RESEARCH ARTICLE
doi: 10.2306/scienceasia1513-1874.2010.36.210

ScienceAsia 36 (2010): **210–215**

Long-term prediction of Zhonghua kiwifruit dry matter by near infrared spectroscopy

Qiang Lü[a,b], **Ming-jie Tang**[b], **Jian-rong Cai**[b,*], **Hua-zhu Lu**[b]

[a] School of Information Science and Engineering, Henan University of Technology, ZhengZhou 450052, China
[b] School of Food & Biological Engineering, Jiangsu University, Zhenjiang 212013, China

*Corresponding author, e-mail: jrcai66@gmail.com

Received 5 Mar 2010
Accepted 27 Aug 2010

ABSTRACT: Synergy interval partial least square (siPLS) was proposed to select efficiently the characteristic wavelength regions of dry matter against kiwifruit near-infrared spectra for dry matter prediction. Four data sets (NIR spectra and dry matter of unripe fruit (UU), NIR spectra of unripe fruit and dry matter of ripe fruit (UR), NIR spectra and dry matter of ripe fruit (RR), and UU&UR&RR) were obtained in the experiment. They were used to develop models for predicting dry matter of unripe and/or ripe kiwifruits. The results of cross-validation showed that the change of characteristic wavelength regions was caused by chemical conversion of organic compounds included in the dry matter at different storage periods of kiwifruits. Compared with the global spectra data models, the siPLS method could simplify the models with efficiently selecting characteristic wavelength regions. The root mean square error of cross-validation and correlation coefficient (r) of the UR model were 0.47% and 0.92, respectively, in calibration set. The root mean square error of prediction and r were 0.53% and 0.90, respectively, in the prediction set. This study demonstrated that NIR spectroscopy of unripe kiwifruits could predict the dry matter.

KEYWORDS: synergy interval partial least square (siPLS)

INTRODUCTION

Timing of the harvest is an important factor to the subsequent postharvest shelf life and fruit quality[1,2]. Dry matter (DM) in kiwifruit (*Actinidia deliciosa*) has been proposed as being both a maturity indicator for the proper time of harvest and also as a predictor of the sensory quality of the fruit once ripe[3–5]. It comprises both the soluble solids (largely fructose, glucose, and sucrose) and insoluble solids (mainly the structural carbohydrates and starch). Kiwifruit are harvested unripe, although physiologically mature, but must be put in natural storage to ripen (i.e., conversion of the stored starch into soluble solids) before consumption[6]. The kiwifruit DM is reasonably constant during ripening with only small losses due to respiration. It is dominated by the large carbohydrate component (around 75% of DM), most of which is sugar and starch at harvest, that becomes sugar when eaten ripe. Hence the DM indicates either the potential or actual sugar level of the fruit. The only reliable method to measure DM is by drying slices of fruit to drive off the water. But this is obviously destructive[5].

Near infrared (NIR) spectroscopy is a fast, accurate, and non-destructive technique that can be adopted as a replacement of individual labour skill and time-consuming methods. NIR spectroscopy has been used to grade fruits[7,8], predict the fruit maturity[9], and indicate the optimal harvesting time[10]. Kiwifruit is one commodity where sorting based on pre-selected NIR spectral features at harvest can be used to grade fruit on the basis of DM. Recent research has established that NIR spectroscopic analysis can be used to assess kiwifruit DM and/or ripened soluble solid content using spectra below 1100 nm for greater penetration[2,5,6,11,12]. The spectra, ranging from 10 000 to 4000 cm^{-1} (1000–2500 nm), were composed of overtones and combinations of fundamental vibrations of –CH, –NH, and –OH groups from the mid-infrared. Therefore, many spectral regions may contain useful information about the chemical vibrations of DM in the samples. So in this work, the NIR spectra (10 000–4000 cm^{-1}) were explored to predict DM in kiwifruit.

NIR spectral data calibrations have been made with the classical multivariate calibration analysis, e.g., partial least squares (PLS) regression. However, the fact that there might be spectral regions that do not contain any information about the chemical variations in the samples has not been addressed in the analysis. In fact, one of the major problems in multivariate data

analysis is to select appropriate spectral regions in order to achieve the best performance.

This study investigated and compared the results provided by PLS and siPLS procedures for NIR quantitative analysis of the DM in kiwifruit. Three specific objectives of this study were (1) to establish relationships between the NIR measurements and the DM of kiwifruit, (2) to optimize spectral intervals at different storage periods of kiwifruit, and (3) to predict the DM of ripe kiwifruit (edible phase) based on the NIR spectra collected when the kiwifruit was unripe.

