Evaluation of sugar content in potatoes using NIR reflectance and wavelength selection techniques

Ahmed M. Rady *, Daniel E. Guyer

Department of Biosystems and Agricultural Engineering, 524 S. Shaw Lane, Michigan State University, East Lansing, MI 48824, USA

ABSTRACT

Near-infrared (NIR) diffuse reflectance has been extensively and successfully applied on quality assurance for fruits, vegetables, and food products. This study is principally aimed to extract the primary wavelengths related to the prediction of glucose and sucrose for potato tubers (of Frito Lay 1879 (FL), a chipping cultivar, and Russet Norkotah (RN), a table use cultivar, and investigating the potential of classification of potatoes based on sugar levels important to the frying industry. Whole tubers, as well as 12.7 mm slices, were scanned using a NIR reflectance spectroscopic system (900–1685 nm). To extract the most influential wavelength in the studied range, interval partial least squares (IPLS), and genetic algorithm (GA) were utilized. Partial least squares regression (PLSR) was applied for building prediction models. Prediction models for RN showed stronger correlation than FL with r(RPD) (correlation coefficient (ratio of reference standard deviation to root mean square error of the model)) values for whole tubers for glucose being as high as 0.81(1.70), and 0.97(3.91) for FL and RN; in the case of sliced samples the values were 0.74(1.49) and 0.94(2.73) for FL and RN. Lower correlation was obtained for sucrose with r(RPD) for whole tubers as high as 0.75(1.52), 0.92(2.57) for FL and RN; and the values for sliced samples were 0.67(1.31) and 0.75(1.41) for FL and RN respectively. Classification of potatoes based on sugar levels was conducted and training models were built using different classifiers (linear discriminant analysis (LDA), K-nearest neighbor (Knn), partial least squares discriminant analysis (PLSDA), and artificial neural network (ANN)), in addition to classifier fusion. To obtain more robust classification models for the training data, 4-fold cross validation was used and results were tested using separate sets of data. Classification rates of the testing set for whole tubers, based on glucose, were as high as 81% and 100% for FL and RN. For sliced samples, the rates were 83% and 81% for FL and RN. Generally, lower classification rates were obtained based on sucrose with values of whole tubers of 71%, and 79% for FL and RN, and for sliced samples the rates were 75%, and 82% which follows a similar trend as PLSR results. This study presents a potential of using selected wavelengths and NIR reflectance spectroscopy to effectively evaluate the sugar content of potatoes and classify potatoes based on thresholds that are crucial for the frying industry.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Potato is an important crop in developed and developing countries for human diet as a source of carbohydrates, protein, and vitamins (C, and B9) (Navarre et al., 2009). Processed potato products (French fries, chips, dehydrated, sliced, etc.) dominate the use of potatoes in developed countries. The US potato production for 2012 was 23.5 million metric ton, 61.1% used for processing, 25.7% used as table or fresh tubers, 5% used as seeds, and less than 1% were used for livestock feed (USDA-NASS, 2013). Among processed potato products, 8.46, and 6.11 million metric tons were used by French fries, and chips in 2012.

Quality of chips, and French fries drastically depend on several factors including dry matter that is strongly related to starch content, and specific gravity (Storey, 2007; Stark and Love, 2003; Storey and Davis, 1992). In addition, sugar content significantly determines the internal and/or external quality attributes of fried products. Glucose, and fructose are the major monosaccharide sugars in potato tubers with a concentration of 0.15–1.5%, and either one is considered a reducing sugar. Sucrose (0.4–6.6%) is a non-reducing disaccharide (Storey, 2007).

Sugar content of potato tubers varies between cultivars, and during storage time. Consequently, managing and monitoring

* Corresponding author.
  E-mail addresses: radyahme@msu.edu (A.M. Rady), guyer@msu.edu (D.E. Guyer).

sugar levels are crucial steps for tubers dedicated for processing. Excess levels of reducing sugars cause unacceptable non-enzymatic browning color for fried products due to the reaction between the reducing sugars and the α-amino groups of amino acids which is known as the Maillard reaction (Schallenberger et al., 1959). Maximum reducing sugar levels in potatoes for processing are 0.2–0.3% for chips, and 0.3–0.5% for French fries (Storey, 2007). More precise thresholds were stated by Stark, et al. (2003), as of 0.035% and 0.12% for chips, and French fries. Moreover, higher levels of sucrose cause sweetening flavor for fried products and boiled tubers. Such unacceptable sucrose levels are associated with low storage temperatures, and long storage periods (Storey, 2007; Blenkinsop et al., 2002). Maximum accepted sucrose level for processing at harvest is 0.15% for chips, or French fries, and the levels in storage are 0.15% and 0.10% for chips, and French fries respectively (Stark et al., 2003). Other studies recommended different sugars' thresholds for processing (Kadam et al., 1991; Smith and Davis, 1968).

Near-infrared spectroscopy (NIRS) has become a promising technology that has been extensively applied in quality control and monitoring in chemical, petrochemical, pharmaceutical, agricultural, and food industries. According to Burns and Ciurczak (2001), the use of NIR diffuse reflection for quantitative analysis of biological products is widely applicable. In the reflectance mode, the incident light penetrates the sample surface and a portion of such light passes within the sample tissue and is then reflected back, known as diffuse reflectance, and detected with information about the internal composition of the tubers (Chen, 1978). As a rapid, and/or noninvasive method, NIRS is suitable for on-line applications with it being less time consuming, more robust, more reproducible, and more cost effective than human labor or other laboratory destructive methods used for quality assurance purposes. Fruits and vegetables, as high moisture products and having a relatively large size, were not initially suitable for NIRS applications. After development of high performance hardware, quality measurements of intact fruits and vegetables using NIRS were made possible with interactance and transmission modes (Kawano, 2002). NIRS interactance was developed in a USDA laboratory at Beltsville by Conway et al. (1984), to measure human body fat. Later, the practice of NIR interactance in the agricultural field became more intensive.

Several studies were conducted for measuring sugars in potatoes using NIRS. Mehrubeoglu and Cote (1997), conducted a study to build calibration models for total reducing sugars for a Russet variety and a chipping variety. NIR transmittance (2050–2400 nm) was applied on sliced samples. The best PLSR calibration model was obtained using the cross validation technique, by which the coefficient of correlation r(root mean square error of calibration) for correlation model or r(RMSEC) values for the Russet variety were 0.77(0.0387%), and for the chipping variety the values were 0.80(0.0173). When merging both cultivars' data, the r(RMSEC) values were 0.52(0.040%). Hartmann and Büning-Pfaue, (1998), used the NIR diffuse reflectance (1100–2500 nm) mode to measure fructose, glucose, sucrose, total reducing sugars, starch, and crude protein in potatoes (homogenized samples). Using partial least squares regression (PLSR), the cross-validated models for starch, crude protein, glucose, fructose, sucrose, and the total reducing sugars resulted in r(root mean square error of prediction) or r(RMSEP) values of 0.96(0.28%), 0.93(0.06%), 0.84(0.041%), 0.94 (0.028%), 0.79(0.037%), and 0.91(0.061%) respectively. Other studies were utilized to estimate sugars in potatoes using VIS/NIR interactance (400–1000 nm) by Yaptenco et al. (2000), VIS/NIR interactance (400–1100 nm) by Chen et al. (2010), and NIR reflectance (850–2500 nm) by Haase (2011). While the previous studies provided good prediction performance, most did not include separate prediction data sets and did not include whole tubers for measurements. They also included sampling methods that required longer time than working with whole tubers or sliced samples that require relatively short preparation time.

Sorting of fruits and vegetables based on internal constituents is still a relatively new, and open point of research. Specific gravity and dry matter were used to indirectly recognize hollow heart infected tubers by Kang et al. (2008), based on the fact that low specific gravity tubers are more likely infected. Visible/near-infrared or VIS/NIR transmittance measurements (650–1000 nm) were acquired and PLSR calibration for specific gravity had r(RMSEP) values 0.87(0.0045%) and those values were 0.83 (0.0050) for prediction. Using dry matter to detect hollow heart based on the same principal as specific gravity, r(RMSEP) values for calibration model were 0.83(0.0062%) and for validation 0.80 (0.0067%). However, developing an on-line system for discarding tubers with higher sugar content than the allowable levels is not yet conducted.

