this technology in assessing a sample of fruit ahead of harvest, when management decisions can be made that improve fruit quality (principally time of harvest), and also for use in quality control of consignments further down the value chain. These use cases tend to involve lower-cost instrumentation and less technically trained personnel than in the packing line application, reinforcing a need for efficient calibration and quality control protocols.

Furthermore, while the field of sorting statistics is well established academically, it remains to be practically developed and implemented in commercial postharvest activities.

8.2. Sensor fusion

The fusion of data from NIRS assessment with those from other technologies to yield information on an attribute relevant to postharvest management can exceed that possible with a single method. For example, Zou and Zhao (2005) reported on fusion of NIRS, machine vision and an electronic nose data for grading of apple fruit into two classes, Mendoza et al. (2012) reported on fusion of acoustic, fruit deformation under applied force, Vis-SWNIRS and spectral scattering in assessment of apple firmness, and Cortés et al. (2017) report on use of tactile sensing and NIRS in control of robotic gripper for selective fruit harvesting.

The promise of data fusion for assessment of a given attribute is high. For example, as indicated in case study 6, a combination of firmness, NIRS and image (machine vision) assessment on grader, and destructive at-line assessment of population subsamples to reveal biochemical properties may help inform and/or improve storage prediction. However, examples of practical deployment are elusive. The typical electronic pack-line on which a NIRS device might be employed only provides load cells and RGB camera capacities. Further, the successful integration of such diverse information and measures remains a challenge. Techniques involving deep neural networks and other nonlinear modelling (Artificial Intelligence) techniques are likely to see increased use for such applications.

Declaration of Competing Interest

Co-author Walsh acknowledges involvement in the development of the F750 instrument with Felix Instruments Inc. Authors Walsh, McGlone and Han acknowledge collaboration with manufacturers of inline grading equipment, MAF (Montabaun, France), Tomra (Auckland, NZ) and various grading companies in China and Japan, respectively.

Acknowledgements

KW acknowledges funding from the Australian Government Department of Agriculture as part of its Rural R&D for Profit program through Horticulture Innovation Australia ST19009. We thank Mark Wohlers of Plant and Food Research Limited for valuable discussions around the topic of sampling statistics.

References

Acharya, U.K., Subedi, P.P., Walsh, K.B., 2012. Evaluation of a dry extract system involving NIR spectroscopy (DESIR) for rapid assessment of pesticide contamination of fruit surfaces. Am. J. Anal. Chem. 3, 524–533. https://doi.org/10.4236/ajac.2012.38070.

Amodio, M., Ceglie, F., Chaudhry, M., Piazzolla, F., Colelli, G., 2017. Potential of NIR spectroscopy for predicting internal quality and discriminating among strawberry fruits from different production systems. Postharvest Biol. Technol. 125, 112–121. https://doi.org/10.1016/j.postharvbio.2016.11.013.

Anderson, N.T., Subedi, P.P., Walsh, K.B., 2017. Manipulation of mango fruit dry matter content to improve eating quality. Sci. Hortic. 226, 316–321. https://doi.org/10.1016/j.scienta.2017.09.001.

Birth, G.S., Dull, G.G., Renfroe, W.T., Kays, S.J., 1985. Nondestructive spectrophotometric determination of dry matter in onions. J. Am. Soc. Hortic. Sci. 110, 297–303.

Bollen, A.F., Prussia, S.E., 2009. Sorting for defects and visual quality attributes. In: Florkowski, W.J., Shewfelt, R.L., Brueckner, B., Prussia, S.E. (Eds.), Postharvest Handling (Second Edition). Academic Press, Boston, pp. 399–420. https://doi.org/10.1016/B978-0-12-374112-7.X0001-7.

Bureau, S., Cozzolino, D., Clark, C.J., 2019. Contributions of Fourier-transform mid infrared (FT-MIR) spectroscopy to the study of fruit and vegetables: a review. Postharvest Biol. Technol. 148, 1–14. https://doi.org/10.1016/j.postharvbio.2018.10.003.