MATERIALS AND METHODS

Sample preparation

The 221 '*Zhonghua*' kiwifruit samples used in this study were purchased from a farm in Zhouzhi, Shaanxi Province, China. All the kiwifruits stored for one month were sent to our laboratory in October 2008. They were stored at 1 °C, 90% RH. They were divided into two groups at random. In the first group (109 kiwifruits), 72 samples were taken as the calibration set and 37 samples were taken as the prediction set. In the second group (112 kiwifruits), 74 samples were taken as the calibration set and 38 samples were taken as the prediction set. Experiments were done under controlled conditions (20 °C, 68% RH). Before being examined by the NIR technique, kiwifruits were acclimatized to equilibrium for 12 h in controlled conditions. After the kiwifruits had been stored for one day, the NIR spectra and DM were determined for the first group. Only the NIR spectra was collected from the second group. The DM determination and the second NIR spectra collection of the second group were done when the samples were stored in a refrigerator (4 °C, 90% RH) for 30 days. The samples were equilibrated for 2–4 days in order to soften (edible phase) in the controlled conditions (20 °C, 68% RH) before the experiment of the second group.

Four data sets for analysis were created using combinations of these unripe and ripe fruit measurements. UU data set containing non-destructive NIR spectra and destructive DM measurements all made on unripe fruit. The sample number was 109. UR data set containing non-destructive NIR spectra made on unripe fruit, with corresponding destructive DM measurements after ripening. The sample number was 112. RR data set containing non-destructive NIR spectra and destructive DM measurements all made on ripe fruit. The sample number was 112. UU&UR&RR data set containing UU, UR and RR data set. The sample number was 333.

Spectra collection

The NIR spectra were collected in the reflectance mode using the FT-NIR spectrophotometer (AntarisII, Thermo Electron Co., USA) with an integrating sphere. Each spectrum was from the average of 32 scans. The range of spectra was from 10 000 to 4000 cm^{-1}, and the data were measured in 1.928 cm^{-1} intervals, which resulted in 3112 data points. The average of the three spectra, which were collected from the equator of each kiwifruit, was used in the sequence analysis.

Determination of kiwifruit DM

Fruit DM was measured by cutting two equatorial slices, approximately 3 mm thickness, and drying them at 65 °C to constant weight (approximately 24 h). DM (%) is actually 1 − water content. The kiwifruit DM was calculated from the final dry weight and initial wet weight of the slices, recorded as % fresh weight in this experiment[6]. The average of the DM of two slices was used in the sequence analysis. The maximum DM was 18.76%, the minimum was 13.53%, and the average was 16.20%. The standard deviation and coefficient of variation of the data were 1.08 and 6.69%, respectively.

Calibration model

The PLS models were first calibrated to predict the DM for the four data sets. In the application of the PLS algorithm it is generally known that the number of PLS components is a critical parameter in calibrating the model. The optimum number of PLS components is determined by the lowest root mean square error of cross-validation (RMSECV) and the root mean square error of prediction (RMSEP).

Synergy interval PLS (siPLS) were performed to extract near-infrared spectral intervals of kiwifruit DM due to the disadvantages of PLS. The siPLS algorithm used here was developed by Nørgaard et al[13]. First the data set was split into a number of intervals (variable-wise). Next, PLS regression models were established for all possible combinations of two, three, or four intervals. Thereafter RMSECV was calculated for every combination of intervals. The combination of intervals with the lowest RMSECV was then chosen. The number of intervals was also optimized according to RMSECV in the siPLS model calibration.

Software

All algorithms were implemented in MATLAB V7.0 (Mathworks) under Windows XP. RESULT Software (Antaris System, Thermo Electron Co., USA) was used in NIR spectral data acquisition. The siPLS algorithm used in this work was downloaded from www.models.kvl.dk/.

RESULTS AND DISCUSSION

Spectral preprocessing

Each mean spectrum was recorded as $\log(1/R)$, where R is the reflectance (Fig. 1). All the spectral data were analysed with a multiplicative scatter correction preprocessing technique used to correct for additive and multiplicative effects in the spectra [14].

PLS model results

Table 1 shows the results of the PLS model of four data sets. The correlation coefficient (r) was more than 0.86 in the calibration set and more than 0.85 in the prediction set. The result of maturity prediction (RR, edible phase) was the best; the correlation coefficient (r) was 0.94 in the calibration set and 0.91 in the prediction set. The results of long-term (UR) and hybrid prediction (UU&UR&RR) were slightly inferior. The obtained results were slightly inferior to the results of PLS models performed on some selected spectral regions by McGlone et al [2,5,6]. For this reason PLS was performed to the calibrate the global model on the full spectral region (10 000–4000 cm^{-1}), but some noisy spectral information inevitably weakened the modelling performance. It would also spend more computational time for more input variables. Therefore, in the following study, it was necessary to further optimize the model.