Current technologies for measuring sugars in potatoes include: HPLC (high performance liquid chromatography), HPAEC (high performance anion chromatography), gas–liquid chromatography, and the YSI Analyzer that was invented by Yellow Springs Instruments (Yellow Springs Instrument, Yellow Springs, Ohio, USA). While these techniques provide acceptable accuracy for industry, they are destructive, time consuming, and are consequently not suitable for rapid monitoring or on-line sorting applications.

The proposed objectives of this study were to:

1. Apply NIR diffuse reflectance spectroscopy along with variable selection techniques to extract the wavelengths in the range (900–1965 nm) associated with best predicting glucose and sucrose in potatoes.
2. Develop calibration and prediction models of glucose and sucrose in potatoes using NIR diffuse reflectance spectroscopy based on the selected wavelengths.
3. Build classification models of potatoes based on sugar levels that are critical to chipping and French frying industries.

2. Materials and methods

2.1. Raw materials and sample preparation

Experiments were conducted in 2009 and 2011, and in both seasons, two common cultivars were used in the experiments, Frito Lay 1879 (FL) as a chipping cultivar, and Russet Norkotah (RN) which is usually used for baking. In the 2009 season, cultivars were hand-harvested from two locations; a research farm at Montcalm, MI. (sandy soil) for RN and FL, and additionally the MSU Muck experimental farm, Bath, MI. for FL. Samples were stored in three temperatures of 4, 7, and 10 °C. Tubers were then monthly sampled for experimentation starting in November, 2009 until April, 2010 (except at March) with a total number of 540 tubers from FL and 180 tubers from RN. In the 2011 season, both cultivars were obtained from a commercial production field (sandy soil) in Southwest Michigan. Two more storage temperatures (1 °C, and 13 °C) were added in order to obtain more uniform sugar distribution and simulate the various uses of potato tubers. In general, lower storage temperature is desired for cultivars that are used as seeds or for cooking, while relatively higher temperatures are used for chip cultivars. Tubers were first stored at 4 °C for three weeks and an initial electronic measurement was conducted. Tubers were then distributed over five different cold storage rooms with the following temperatures: 1, 4, 7, 10, and 13 °C. They were then sampled for experimentation starting in November 2011, and

each month until May 2012 (except at April) with a total number of 195 tubers from FL, and 75 tubers from RN.

In both seasons, tubers were cleaned prior to the imaging, and any defective samples were discarded. Several vine killing and harvest dates, different locations, and storing at different temperatures were used to obtain broad, and uniform, sugar distribution, rather than evaluating the growing condition, and other pre- and post-harvest practices that were conducted on tubers. Consequently, results representing different locations for FL were not compared.

Two types of samples were utilized for the electronic measurements: whole tubers, and a 12.7 mm thick slice which was obtained by cutting in a direction that is perpendicular to its longitudinal axis, starting from the stem end of the tuber. The tested slice was the third slice in the cutting routine. Each whole tuber was scanned only once, whereas each slice was scanned on both sides with each scan representing a sample measurement. The number of samples in the 2009 season was then 540, and 180 for FL, and RN in the case of whole tubers, and 1080, and 360 for FL, and RN in the case of sliced samples. In the 2011 season, there were 195 and 75 whole tubers for FL, and RN, whereas there were 390 and 150 sliced samples for FL, and RN respectively.

2.2. Wet chemistry experiments

After electronically scanning whole tubers, sliced samples were obtained as described in Section 2.1, and scanned. Each sliced sample was then immediately transferred into a plastic bag and placed in a foam box containing ice to maintain its freshness, and minimize any chemical deterioration during scanning the remaining samples. To ensure consistency between the slice electronic and wet chemistry measurements, a sufficient amount of potato tuber juice from the specific areas that had already been electronically tested was obtained by using a 25.4 mm cylindrical metal core borer to extract tissue primarily from the middle of the slice. This tissue was then put in a pre-sterilized 0.1985 kg Whirl–Pak filter bag, 9.5 × 18 cm (Nasco, Fort Atkinson, Wisconsin, USA). The bag was then hammered by hand using a 0.907 kg weight for juicing and then homogenized using a stomacher for 1 min. The juice was filtered by the Whirl–Pak filter bag and transferred with a pipette to a polystyrene tube with cap. This juice was stored at

$-20\,^\circ$C to reduce any variation of constituents and allow subsequent use and analysis of the juice at a later time.

The enzymatic approach is a common technique for measuring carbohydrates in agricultural products (BeMiller, 2010). Using the Megazyme sucrose/D-glucose assay procedure (Megazyme International Ireland Ltd Wicklow, Ireland), the concentrations of glucose and sucrose in the extracted juice, expressed as gram per gram fresh tuber weight or percentage, were measured as explained in Rady et al. (2014), using the following equations:

$$D - Glucose = \Delta A \times F \times 0.005 \times 0.1 \tag{1}$$

$$Sucrose = (\Delta B - \Delta A) \times F \times DL \times 0.0095 \times 0.1 \tag{2}$$

Where:
  D-Glucose: concentration of glucose (%).
  Sucrose: concentration of sucrose (%).
  ΔA: GOPOD (glucose oxidase/peroxidase) absorbance for D-glucose.
  ΔB: GOPOD absorbance for sucrose.
  F: factor to convert from absorbance to μg of D-glucose (nm^{-1}). F is calculated as follow:

$$F = \frac{100}{absorbance\ of\ the\ control\ sample}.$$

The control sample is prepared using the D-glucose standard solution provided with the Megazyme assay kit.

DL: Dilution ratio, 1 in case of Frito Lay 1879 and 10 in case of Russet Norkotah.

0.1: Unit conversion factor for glucose and sucrose to convert from g/L into %.

During the wet chemistry experiments, each tuber had 2 replicates, so each replicate was assigned to a slice side, while each whole tuber had the average value of both replicates.

2.3. NIR reflectance system

The NIR diffuse reflectance system used in this study is represented by a schematic diagram in Fig. 1. The light source probe tip and the detector tip were approximately 3 cm from the sample upper surface. An InGaAs spectrometer (model no. NIR512L-1.7T1,
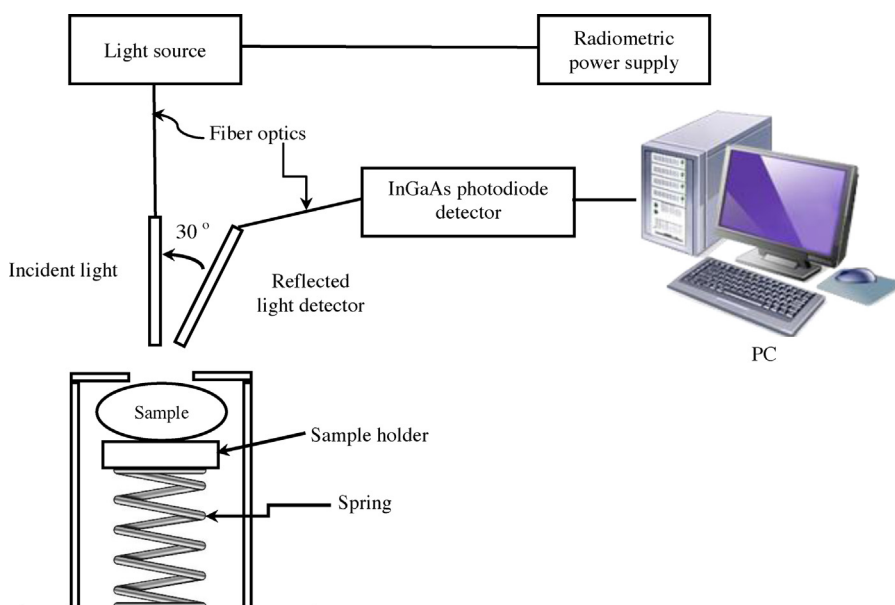

Fig. 1. Schematic representation of NIR reflectance system utilized in the study.