Clark, C.J., McGlone, V.A., De Silva, H.N., Manning, M.A., Burdon, J., Mowat, A.D., 2004. Prediction of storage disorders of kiwifruit (*Actinidia chinensis*) based on visible-NIR spectral characteristics at harvest. Postharvest Biol. Technol. 32, 147–158. https://doi.org/10.1016/j.postharvbio.2003.11.004.

Cortés, V., Blanes, C., Blasco, J., Ortiz, C., Aleixos, N., Mellado, M., Cubero, S., Talens, P., 2017. Integration of simultaneous tactile sensing and reflectance visible and near-infrared spectroscopy in a robot gripper for mango quality assessment. Biosyst. Eng. 166 (1), 12–123. https://doi.org/10.1016/j.biosystemseng.2017.08.005.

Faul, F., Erdfelder, E., Lang, A.-G., Bucher, A., 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39, 175–191.

Feng, J., 2004. Segregation of' Hayward' Kiwifruit for Storage Potential. Ph.D. Thesis. Massey University, Palmerston North, New Zealand, pp. 250.

Feng, J., Maguire, K.M., MacKay, B.R., 2006. Discriminating batches of 'Hayward' kiwifruit for storage potential. Postharvest Biol. Technol. 41, 128–134. https://doi.org/10.1016/j.postharvbio.2006.03.010.

Feng, J., Kosasih, M., Keesing, R., Woolf, A., White, A., McGlone, A., Snelgar, B., Hall, A., 2010. Use of non-destructive measurements to define maturation and to predict softening of' Hayward' kiwifruit. Report to ZESPRI Group Ltd, Project PE1003. Plant & Food Research Client Report No: 31264, PFR SPTS No. 3571. PFR Contract No:23684. pp32.

Feng, J., McGlone, V.A., Currie, M., Clark, C.J., Jordan, R.B., 2011. Assessment of yellow-fleshed kiwifruit (*Actinidia chinensis*' Hort16A') quality in pre- and post-harvest conditions using a portable near-infrared spectrometer. HortScience 46, 57–63.

Feng, J., Wohlers, M., Olsson, S.R., White, A., McGlone, V.A., Seelye, R.J., Tanner, D., Bollen, F., 2016. Comparison between an acoustic firmness sensor and a near-infrared spectrometer in segregation of kiwifruit for storage potential. Acta Hortic. 1119, 279–288. https://doi.org/10.17660/ActaHortic.2016.1119.39.

Fornasini, P., 2008. The Uncertainty in Physical Measurements: an Introduction to Data Analysis. Springer-Verlag, New York, New York. https://10.1007/978-0-387-78650-6.

Garrido-Varo, A., Garcia-Olmo, Fearn T., 2019. A note on Mahalanobis and related distance measures in WinISI and the Unscrambler. J. Near Infrared Spectrosc. 27, 253–258. https://doi.org/10.1177/0967033519848296.

Harrison, R.L., 2010. Introduction to Monte Carlo simulation. In: AIP Conference Proceedings. Bratislava. pp. 17–21. https://doi.org/10.1063/1.3295638.

Henriod, R., Sole, D., Wright, C., Campbell, T., 2015. Mango quality standards. Horticulture Industry Australia Project MG14504 Final Report. HIA, Sydney, Australia.

Jacobs, B.A.J.G., Verlinden, B.E., Bobelyn, E., Decombel, A., Bleyaert, P., Van Lommel, J., Vandevelde, I., Saeys, W., Nicolai, B.M., 2016. Estimation of the prior storage period of lamb's lettuce based on visible/near infrared reflectance spectroscopy. Postharvest Biol. Technol. 113, 95–105. https://doi.org/10.1016/j.postharvbio.2015.11.007.

Jordan, R.B., Loeffen, M., 2013. A new method for modelling biological variation using quantile functions. Postharvest Biol. Technol. 86, 387–401. https://doi.org/10.1016/j.postharvbio.2013.07.008.