SiPLS model results

Table 2 shows the results of the siPLS model calibration when the spectra were split into different numbers of intervals. By optimizing the spectral intervals, the number of variables was reduced by more than 73%. The number of components was slightly reduced. The results of two groups (UU, UU&UR&RR) slightly fluctuated. The results of the other two groups (UR, RR) were slightly improved. For the UR data set, the optimal siPLS model was obtained with 15 intervals and 10 PLS components because the lowest RMSECV was 0.53% in Table 2. The optimal combinations of intervals selected contained numbers 3, 4, 8, and 12, corresponding to the spectral regions 4802.04–5201.14, 5203.07–5602.16, 6807.16–7204.33, and 8403.55–8800.71 cm^{-1} (Fig. 2). Compared with the full spectrum model, the siPLS model was simplified. The numbers of variables and components (830, 10) were fewer than those of the full spectrum model (3112, 12). For the optimal model, RMSECV was 0.47%, and the correlation coefficient (r) was 0.92 in the calibration set. When the performance of siPLS model was evaluated by the samples in the prediction set, RMSEP was 0.53% and the correlation coefficient (r) was 0.90 in the prediction set. So it is feasible to predict the DM of ripe kiwifruit (edible phase) based on the NIR spectra collected when the kiwifruit was unripe.

As can be seen from Table 2, the NIR spectra collected from unripe kiwifruit of the UU and UR data sets had the spectral intervals optimized by

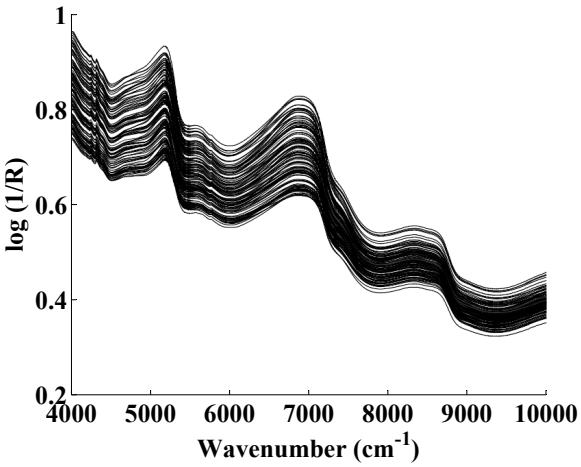

Fig. 1 NIR spectra of UR samples.

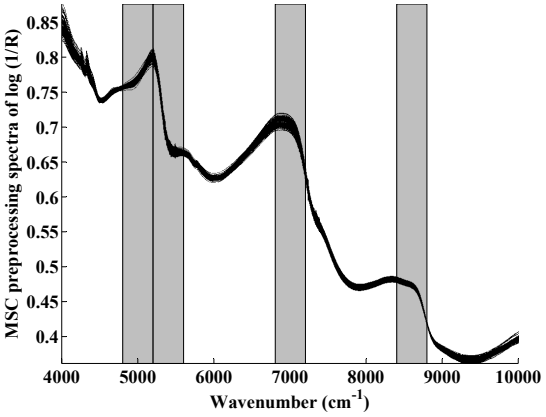

Fig. 2 Optimal spectral regions selected by siPLS with wavenumbers 4802.04–5201.14, 5203.07–5602.16, 6807.16–7204.33, and 8403.55–8800.71 cm^{-1} (UR).

Table 1 Results for PLS model of different data sets.

Data set	VN	CN	Calibration set		Prediction set	
			r	RMSECV/%	r	RMSEP/%
UU	3112	14	0.87	0.48	0.86	0.48
UR	3112	12	0.91	0.51	0.89	0.57
RR	3112	12	0.94	0.44	0.91	0.49
UU&UR&RR	3112	15	0.90	0.50	0.89	0.51

VN = variable number; CN = component number

Table 2 Best results of siPLS calibration model of different data sets.