Control Development Inc. South Bend, IN, USA) with spectral resolution of 3.25 nm FWHM and linear dispersion of 1.625 nm/pixel was used in the reflectance mode, in the wavelength range of 900–1685 nm along with an Oriel radiometric power supply with a 300 W maximum power (model no. 68931, Oriel Inst., Irvine, CA, USA), and an Oriel light source (model no. 66881, Oriel Inst., Irvine, CA, USA) having 250 W maximum power, and with a quartz tungsten halogen lamp. In the diffuse reflectance experiments, the sample area covered by the light source had a diameter of 25.5 mm. The integration time was set as 4 ms, and each measurement is the average of 4 individual measurements. The incident light was directed on the middle area of the cut side of the slice. For the whole tubers, the light was directed on the surface approximately in the center area where the longitudinal, and perpendicular axes intersect. The detector covers an area on the sample surface of 12.7 mm diameter. Each sample spectrum was normalized using Teflon® as a reference material, and the calculation of the relative reflectance was done using the following equation:

$$\text{Relative reflectance} = \frac{I_s - I_b}{I_r - I_b} \qquad (3)$$

Where

I_s: intensity of reflected light for sample.
I_r: intensity of reflected light for reference (Teflon®).
I_b: intensity of reflected light for background.

2.4. Data analysis

2.4.1. Partial least squares regression (PLSR)

Partial least squares regression (PLSR) SIMPLS algorithm, presented by De Jong (1993), was applied in this study for building calibration and prediction models. To reduce the noise resulting from various sources, it is often advisable to pretreat spectroscopic data before building a calibration model (Varmuza and Filzmoser, 2009; Christy and Kvalhiem, 2007; Martens and Naes, 2001; Wold et al., 2001; Bjørsvik and Martens, 2001) and thus, two stages of preprocessing were implemented in this study for the spectral data. The first stage was the primary processing, which included separate evaluations of no-preprocessing, weighted baseline, smoothing with first derivative, smoothing with second derivative, normalization, standard normal deviate (SNV) correction, multiplicative signal correction (MSC), and median center. After these first stage preprocessing methods (tried separately), further preprocessing was applied to the spectral data during the PLSR algorithm development (Eigenvector Research, Inc. WA, USA), which included mean center or orthogonal signal correction (osc) (Wise et al., 2006).

Transformation of the reference data was also conducted to get the constituents' distribution as uniform as possible, which helps avoid overfitting in prediction models especially with the relatively low number of samples (Bjørsvik and Martens, 2001). Such transformations included the log and power transformation, with 2.0 as the exponent, in addition to using the non-transformed data, to study the effect of constituent value transformation.

Calibration and prediction sets of data were formed such that the calibration set contained 75% of the data and the prediction set contained 25% of the data. Cross validation (4-fold) was then applied on the calibration set of data to obtain the best calibration model based on the minimum value of root mean square of error of calibration using cross validation (RMSECcv) and the calibration model was subsequently applied to the separate prediction set to evaluate the model performance. It should be noted that the configuration of PLSR models was based on preliminary analysis to obtain the best prediction performance. The best prediction models have been selected based on the values of RMSEP, RPD, and the number of the latent variables (LVs).

2.4.2. Wavelengths selection

Interval partial least squares (IPLS) and genetic algorithm (GA) were applied in this study as wavelength selection techniques commonly used in the case of relatively high dimension data (i.e. spectroscopic data) to improve correlation robustness and computation time without considerable decrease in performance (Heise and Winzenm, 2002; Mark, 2001). Configurations of both IPLS and GA used in this study were based on preliminary results (not shown). The IPLS method is a known variable selection method for spectroscopic data and for optimizing the performance of PLSR models. IPLS uses sequential (either forward, backward or both) and exhaustive methods of search for the best subset of variables and different window width values (number of variables per window). In this research, forward mode, window width of 1, 2, and 3 variables, and number of latent variable for the PLSR model being 20, were used.

In the genetic algorithm (GA) technique, the RMSEC_{CV} was used as a model fitness scale, and window width of 1 variable, double crossover, maximum number of generations of 300, maximum number of latent variables of 20, and three iterations were used to obtain the wavelengths associated with best correlation between spectra, and reference values. More details regarding IPLS, and GA can be found in Leardi and Nørgaard, (2004); and Westad et al. (2013).

2.4.3. Classification of potatoes based on sugar levels

Classification of sliced samples and whole tubers based on sugar levels was conducted using several techniques. Linear discriminant analysis (LDA), K-nearest neighbor (Knn), partial least squares discriminant analysis (PLSDA), and artificial neural networks (ANN) were individually applied. Euclidean, as well as mahanalobis distances, were used in the current study for assigning each sample to the appropriate class. Prior to LDA classification, principal component analysis was conducted on the spectral data to overcome the problem of colinearity associated with spectroscopic measurements. To sustain the variance between samples, the first 20 components, responsible for >99% of the total variance, were used for classification task. In the case of Knn PLSDA and ANN, spectral data containing the selected wavelengths resulted from IPLS (window width or W = 1) were used. In the case of the Knn method, the Euclidean distance, and k values of 3 and 5 were selected based on preliminary analysis. Samples were then classified to the nearest class or to the class having the majority vote (Bishop, 2007; Duda et al., 2001). PLSDA has the advantages, as in PLSR, of colinearity tolerance and noise reduction (Wise et al., 2006). ANN classification was based on the feed forward neural network with back propagation (FFNN), commonly used in classification tasks, and

Table 1
Statistical summary of glucose (GL) and sucrose (SU) measurements based on wet chemistry for Frito Lay 1879 and Russet Norkotah cultivars in the 2009 and 2011 seasons.

Season	Statistics	FLGL (%)	FLSU (%)	RNGL (%)	RNSU (%)
2009	Minimum	0.0028	9.1e−5	0.0031	0.0045
	Maximum	0.1514	0.1607	0.3574	0.4205
	Mean	0.0457	0.0330	0.0591	0.10253
	Median	0.0391	0.0275	0.0338	0.0836
	Standard Deviation	0.0281	0.0239	0.0688	0.0806
	Skewness	1.4003	2.819	2.0412	1.2472
	Kurtosis	6.1725	17.1841	6.0141	4.6476
2011	Minimum	0.0224	7e−5	0.1544	3e−4
	Maximum	0.2394	0.4107	1.1355	2.1842
	Mean	0.0951	0.0765	0.5296	0.3012
	Median	0.0874	0.0605	0.5411	0.1760
	Standard Deviation	0.0470	0.0685	0.1926	0.3475
	Skewness	0.9401	1.5135	0.2741	2.2217
	Kurtosis	3.3262	5.6421	3.0054	9.8636