Khatiwadi, B., Subedi, P., Hayes, C., Cunha Carlos, L.C., Walsh, K.B., 2016. Assessment of internal flesh browning in intact apple using visible-short wave near infrared spectroscopy. Postharvest Biol. Technol. 120, 103–111. https://doi.org/10.1016/j.postharvbio.2016.06.001.

Li, X., Huang, J., Xiong, Y., Zhou, J., Tan, X., Zhang, B., 2018. Determination of soluble solid content in multi-origin' Fuji' apples by using FT-NIR spectroscopy and an origin discriminant strategy. Comput. Electron. Agric. 155, 23–31. https://doi.org/10.1016/j.compag.2018.10.003.

Magwaza, L.S., Opara, U.S., Terry, L.A., Landahl, S., Cronje, P.J., Nieuwoudt, H., Mouazen, A., Saeys, W., Nicolaï, B.M., 2012. Prediction of 'Nules Clementine' mandarin susceptibility to rind breakdown disorder using Vis/NIR spectroscopy. Postharvest Biol. Technol. 74, 1–10. https://doi.org/10.1016/j.postharvbio.2012.06.007.

Magwaza, L.S., Ford, H.D., Cronje, P.J.R., Opara, U.L., Landahl, S., Tatam, R.P., Terry, L.A., 2013. Application of optical coherence tomography to non-destructively characterise rind breakdown disorder of 'Nules Clementine' mandarins. Postharvest Biol. Technol. 84, 16–21. https://doi.org/10.1016/j.postharvbio.2013.03.019.

McGlone, V.A., 2009. Kiwifruit maturity screening by NIR; Dealing with genotype variation. In: Proceedings of 14th International Conference on Near Infrared Spectroscopy. Bangkok. pp. 251.

McGlone, V.A., Kawano, S., 1998. Firmness, dry-matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy. Postharvest Biol. Technol. 13, 131–141. https://doi.org/10.1016/S0925-5214(98)00007-6.

McGlone, V.A., Wohlers, M.W., 2016. NIR Sorting of Gold Kiwifruit During the 2016 Season. Client Report SPTS No.13971. Plant and Food Research, NZ.

McGlone, V.A., Fraser, D.G., Jordan, R.B., Künnemeyer, R., 2003. Internal quality assessment of mandarin fruit by vis/NIR Spectroscopy. J. Near Infrared Spectrosc. 5, 323–332. https://doi.org/10.1255/jnirs.383.

Mendoza, F., Renfu, L., Cen, H., 2012. Comparison and fusion of four nondestructive sensors for predicting apple fruit firmness and soluble solids content. Postharvest Biol. Technol. 73, 89–98. https://doi.org/10.1016/j.postharvbio.2012.05.012.

Nicolai, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theronb, K., Lammertyn, J., 2007. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. Postharvest Biol. Technol. 46, 99–118. https://doi.org/10.1016/j.postharvbio.2007.06.024.

Norris, K.H., 1992. Early history of near infrared for agricultural applications. NIR News 3, 12–13. https://doi.org/10.1255/nirn.105.

Ooms, D., Palm, R., Leemans, V., Destain, M.-F., 2010. A sorting optimization curve with quality and yield requirements. Pattern Recogn. Lett. 31, 983–990. https://doi.org/10.1016/j.patrec.2009.12.015.

Peirs, A., Lammertyn, J., Ooms, K., Nicolaï, B.M., 2001. Prediction of the optimal picking date of different apple cultivars by means of VIS/NIR-spectroscopy. Postharvest Biol. Technol. 21, 189–199. https://doi.org/10.1016/S0925-5214(00)00145-9.

Peirs, A., Tirry, J., Verlinden, B., Darius, P., Nicolaï, B., 2003. Effect of biological variability on the robustness of NIR models for soluble solids content of apples. Postharvest Biol. Technol. 28, 269–280. https://doi.org/10.1016/S0925-5214(02)00196-5.