Data set	IN	VN	Selected Intervals (cm^{-1})	CN	Calibration set		Prediction set	
					r	RMSECV/%	r	RMSEP/%
UU	14	667	5289.83–5717.84 7003.82–7429.91 8287.87–8713.95	11	0.86	0.48	0.86	0.49
UR	15	830	4802.04–5201.14 5203.07–5602.16 6807.16–7204.33 8403.55–8800.71	10	0.92	0.47	0.90	0.53
RR	19	656	4000.00–4314.26 4948.57–5262.83 7478.11–7792.37 8426.68–8740.95	11	0.95	0.39	0.94	0.41
UU&UR&RR	15	667	4000.00–4399.09 7605.35–8002.52 8403.55–8800.71 8802.64–9199.81	12	0.91	0.48	0.88	0.53

IN = interval number; VN = variable number; CN = component number

siPLS. These optimal spectral intervals had many cross-regions. It means that the DM did not basically change before and after maturing. The NIR spectra collected kiwifruit of the UU and RR data sets were optimized spectral intervals by siPLS. For the UU and UR data sets optimized spectra for ripe fruit there was only one cross-region indicating that some chemical components of the DM changed as the kiwifruit matured. As the kiwifruit matures, the starch, one component of the DM, is gradually converted into soluble saccharides (glucose, fructose, sucrose, etc.)[6]. Thus the characteristic wavelength regions of DM greatly changed.

From Table 1 and Table 2, it can be seen that the prediction accuracy of the siPLS model is not obviously superior to that of the PLS. Dry matter in kiwifruit is a mixture of soluble solids (largely fructose, glucose, and sucrose) and insoluble solids (mainly the structural carbohydrates and starch). The corresponding NIR bands are composed of overtones and combinations of fundamental vibrations of –CH, –NH, and –OH groups from the mid-infrared. Therefore, the relationship between the NIR spectra and dry matter may be complicated due to the presence of many relevant spectral regions. For the siPLS model, it was hard to select specific regions that contained all information of the DM of kiwifruit. Hence the results of siPLS models were not improved. Nevertheless, the siPLS model had fewer variables of the prediction model than PLS.

Table 3 and Table 4 list two cross-validation methods based on the siPLS model of different data sets. One was the siPLS calibration model with the optimal spectral intervals and optimal number of components of different data sets to predict the DM. For example, the spectral regions (4802.04–5201.14, 5203.07–5602.16, 6807.16–7204.33, and 8403.55–8800.71 cm^{-1}, obtained from UR model) of four data sets were used to calibrate four models (with 10 components, obtained from UR model) to predict DM

Table 3 Cross-validation of different siPLS calibration models (using different modelling parameter resources) for prediction sets.

Data set	UU		UR		RR		UU&UR&RR	
Parameter resource	r	RMSEP/%	r	RMSEP/%	r	RMSEP/%	r	RMSEP/%
UU	0.86	0.49	0.86	0.64	0.89	0.56	0.87	0.56
UR	0.83	0.52	0.90	0.53	0.92	0.48	0.87	0.55
RR	0.83	0.50	0.91	0.53	0.94	0.41	0.85	0.60
UU&UR&RR	0.79	0.57	0.87	0.61	0.91	0.50	0.88	0.53

Table 4 Cross-validation of different siPLS calibration models for prediction sets.

Data set	UU		UR		RR		UU&UR&RR	
Model	r	RMSEP/%	r	RMSEP/%	r	RMSEP/%	r	RMSEP/%
UU	0.86	0.49	0.82	0.71	0.82	0.93	0.77	0.73
UR	0.71	0.66	0.90	0.53	0.83	0.75	0.82	0.66
RR	0.34	4.87	0.62	4.81	0.94	0.41	0.32	3.95
UU&UR&RR	0.77	0.59	0.90	0.53	0.92	0.47	0.88	0.53

for every data set (Table 3). As can be seen from the statistical results, the siPLS could effectively select optimal spectral intervals according to the kiwifruit at different storage periods. Comparing with the UU, UR, and UU&UR&RR data sets, the results of the four models (calibrated by the parameters obtained from the four data sets) on the RR data set were the best. Therefore, the NIR spectra collected from unripe or ripe kiwifruit could effectively predict the DM of ripe kiwifruit.

The other way was to use the siPLS calibration model with different data sets to predict the DM. For example, the siPLS model based on UR data set was used to predict DM for every data set (Table 4). It shows that the NIR spectra were greatly changed as the hydrolysis of starch into soluble sugar occurred during storage. As the unripe kiwifruit contains more starch, the optical density is higher due to stronger light scattering. As the starch was converted into soluble sugars, the ripe kiwifruit becomes more transparent. The model UU worked to a limited degree on RR data set ($r = 0.82$, RMSEP $= 0.93$). The result of the model RR on UU data set was very poor ($r = 0.34$, RMSEP $= 4.87$). The reason was that the UU model trained on unripe fruit might have confounding problems due to starch and high optical density, and that might make the UU model more robust to handle DM in ripe fruit (such as little starch and more soluble sugars). However, the RR model trained on ripe fruit was less robust.