a

b

c

d

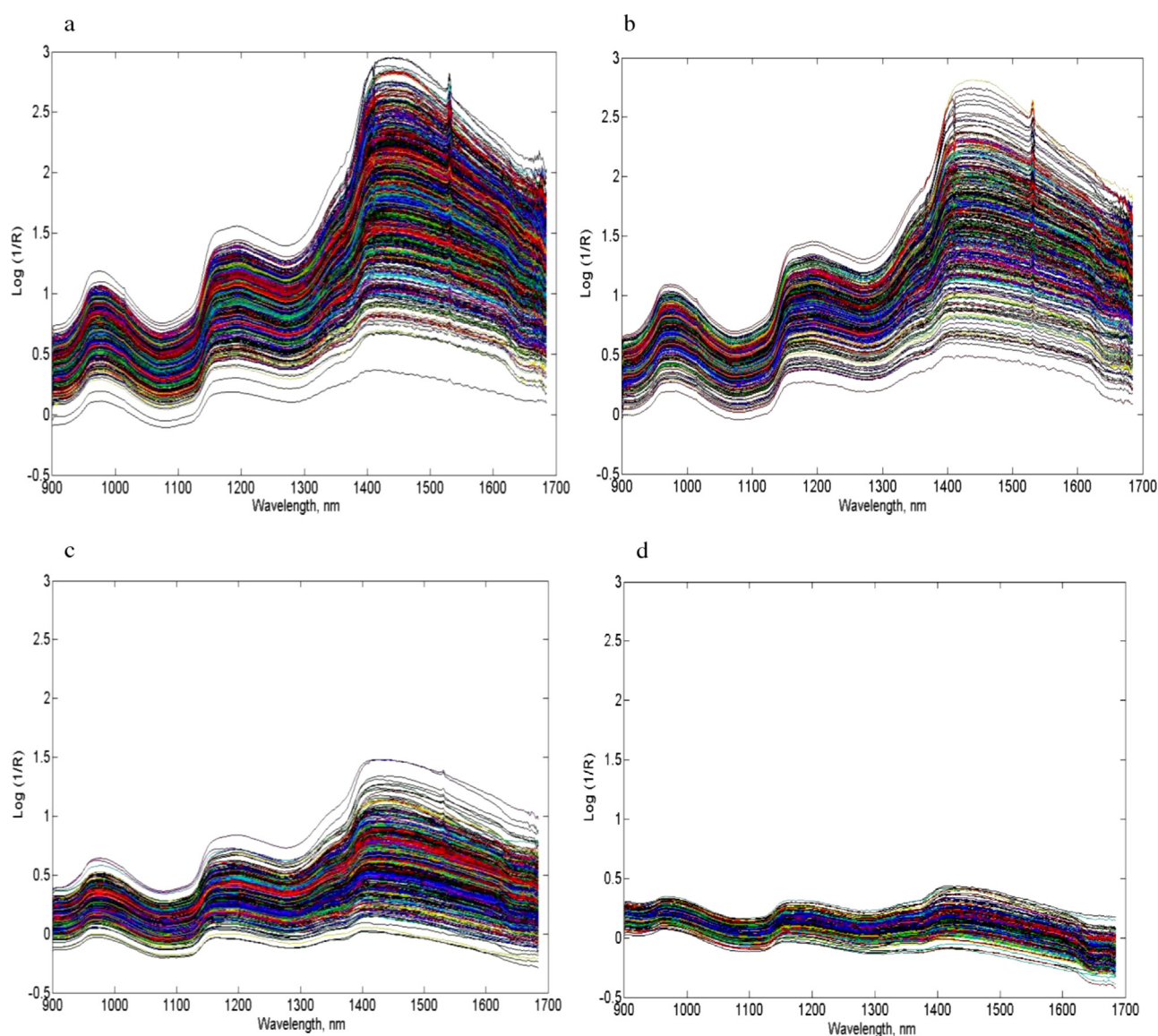**Fig. 2.** Signals of absorbance (log(1/reflectance)) of 2009 season data for sliced samples of (a) Frito Lay 1879, (b) Russet Norkotah, and whole tubers of (c) Frito Lay 1879, and (d) Russet Norkotah.

consisted of an input layer containing the pretreated spectral data, a hidden layer with 50 neurons chosen based on results obtained from preliminary analysis, and an output layer that contained the assigned class. Transfer functions were chosen as log-sigmoid, and scaled conjugate gradient back propagation for hidden and output layers respectively.

Spectral and reference data were preprocessed as mentioned in Section 2.4.1. In this study, samples were divided into a training set (80%), and a testing set (20%). A 4-fold cross validation technique was conducted on the training set to increase the robustness of the training models for all classification methods. Samples in both seasons were divided into two classes based on the cut-off glucose values in the 2009 season of 0.035% for both FL and RN, whereas the values for sucrose were 0.03% and 0.10%. In the 2011 season, and based on sugar distribution, the threshold values for glucose were 0.09% and 0.5% for FL and RN, while the values for sucrose were 0.08% and 0.15%. Cut-off levels were adopted from recommended thresholds listed by Stark and Love (2003), for

both sugars except for the glucose level for RN which was chosen to create two balanced classes. Classification of samples using LDA and ANN was conducted using the Matlab® statistical toolbox. For Knn and PLSDA, the classification toolbox for Matlab created by Davide Ballabio (Milano Chemometrics and QSAR Research Group, University of Milano-Bicocca, Milan, Italy) was used, and the PLS routine used to compute PLSDA was written by Frans W.J. van den Berg (Quality & Technology group, section Spectroscopy and Chemometrics, Department of Food Science, University of Copenhagen, Denmark).

Classifier fusion was also conducted in attempt to increase the overall classification accuracy. Weighted majority voting was used for setting each sample in the correct class. In majority voting, based on results obtained from individual classifiers, PLSDA was given the highest weight of 0.40, and weights of 0.20, 0.10, and 0.15 were given to LDA, and Knn and ANN respectively. Each sample was assigned to the class having the highest total voting resulted from all classifiers.

3. Results and discussions

3.1. Reference distribution

The basic statistics for glucose and sucrose obtained from the wet chemistry experiments are shown in Table 1 after discarding outliers and other experimental error values (<0). According to Storey (2007), glucose, and sucrose contents in potatoes are as high as 1.5% and 6.6% respectively. Thus, and based on the obtained sugar concentrations, the cut-off values for glucose, and sucrose were chosen as 1.5% and 2.5%. In the case of sucrose, the cut-off level was chosen as there were relatively few samples (less than 10 samples for both cultivars) that had values higher than 2.5%. Thus, they were considered outliers. Skewness resulted in both seasons especially in the case of sucrose even though the experiment was designed to minimize such. Maximum values of glucose and sucrose obtained from the 2011 season were higher than values in 2009 which is a direct result of the lower temperature (1 °C) added to the 2011 season.

3.2. Diffuse reflectance spectra

The relative values of absorbance (log(1/reflectance)) resulting from the 2009 season NIR measurements in the case of sliced samples and whole tubers are shown in Fig. 2a–d for both Frito Lay 1879 and Russet Norkotah. For the sliced samples, similar trend of signals for FL and RN was observed with multiple water absorption peaks around 970, 1200, and 1450 nm (Workman and Weyer, 2008). Moreover, another absorption peak, in both cultivars, is noted at 1530 nm which is a possible indication of an OH polymeric group located in starch (Workman and Weyer, 2008). In general, signals collected from whole tubers showed less absorption than sliced samples especially for RN which yielded a more condensed response in the raw signals than FL. The possible explanation for such a result is the thicker periderm layer of RN that reduces the captured diffuse reflectance signals. It is worth stating that the absorption peaks for sliced samples (around 970, 1200, and 1450 nm) were noted in the whole tubers in both cultivars with

more clarification in FL. Another effect of the thick skin for RN was the absence of the absorption peak at 1530 nm and the relatively slight presence, compared to sliced samples, of the same peak in FL samples. Results of the 2011 season, not shown, showed similar trend as in the 2009 season, with more condensed raw signals for both cultivars in the case of sliced samples and whole tubers, which refers to less variation between samples.

3.3. Partial least squares regression (PLSR) using all or selected wavelengths

Results for calibration and prediction models of glucose and sucrose using all and selected wavelengths for FL and RN cultivars in the 2009 and 2011 seasons are shown in Table 2 and Table 3 respectively (only the best models are shown). Williams (2007), illustrated that in NIR spectroscopy, correlation coefficient (r) values of 0.81-0.90 can be used for screening, and approximate calibration, values of 0.91–0.95 can be used for most applications including research, values of 0.96–0.98 are workable for most industrial applications, and finally $r \geq 0.99$ are suitable for any application. In the case of RPD, values of 1.5–2.0 refers to the model capability to differentiate between high and low constituent values. Whereas values of RPD in the range of 2.0–2.5 means a possibility of coarse prediction of reference values. Values of RPD of 2.5–3.0 or higher can be used for good and excellence prediction respectively (Nicolai et al., 2007). Results of PLSR showed general improvement of prediction performance when comparing models obtained using all versus selected wavelengths. For the 2009 season and in the case of sliced samples, glucose prediction models, obtained using IPLS selected wavelengths, yielded r(RPD) values of 0.74(1.49) and 0.94(2.73) for FL, and RN. Sucrose models showed less correlation performance for both cultivars with r(RPD) values of 0.50(1.15), and 0.41(1.01) for FL, and RN respectively. In the case of whole tubers, a similar trend of results was obtained with r(RPD) values for glucose models being 0.73(1.46), and 0.95(2.95) for FL, and RN. For sucrose prediction models, r(RPD) values were 0.47 (1.13), and 0.40(1.02) for FL, and RN. In the 2011 season, general improvement of prediction performance was obtained, especially

Table 2
Best performance PLSR results for predicting glucose and sucrose of potato tubers using NIR reflectance mode (sliced samples and whole tubers) obtained from all wavelengths for Frito Lay1879 and Russet Norkotah cultivars in the 2009 and 2011 seasons.