Qin, J., Kim, M.S., Chao, K., Dhakal, S., Cho, B.K., Lohumi, S., Mo, C., Peng, Y., Huang, M., 2019. Advances in Raman spectroscopy and imaging techniques for quality and safety inspection of horticultural products. Postharvest Biol. Technol. 149, 101–117. https://doi.org/10.1016/j.postharvbio.2018.11.004.

Subedi, P.P., Walsh, K.B., 2008. Non-invasive techniques for measurement of fresh fruit firmness. Postharvest Biol. Technol. 51, 297–304. https://doi.org/10.1016/j.postharvbio.2008.03.004.

Subedi, P.P., Walsh, K.B., 2011. Assessment of sugar and starch in intact banana and mango fruit by SWNIR spectroscopy. Postharvest Biol. Technol. 62, 238–245. https://doi.org/10.1016/j.postharvbio.2011.06.014.

Subedi, P., Walsh, K.B., Owens, G., 2007. Prediction of mango eating quality at harvest. Postharvest Biol. Technol. 43, 326–334. https://doi.org/10.1016/j.postharvbio.2006.09.012.

Subedi, P.P., Walsh, K.B., Hopkins, D.W., 2012. Assessment of titratable acidity in fruit using short wave near infrared spectroscopy. Part B: intact fruit studies. J. Near Infrared Spectrosc. 20, 459–463. https://doi.org/10.1255/jnirs.1010.

Uwadaira, Y., Sekiyama, Y., Ikehata, A., 2018. An examination of the principle of non-destructive flesh firmness measurement of peach fruit by using VIS-NIR spectroscopy. Heliyon 4, e00531. https://doi.org/10.1016/j.heliyon.2018.e00531.

Walsh, K.B., 2005. Commercial adoption of technologies for fruit grading, with emphasis on NIRS. Information and Technology for Sustainable Fruit and Vegetable Production. FRUTIC 05, Montpellier, pp. 399–408. http://symposcience.lyon.cemagref.fr/exl-doc/colloque/ART-00001679.pdf.

Walsh, K.B., 2014. Chapter 9. Postharvest regulation and quality standards on fresh produce. In: Florkowski, W.J., Shewfelt, R.L., Brueckner, B., Prussia, S.E. (Eds.), Postharvest Handling – A Systems Approach, third edition. Academic Press, Boston, pp. 205–245. https://doi.org/10.1016/B978-0-12-408137-6.00008-9.

Walsh, K.B., 2015. Nondestructive assessment of fruit quality. In: Wills, R.B.H., Golding, J.B. (Eds.), Advances in Postharvest Fruit and Vegetable Technology. CRC Press, Florida, Amsterdam, pp. 40–61. https://doi.org/10.1201/b18489.

Walsh, K.B., Wang, Z., 2018. Monitoring fruit quality and quantity. In: Sauco, V.G., Lu, P. (Eds.), Achieving Sustainable Cultivation of Mangoes. Bureligh Dodds Publishing Limited, London, pp. 313–338. https://doi.org/10.19103/as.2017.0026.14.

Walsh, K.B., Subedi, P., Tijskens, P., 2015. A case study of a decision support system on mango fruit maturity. Acta Hortic. 1091, 195–204. https://doi.org/10.17660/ActaHortic.2015.1091.24.

Wang, J., 2007. Information Processing and Mathematic Modelling for Nondestructive Determination of Internal Quality of Apple and European Pear. PhD thesis. China Agricultural University.

Wang, H., Peng, J., Xie, C., Bao, Y., He, Y., 2015. Fruit quality evaluation using spectroscopy technology: a review. Sensors 15, 11889–11927. https://doi.org/10.3390/s150511889.

Woolworths, 2019. Produce Specifications. (accessed 22 January 2019). http://www.wowlink.com.au.

Zou, X., Zhao, J., 2005. Apple Quality Assessment by Fusion of Three Sensors. Sensors IEEE, Irvine, pp. 389–392. https://doi.org/10.1109/ICSENS.2005.1597717.