CONCLUSIONS

From the results, DM before and after maturing is reasonably constant with only small losses due to respiration, but it took place as the insoluble solids (i.e., starch) turned into soluble solids (i.e., fructose, glucose, and sucrose), leading to changes in NIR spectra. The siPLS could optimize spectral intervals at different storage periods of kiwifruit. In the siPLS model of four data sets, the number of variables was reduced by more than 70% and the number of components was slightly reduced. When the NIR spectra collected from unripe kiwifruit was used to predict the DM of ripe kiwifruit, the number of variables was 830, the number of components was 10, correlation coefficient (r) was 0.90, and the RMSEP was 0.53% in the prediction set. So it is feasible to predict the DM of ripe kiwifruit (edible phase) based on the NIR spectra collected when the kiwifruit was unripe. The siPLS model of RR data set achieved the best performance with 656 variables, 11 components, a correlation coefficient (r) of 0.94 and a RMSEP of 0.41% in the prediction set. Thus, NIR spectroscopy is a suitable tool for long-term prediction of the most important fruit inner qualities such as DM in kiwifruits.

Acknowledgements: This work was financially supported by the National High Technology Research and Development Program of China (863 Project, No. 2006AA10Z263) and the Key Natural Science Foundation of Jiangsu Province (Grant No. BK2006707-1). We are grateful to the website www.models.kvl.dk for their generous contribution of freely downloadable software for siPLS.

REFERENCES

1. Slaughter DC, Crisosto CH (1998) Nondestructive internal quality assessment of kiwifruit using near-infrared spectroscopy. *Semin Food Anal* **3**, 131–40.
2. McGlone VA, Clark CJ, Jordan RB (2007) Comparing density and VNIR methods for predicting quality parameters of yellow-fleshed kiwifruit (*Actinidia chinensis*). *Postharvest Biol Tech* **46**, 1–9.
3. Burdon J, McLeod D, Lallu N, Gamble J, Petley M, Gunson A (2004) Consumer evaluation of "Hayward" kiwifruit of different at-harvest dry matter contents. *Postharvest Biol Tech* **34**, 245–55.
4. Jordan RB, Walton EF, Klages KU, Seelye RJ (2000) Postharvest fruit density as an indicator of dry matter and ripened soluble solids of kiwifruit. *Postharvest Biol Tech* **20**, 163–73.
5. McGlone VA, Kawano S (1998) Firmness, dry-matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy. *Postharvest Biol Tech* **13**, 131–41.
6. McGlone VA, Jordan RB, Seelye R, Martinsen PJ (2002) Comparing density and NIR methods for measurement of kiwifruit dry matter and soluble solids content. *Postharvest Biol Tech* **26**, 191–8.
7. Budiastra IW, Ikeda Y, Nishizu T (1998) Prediction of individual sugars and malic acid concentrations of apples and mangoes by the developed NIR reflectance system. *J Jpn Soc Agr Machine* **60**, 117–27.
8. Steinmetz V, Roger JM, Molto E, Blasco J (1999) On-line fusion of colour camera and spectrophotometer for sugar content prediction of apples. *J Agr Eng Res* **73**, 207–16.
9. Peirs A, Scheerlinck N, Nicolaï BM (2003) Temperature compensation for near infrared reflectance measurement of apple fruit soluble solids contents. *Postharvest Biol Tech* **30**, 233–48.
10. Park B, Abbott JA, Lee KJ, Choi CH, Choi KH (2003) Near-infrared diffuse reflectance for quantitative and qualitative measurement of soluble solids and firmness of delicious and dala apples. *Trans Am Soc Agr Eng* **46**, 1721–31.
11. Clark CJ, McGlone VA, De Silva HN, Manning MA, Burdon J, Mowat AD (2004) Prediction of storage disorders of kiwifruit (*Actinidia chinensis*) based on visible-NIR spectral characteristics at harvest. *Postharvest Biol Tech* **32**, 147–58.
12. Schaare PN, Fraser DG (2000) Comparison of reflectance, interactance and transmission modes of visible-near infrared spectroscopy for measuring internal properties of kiwifruit (*Actinidia chinensis*). *Postharvest Biol Tech* **20**, 175–84.
13. Nørgaard L, Saudland A, Wagner J, Nielsen JP, Munck L, Engelsen SB (2000) Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy. *Appl Spectros* **54**, 413–9.
14. Chu XL, Yuan HF, Lu WZ (2004) Progress and application of spectral data pretreatment and wavelength selection methods in NIR analytical technique. *Progr Chem* **4**, 528–42.