Season	Sample type	CV_Y	Sample size	Preprocessing	Calibration model			Prediction model		
					r_{cal}	$RMSEC_{cv}$ (%)	LVs	r_{pred}	RMSEP (%)	RPD
2009	Slice	FL_{GL}	1007	A_4,B_1,C_2	0.73	0.0574	15	0.72	0.0508	1.44
		FL_{SU}	981	A_2,B_1,C_0	0.51	0.0311	18	0.36	0.0311	1.06
		RN_{GL}	336	A_7,B_1,C_2	0.96	0.0589	18	0.90	0.0567	2.32
		RN_{SU}	334	A_2,B_1,C_0	0.50	0.0786	20	0.44	0.0729	1.11
	Whole	FL_{GL}	503	A_4,B_1,C_0	0.77	0.0231	20	0.73	0.0199	1.44
		FL_{SU}	495	A_3,B_2,C_0	0.57	0.0332	6	0.36	0.0316	1.04
		RN_{GL}	167	A_7,B_2,C_2	0.99	0.0542	16	0.97	0.0333	3.95
		RN_{SU}	155	A_2,B_1,C_2	0.40	0.1214	14	0.37	0.1126	1.08
2011	Slice	FL_{GL}	390	A_0,B_1,C_0	0.74	0.0511	13	0.61	0.0476	1.25
		FL_{SU}	364	A_2,B_1,C_0	0.62	0.0630	12	0.61	0.0494	1.26
		RN_{GL}	150	A_3,B_1,C_2	0.93	0.0662	17	0.66	0.0839	1.26
		RN_{SU}	140	A_1,B_1,C_2	0.72	0.2028	9	0.37	0.2065	0.96
	Whole	FL_{GL}	193	A_4,B_2,C_0	0.97	0.0452	19	0.71	0.0441	1.28
		FL_{SU}	193	A_2,B_1,C_0	0.66	0.0523	13	0.66	0.0471	1.34
		RN_{GL}	75	A_2,B_1,C_2	0.82	0.0754	12	0.77	0.0665	1.56
		RN_{SU}	74	A_5,B_1,C_2	0.76	0.1700	7	0.75	0.1324	1.52

Nomenclature: CV_Y (CV = Cultivar, Y = Constituent); A_x: First stage spectrum preprocessing; B_x: Second stage spectrum preprocessing; A_0: No preprocessing; B_1: Mean center; A_1: Weighted baseline; B_2: Orthogonal signal correction; A_2: 1st derivative; C_x: Reference data preprocessing; A_3: 2nd derivative; C_0: No reference transformation; A_4: Normalization; C_1: Log reference transformation; A_5: Standard normal variate (SNV); C_2: Power reference transformation; A_6: Multiplicative signal correction (MSC); A_7: Median center; Preproc.: Optimal preprocessing; R_{cal}: correlation coefficient for calibration set of data; $RMSEC_{cv}$: root mean square error of calibration set of data using cross validation; LVs: number of latent variables selected for the calibration model; R_{pred}: correlation coefficient for prediction set of data; RMSEP: root mean square error of predation set of data; RPD: ratio of reference standard deviation to root mean square error of the model.

Table 3
Best performance PLSR results for predicting glucose and sucrose of potato tubers using NIR reflectance mode (sliced samples and whole tubers) and selected wavelengths using IPLS and GA for Frito Lay1879 and Russet Norkotah cultivars in the 2009 and 2011 seasons.

Selection	Season	Sample type	CV_Y	Number of wavelengths	Preprocessing	Window width	Calibration model			Prediction model		
							r_{cal}	$RMSEC_{cv}$ (%)	LVs	r_{pred}	RMSEP(%)	RPD
IPLS	2009	Slice	FL_{GL}	84	A_4,B_1,C_0	W3	0.72	0.0218	20	0.74	0.0193	1.49
			FL_{SU}	68	A_4,B_1,C_0	W2	0.57	0.0306	18	0.50	0.0286	1.15
			RN_{GL}	123	A_0,B_1,C_0	W1	0.96	0.0332	20	0.94	0.0353	2.73
			RN_{SU}	40	A_0,B_1,C_0	W2	0.61	0.0800	20	0.41	0.0662	1.01
		Whole	FL_{GL}	48	A_0,B_1,C_0	W2	0.72	0.0215	20	0.73	0.0197	1.46
			FL_{SU}	48	A_4,B_1,C_0	W2	0.52	0.0304	20	0.47	0.0292	1.13
			RN_{GL}	78	A_4,B_2,C_0	W2	0.95	0.0330	20	0.95	0.0327	2.95
			RN_{SU}	16	A_4,B_1,C_0	W3	0.56	0.0847	20	0.40	0.0657	1.02
	2011	Slice	FL_{GL}	114	A_4,B_3,C_0	W3	0.86	0.0425	17	0.74	0.0403	1.48
			FL_{SU}	55	A_0,B_1,C_0	W1	0.75	0.0567	17	0.67	0.0476	1.31
			RN_{GL}	102	A_4,B_3,C_2	W2	0.98	0.0577	20	0.90	0.0468	2.25
			RN_{SU}	52	A_0,B_2,C_2	W2	0.91	0.1575	20	0.75	0.1414	1.41
		Whole	FL_{GL}	53	A_0,B_1,C_0	W1	0.91	0.0351	20	0.81	0.0333	1.70
			FL_{SU}	52	A_1,B_2,C_0	W2	0.85	0.0427	20	0.75	0.0346	1.52
			RN_{GL}	30	A_6,B_2,C_2	W2	0.96	0.0517	15	0.97	0.0266	3.91
			RN_{SU}	36	A_0,B_2,C_2	W2	0.93	0.1572	20	0.92	0.0790	2.57
GA	2009	Slice	FL_{GL}	116	A_4,B_1,C_2		0.69	0.0591	12	0.70	0.0523	1.40
			FL_{SU}	107	A_4,B_1,C_0		0.47	0.0319	10	0.38	0.0305	1.08
			RN_{GL}	128	A_4,B_1,C_2		0.93	0.0609	13	0.87	0.0644	2.04
			RN_{SU}	105	A_0,B_1,C_0		0.45	0.0836	8	0.23	0.0708	0.95
		Whole	FL_{GL}	128	A_4,B_2,C_0		0.74	0.0226	19	0.68	0.0211	1.36
			FL_{SU}	123	A_0,B_1,C_0		0.52	0.0323	20	0.35	0.0315	1.05
			RN_{GL}	169	A_4,B_2,C_0		0.97	0.0542	18	0.94	0.0326	2.85
			RN_{SU}	125	A_4,B_2,C_0		0.74	0.0720	4	0.53	0.0710	1.14
	2011	Slice	FL_{GL}	105	A_7,B_1,C_0		0.69	0.0499	10	0.52	0.0516	1.16
			FL_{SU}	95	A_4,B_1,C_0		0.55	0.0631	8	0.52	0.0530	1.17
			RN_{GL}	113	A_7,B_2,C_2		0.85	0.0756	9	0.65	0.0802	1.31
			RN_{SU}	104	A_5,B_1,C_2		0.67	0.1975	6	0.42	0.1858	1.07
		Whole	FL_{GL}	131	A_4,B_1,C_0		0.74	0.0481	10	0.72	0.0386	1.46
			FL_{SU}	91	A_4,B_1,C_0		0.64	0.0539	9	0.59	0.0506	1.25
			RN_{GL}	116	A_5,B_1,C_2		0.95	0.0770	12	0.77	0.0718	1.55
			RN_{SU}	101	A_4,B_1,C_2		0.90	0.1693	11	0.71	0.1444	1.41

Nomenclature: CV_Y (CV = Cultivar, Y = Constituent); A_x: First stage spectrum preprocessing; B_x: Second stage spectrum preprocessing; A_0: No preprocessing; B_1: Mean center; A_1: Weighted baseline; B_2: Orthogonal signal correction; A_2: 1st derivative; C_x: Reference data preprocessing; A_3: 2nd derivative; C_0: No reference transformation; A_4: Normalization; C_1: Log reference transformation; A_5: Standard normal variate (SNV); C_2: Power reference transformation; A_6: Multiplicative signal correction (MSC); W1: Window size of 1 wavelength; A_7: Median center; W2: Window size of 2 wavelengths; W3: Window size of 3 wavelengths; R_{cal}: correlation coefficient for calibration set of data; $RMSEC_{cv}$: root mean square error of calibration set of data using cross validation; LVs: number of latent variables selected for the calibration model; R_{pred}: correlation coefficient for prediction set of data; RMSEP: root mean square error of predation set of data; RPD: ratio of reference standard deviation to root mean square error of the model.

for sucrose models, possibly due to the broader and more uniform sugar distribution obtained in such season compared with the 2009 season. In the case of sliced samples, glucose prediction models showed r(RPD) values of 0.74(1.48), and 0.90(2.25) for FL, and RN. Whereas, sucrose models yielded r(RPD) values of 0.67 (1.31), and 0.75(1.41) for FL, and RN. In the case of whole tubers, considerable enhancement of prediction was achieved with r(RPD) values for glucose models being 0.81(1.70) for FL, and 0.97(3.91) for RN. Sucrose prediction models yielded r(RPD) values of 0.75(1.52), and 0.92(2.57) for FL, and RN respectively. The best relationships between the measured, and predicted sugar values for FL, and RN for sliced samples and whole tubers in the 2011 season is shown in Fig. 3.

The improvement of results for whole tubers compared to sliced samples, especially for sucrose models in the 2011 season, is possibly a result of the sugar distribution inside tubers. According to Kumar and Ezekiel (2004), Rama and Narasimham (2003) and Rastovski et al. (1987), sugars inside potato tubers tend to concentrate more on the vascular ring than on other tuber parts. Consequently, the diffuse reflected light is expected to hold information of the tissue closer to the skin than to the pith. Thus, measuring the glucose and sucrose content for whole tubers using a rapid NIR device or system should be conducted on the above referred area.

Results also showed that prediction models of glucose and sucrose using selected wavelengths obtained using IPLS yielded better performance than GA models for both cultivars. Moreover, Table 3 indicates that the selected wavelengths using IPLS were less than GA in all models. Consequently, there is more likelihood for over fitting to occur in models built using GA than IPLS when fewer number of samples than variables occurs which was noted in the 2011 season compared to 2009 season (Wise et al., 2006).

Due to the lower sugar concentration for FL than RN, results showed less correlation of prediction models for FL than RN in the case of glucose and sucrose as the detection of certain chemical substances using spectroscopic systems increase with the concentration.

It is worth stating that the prediction of glucose and sucrose for potatoes using selected wavelengths and NIR diffuse reflectance was not previously published and the prediction results obtained
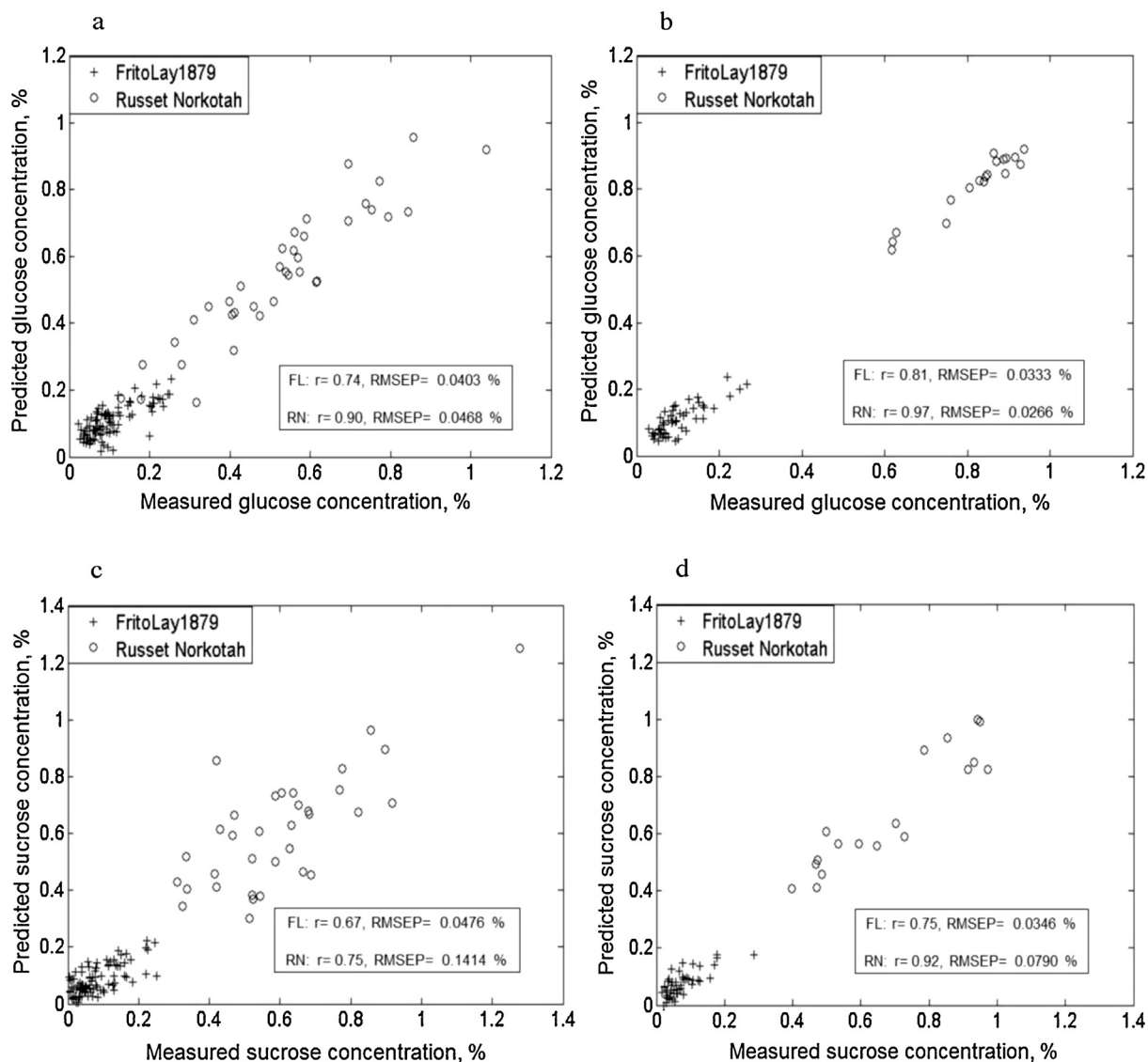
Fig. 3. Best prediction models based on selected wavelengths and PLSR predicted constituents in the 2011 season for Frito Lay 1879 and Russet Norkotah for (a) Glucose for sliced samples, (b) Glucose for whole tubers, (c) Sucrose for sliced samples, and (d) Sucrose for whole tubers.

in this study by PLSR are comparable with others reported by Hartmann and Büning-Pfaue, (1998), on homogenized samples (RMSEP = 0.041% and 0.037% for glucose and sucrose); Yaptenco et al. (2000), on whole tubers (RMSEP = 0.087% and 1.473% for glucose and sucrose); or Haase, (2011), on aliquots samples (SEP = 0.0389%, and 0.0966% for reducing sugars and sucrose). Sampling times in this study are significantly lower than that for all previous studies except for Yaptenco et al. (2000), which did not include a separate prediction data set. This study also confirms the results obtained by Rady et al. (2014), in which a potential investigation of measuring glucose and sucrose of potatoes was shown using different techniques and strong correlation for glucose was achieved for sliced samples (RMSEP = 0.0515%, and 0.0786% for FL, and RN), and whole tubers (RMSEP = 0.0620%, and 0.1529% for FL, and RN) using VIS/NIR interactance spectroscopy. However, no variable selection was applied. Other studies (i.e. Kawano et al., 1993) addressed the soluble solids content, in °B (Brix), and not the individual sugar content. Soluble solids content is not an accurate indication of individual sugars.

Additionally, combining the data sets, including all wavelengths, of the 2009, and 2011 seasons was also conducted and the prediction models were obtained using PLSR. Results (not fully presented) yielded slightly better results in most models. The values of r(RPD) for glucose models were 0.77(1.58) and 0.95(3.31) for FL and RN in the case of sliced samples, and those values were 0.75(1.51) and 0.95(3.33) in the case of whole tubers. For sucrose prediction models, improvement was shown for sliced samples with r(RPD) values of 0.54(1.19) and 0.67(1.34) for FL and RN. In the case of whole tubers, no improvement was achieved where r(RPD) values were 0.52(1.17) and 0.56(1.20) for FL and RN.

3.4. Classification of potatoes based on sugar levels

Spectral and reference data for sliced samples and whole tubers were divided into two classes based on glucose and sucrose thresholds as described in Section 2.4.3 for the 2009 and 2011 seasons with the number of class 1 (sugar level < the threshold) or class 2 (sugar level > the threshold) presented in Table 4. It is worth stating that outliers of sugar concentrations were eliminated as indicated in Section 3.1. The highest classification rates of training and testing groups obtained for slice samples and whole potato tubers of FL and RN cultivars based on glucose and sucrose concentrations are shown in Table 4 with bold fonts to mark the technique(s) producing the best classification rates in the testing

Table 4
Highest classification rates of glucose and sucrose for Frito Lay 1879 and Russet Norkotah for sliced samples and whole tubers using NIR reflectance mode and sliced samples and whole tubers in 2009 and 2011 seasons.

Season	Sample type	CV_Y	Number of samples		Preprocessing for LDA; Knn; PLSDA; ANN; combined classifier	Classification rate for training set (%)				Best classification rate for testing set (%)	
			Class 1	Class 2		LDA	Knn	PLSDA	ANN	Rate	Method(s)
2009	Slice	FL_{GL}	445	562	A_5; A_2; A_1; A_7; A_4	79	64	81	78	83	LDA
		FL_{SU}	523	458	A_4; A_4; A_3; A_5; A_7	63	55	64	56	63	LDA
		RN_{GL}	177	159	A_1; A_1; A_0; A_4; A_6	84	62	94	70	81	LDA, PLSDA
		RN_{SU}	195	139	A_3; A_3; A_6; A_0; A_0	68	53	67	56	68	Combined classifiers
	Whole	FL_{GL}	222	281	A_4; A_4; A_7; A_5; A_4	72	70	81	76	81	Combined classifiers
		FL_{SU}	266	229	A_0; A_2; A_7; A_0; A_7	65	57	70	57	62	Combined classifiers
		RN_{GL}	88	79	A_0; A_2; A_0; A_5;A_4	88	75	88	75	83	LDA
		RN_{SU}	87	68	A_3; A_0; A_4; A_0; A_4	69	64	69	54	69	Knn PLSDA, combined classifiers
2011	Slice	FL_{GL}	204	186	A_4; A_5; A_0; A_5; A_0	73	64	88	66	82	LSDA, combined classifiers
		FL_{SU}	218	146	A_4; A_1; A_4; A_0; A_1	69	55	85	64	75	LDA
		RN_{GL}	66	84	A_7; A_5; A_4; A_7; A_4	81	65	87	70	77	PLSDA, combined classifiers
		RN_{SU}	58	82	A_7; A_1; A_0; A_1; A_0	80	64	89	64	82	PLSDA, class
	Whole	FL_{GL}	136	57	A_7; A_0; A_0; A_0;A_0	81	57	97	74	77	LDA, combined classifiers
		FL_{SU}	122	71	A_5; A_0; A_5; A_9;A_4	83	54	79	69	71	LDA, PLSDA
		RN_{GL}	31	44	A_0; A_0; A_1; A_5;A_0	98	60	100	73	100	PLSDA, combined classifiers
		RN_{SU}	26	48	A_0; A_7; A_4; A_0;A_4	92	71	73	73	79	PLSDA, combined classifiers

Nomenclature: CV_Y (CV = Cultivar, Y = Constituent); A_x: First stage spectrum preprocessing; A_0: No preprocessing; A_1: Weighted baseline; A_2: 1st derivative; A_3: 2nd derivative; A_4: Normalization; A_5: Standard normal variate (SNV); A_6: Multiplicative signal correction (MSC); A_7: Median center.

groups. Classification performance generally followed the PLSR trend stated in Section 3.2 in the 2009 and 2011 seasons. For the 2009 season, classification rates of glucose-based models for the sliced samples (83% for FL and 81% for RN), were similar to values obtained for whole tubers (81%, and 83% for FL and RN). Sucrose-based classification models, however, yielded lower performance for the sliced samples (63% and 68% for FL and RN), and whole tubers (62% and 69% for FL and RN). Following the PLSR results, classification results for glucose in the 2011 showed similar performance compared to 2009 results for the sliced samples (82% and 77% for FL and RN) and even better results in the case of whole tubers, especially for RN (77% and 100% for FL and RN). Moreover, significantly enhanced classification rates were obtained for sucrose models in the case of sliced samples (75% and 82% for FL and RN), and whole tubers (71%, and 79%).

In general, LDA, PLSDA, and classifier fusion yielded better classification results than other techniques. Such trend is a result of the capability of the PLSDA technique, as illustrated in Section 2.4.3, for treating data with colinearity problems, and the application of PCA analysis on spectral data prior to performing classification using LDA. Combined classifiers also resulted in better classification than Knn and ANN classifiers, and slightly similar to results obtained by PLSDA, and LDA.

Broader sugar distribution in the 2011 season is strongly believed to have resulted in better classification results, compared with the 2009 season, especially for sucrose, which follows the same trend obtained in PLSR results. Classification of potatoes based on sugar levels and using noninvasive measurements was not addressed previously in literature and results showed the potential for eliminating tubers with sugar content that is not suitable for frying, allowing these tubers to potentially be reconditioned to reduce sugar content (Sowokinos, 2007). Enhancing classification outputs obtained in this study is feasible by developing a broader sugar distribution, increasing the number of samples, and using kernel-based classification methods (i.e. soft independent modeling of class analogy or SIMCA, Gaussian mixture models, and support vector machines or SVM).

4. Conclusions

IPLS and GA were tested to select the most discerning influencing wavelengths from spectral data obtained by measuring diffuse reflectance of potato tubers and 12.7 mm slices in the NIR region (900–1685 nm). Frito Lay 1879 and Russet Norkotah potato cultivars were used in this study, and glucose and sucrose were ground truth measured using the enzymatic approach. PLSR was utilized to build calibration and prediction models for glucose and sucrose. Selected wavelengths were found to have strong correlation performance for sliced samples with RMSEP of 0.0193%, and 0.0353% for FL and RN in the case of glucose. In the case of sucrose, the best models had RMSEP values of 0.0286% and 0.0662% for FL and RN respectively. Prediction models obtained from whole tubers yielded similar performance for glucose to sliced samples with RMSEP values of 0.0197%, and 0.0327% for FL and RN in the case of glucose, while those values for sucrose were 0.0295%, and 0.0657% for FL and RN.

Monitoring sugar content in tubers dedicated for processing is an important quality assurance step to prevent non-enzymatic browning after frying. Levels of RMSEP obtained in this study are less than thresholds stated in section 1 by Stark et al., 2003. Thus, there is a possibility to more rapidly track sugar levels, especially for whole tubers, which is a crucial practice during storage, and prior to processing. Classification of whole tubers based on sugar levels is considered important to the frying industry and was shown to have feasible application for sorting, especially in the case of glucose in which the classification rate values for testing sets were as high as 81%, and 100% for FL, and RN, and those values were 71%, and 79% for sucrose. Classification rates for sliced samples were similar to those obtained for whole tubers. Performance of classification models can possibly be improved with broader and more uniform distribution of sugars, and scanning the whole tuber in more than one point on the tuber surface so that more robust prediction and classification is feasible. Moreover, to simulate real sorting conditions, it is important to conduct more experiments on moving tubers mixed with clods, and using tubers that have soil attached to their surfaces. Building a

sorting system for potato tubers based on sugar levels is still in the research stage and more work is needed to build up a robust system with performance needed from growers and processors. These results suggest it may be possible to make a handheld device for monitoring and classifying potatoes based on sugars and using whole tubers or sliced samples as the sample preparation for the latter case is very small.

Acknowledgements

References

BeMiller, J.N., 2010. Carbohydrate analysis, In: Nielsen, S.S. (Ed.), Food Analysis. 4th Edn. Springer, NY, USA, pp. 147–177.

Bishop, C.M., 2007. Pattern recognition and machine learning. Springer Science+ Business Media. LLC, NY, USA, pp. 67–136.

Bjørsvik, H.-R., Martens, H., 2001. Data analysis: calibration of NIR instruments by PLS regression, In: Burns, D.A., Ciurczak, E.W. (Eds.), Handbook of Near-Infrared Analysis. 2nd Edn. Marcel Dekker, Inc., NY, USA, pp. 185–208.

Blenkinsop, R.W., Copp, L.J., Yada, R.Y., Marangoni, A.G., 2002. Changes in compositional parameters of tubers of potato (Solanum tuberosum) during low-temperature storage and their relationship to chip processing quality. J. Agric. Food Chem. 50 (16), 4545–4553.

Burns, D.A., Ciurczak, E., 2001. Handbook of Near-Infrared Analysis, 2nd Edn. Marcel Dekker, Inc., NY, USA.

Chen, J.Y., Zhang, H., Miao, Y., Asakura, M., 2010. Nondestructive Determination of sugar content in potato tubers using visible and near infrared spectroscopy. Jap. J. Food Eng. 11 (1), 59–64.

Chen, P., 1978. Use of optical properties of food materials in quality evaluation and materials sorting. J. Food Process. Eng. 2 (4), 307–322.

Christy, A.A., Kvalhiem, O.M., 2007. Latent-variable analysis of multivariate data in infrared spectrometry. In: Ozaki, Y., McClure, W.F., Christy, A.A. (Eds.), Near-Infrared Spectroscopy in Food Science and Technology. Wiley-Interscience, USA, pp. 145–162.

Conway, J.M., Norris, K.H., Bodwell, C.E., 1984. A new approach for the estimation of body composition: infrared interactance. Amer. J. Clin. Nutr. 40 (6), 1123–1130.

De Jong, S., 1993. SIMPLS: an alternative approach to partial least squares regression. Chemom. Intell. Lab. Syst. 18 (3), 251–263.

Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern Classification, 2nd Edn. John Willey & Sons, Inc., NY, USA, pp. 215–281.

Haase, N.U., 2011. Prediction of potato processing quality by near infrared reflectance spectroscopy of ground raw tubers. J. Near Infrared Spectrosc. 19 (1), 37–45.

Hartmann, R., Büning-Pfaue, H., 1998. NIR determination of potato constituents. Potato Res. 41 (4), 327–334.

Heise, H.M., Winzenm, R., 2002. Fundamental chemometric methods. In: Siesbr, H. W., Ozaki, Y., Kawata, S., Heise, H.M. (Eds.), Near-infrared Spectroscopy Principles, Instruments Applications. Wiley-Vch Verlag GmbH., Weinheim, Germany, pp. 125–162.

Kadam, S.S., Wankier, B.N., Adsule, R.N., 1991. Processing. In: Salunkhe, D.K., Kadam, S.S., Jadhav, S.J. (Eds.), Potato: Production, Processing, and Products. CRC press, USA, pp. 111–154.

Kang, S., Lee, K.-J., Son, J.-R., 2008. Online internal quality evaluation system for the processing potatoes. ASABE Publication No. 701P0508cd. Proceedings of the ASABE Food Processing Automation Conference, Providence, Rhode Island, USA.

Kawano, S., 2002. Applications to agricultural products and foodstuffs. In: Siesler, H. W., Ozaki, Y., Kawata, S., Heise, H.M. (Eds.), Near-infrared Spectroscopy Principles, Instruments, Applications. Wiley-Vch Verlag GmbH., Weinheim, Germany, pp. 115–124.

Kawano, S., Fujiwara, T., Iwamoto, M., 1993. Nondestructive determination of sugar content in Satsuma mandarin using near infrared (NIR) transmittance. J. Jpn. Soc. Hort. Sci. 62, 465–470.

Kumar, D.R., Ezekiel, R., 2004. Distribution of dry matter and sugars within a tuber of potato cultivars grown under short day conditions. Potato J. 31 (3–4), 130.

Leardi, Nørgaard, L., 2004. Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions. J. Chemom. 18 (11), 486–497.

Mark, H., 2001. Data analysis: multilinear regression and principal component analysis. In: Burns, D.A., Ciurczak, E.W. (Eds.), Handbook of Near-infrared analysis. Marcel Dekker, Inc., NY, USA, pp. 129–184.

Martens, H., Naes, T., 2001. Multivariate calibration by data compression, In: Phil, P., Norris, K. (Eds.), Near-infrared Technology in the Agricultural and Food Industrious. 2nd Edn. American Association of Cereal Chemists, Inc., MN, USA, pp. 59–100.

Mehrubeoglu, M., Cote, G.L., 1997. Determination of total reducing sugars in potato samples using near-infrared spectroscopy. Cereal Foods World 42 (5), 409–413.

Navarre, D.A., Goyer, A., Shakya, R., 2009. Nutritional value of potatoes: vitamin, phytonutrient, and mineral content. In: Singh, J., Kaur, L. (Eds.), Advances in Potato Chemistry and Technology. Academic Press, Elsevier, UK, pp. 395–424.

Nicolai, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I., Lammertyn, J., 2007. Nondestructive measurement of fruit and vegetable quality by means of nir spectroscopy: a review. Postharvest Biol. Technol. 46, 99–118.

Rady, A.M., Guyer, D.E., Kirk, W., Donis-González, I.R., 2014. The potential use of visible/near infrared spectroscopy and hyperspectral imaging to predict processing-related constituents of potatoes. J. Food Eng. 135, 11–25.

Rama, M.V., Narasimham, P., 2003. Potatoes and related crops. In: Caballero, B. (Ed.), Encyclopedia of Food Science and Nutrition. Academic Press, USA, pp. 4658–4680.

Rastovski, A., Van Es, A., De, D.Z., 1987. Storage of Potatoes Postharvest Behavior, Store Design, Storage Practice, Handling. Center for Agricultural Publishing and Documentation, Wageningen, the Netherlands.

Schallenberger, R.S., Smith, O., Treadaway, R.H., 1959. Role of sugars in the browning reaction in potato chips. J. Agric. Food Chem. 7 (4), 274–277.

Smith, O., Davis, C.O., 1968. Potato processing. In: Smith, O. (Ed.), Potato: Production Storing Processing. The Avi Publishing Company, Inc., Westport, Connecticut, USA, pp. 558–602.

Sowokinos, J.R., 2007. Internal physiological disorders and nutritional and compositional factors that affect market quality, In: Vreugdenhil, D., Bradshaw, J., Gebhardt, C., Govers, F., Mackerron, D.K.L., Taylor, M.A., Ross, H.A. (Eds.), Potato Biology and Biotechnology Advanced Perspectives. 1st Edn. Elsevier Science & Technology Books, UK, pp. 501–523.

Stark, J.C., Love, S.L., 2003. Tuber Quality. In: Stark, J.C., Love, S.L. (Eds.), Potato Production Systems. University of Idaho, Moscow, USA, pp. 329–342.

Storey, R.M.J., 2007. The Harvested Crop. In: Vreugdenhil, D. (Ed.), Potato Biology and Biotechnology Advances and Perspectives. Elsevier, UK, pp. 441–470.

Storey, R.M.J., Davis, H.V., 1992. Tuber quality. In: Harris, P. (Ed.), the Potato Crop. Chapman & Hall, London, UK, pp. 507–569.

Varmuza, K., Filzmoser, P., 2009. Introduction to Multivariate Statistical Analysis in Chemometrics. CRC Press, USA.

Westad, F., Bevilacqua, M., Marini, F., 2013. Regression. In: Marini, F. (Ed.), Chemometrics in Food Chemistry. Elsevier, Oxford, UK, pp. 127–170.

Williams, P., 2007. Near-infrared technology–getting the best out of light (5th Edn.). PDK Grain, Nanaimo, Canada. A short course in the practical implementation of near-infrared spectroscopy for the user.

Wise, B.M., Gallagher, N.B., Bro, R., Shaver, J.M., Windig, W., Kock, R.S., 2006. PLS_Toolbox 4.0 for use with Matlab. Eigenvector Research, Inc., WA, USA.

Wold, S., Sjöström, M., Errikson, M., 2001. PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58, 109–130.

Workman, J., Weyer, L., 2008. Practical Guide to Interpretive Near-infrared Spectroscopy. CRC Press, Boca Raton, FL, USA.

Yaptenco, K.F., Kawakamis, S., Takano, K., 2000. Nondestructive determination of sugar content in 'Danshaku' potato (solanum tuberosum l.) by near infrared spectroscopy. J. Agric. Sci. Tokyo Nogyo Daigaku. 44 (4), 284–294.