Sorting in-shell walnuts using near infrared spectroscopy for improved drying efficiency and product quality

Wang Xiaotuo\textsuperscript{1,3}, Griffiths G. Atungulu\textsuperscript{2*}, Ragab Khir\textsuperscript{1,5}, Gao Zhenjiang\textsuperscript{3}, Pan Zhongli\textsuperscript{1,4}, Shantae A. Wilson\textsuperscript{2}, Gbenga Olatunde\textsuperscript{2}, and David Slaughter\textsuperscript{1}

(1. Department of Biological and Agricultural Engineering, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA; 2. University of Arkansas Division of Agriculture, Department of Food Science, 2650 N Young Avenue, Fayetteville, AR 72704, USA; 3. College of Engineering, China Agricultural University, Beijing, 100083, P.R. China; 4. Processed Foods Research Unit, USDA-ARS-WRRC, 800 Buchanan St., Albany, CA 94710, USA; 5. Department of Agricultural Engineering, Faculty of Agriculture, Suez Canal University, Ismailia, Egypt)

Abstract: In the current walnut drying practice, dryers comingle nuts with varying moisture contents (MC) which results in over-drying of nuts with low MC and thereby decrease product quality. The objectives of this research were to investigate correlations among near infrared (NIR) spectral data and MC of freshly harvested in-shell walnuts and determine the feasibility of using NIR spectroscopic sorting of walnuts to minimize moisture variability in individual nuts within each batch introduced into dryers. NIR transmission spectra of in-shell walnuts of Chandler variety with MCs ranging from 10% to 70% wet basis were determined. Partial least square regression (PLS-R) of the spectral data was performed to establish the correlations among the spectral data and walnut MC. Model validation was also performed to establish the accuracy of MC prediction. The results revealed that there was a strong linear relationship between walnut shell MC and whole nut ($R^2 = 0.968$) and walnut shell MC and kernel MC ($R^2 = 0.851$). The wavelengths at which peak NIR transmission intensity occurred were between 820 and 910 nm. The higher the concentration of moisture in the nut, the lower the NIR transmission value would be. Also, there was a spatial variation in individual nut surface transmission spectra. Model validation indicated that five partial least square (PLS) latent variables were adequate to explain 98% of response variation (MC) and also provided the simplest model with a predicted residual error sum of square (PRESS) statistic of 0.751, which was not significantly different from the absolute minimum PRESS value of 0.744. Predicted and observed MC of walnuts agreed with $R^2$ of 0.978. The results are vital for developing new MC-based sorting methods for walnut using NIR to improve drying efficiency and product quality.

Keywords: walnuts; moisture content; near infrared spectroscopy; sorting


1 Introduction

The average moisture content (MC) of walnuts at harvest could be as high as 30% wet basis (w.b.) while the MC of individual walnut varies significantly due to uneven maturation in the field. Walnuts that fall off the tree before harvest are much dryer because they have attained maturity. Walnuts that are less mature and must be shaken off the tree typically have high MC. It has been observed that during harvesting operation, about 40%-50% of harvested walnut has high MC and fall to the orchard floor with adhering hulls (Romas, 1998). To preserve walnut kernel quality it is essential to dry the nuts to 8% MC (w.b.) (Kader, 2002; Rumsey and Thompson, 1984). However, in the current processing practice, existing dryers comingle all walnuts regardless of the wide range of their MC. Walnuts with higher MC
must be dried longer to bring them to safe storage MC because of the large difference in the MC of the nut (Thompson and Grant, 1992), however exposing the nuts with lower MC may lead to over-drying. Over-drying results in a significant use in energy, prolongs drying time and reduce dried product quality. To overcome over-drying, it is vital to develop a new method that can be used to sort freshly harvested in-shell walnuts based on their MC during drying.

Solid-water interaction is one of the fundamental issues widely studied in food and agricultural product processing (Zografi, 1988). The state of water in a solid material may be characterized using x-ray diffraction, microscopic methods, thermal analysis, vibrational spectroscopy, and nuclear magnetic resonance spectroscopy (Osborne et al., 1993). Near-infrared (NIR) spectroscopy in combination with chemometrics has become an established method for rapid and nondestructive assessment of quality parameters in the food and agricultural sectors. NIR is based on measurements of light absorbed by the sample when it is exposed to electromagnetic radiation in the range from 780 nm to 2500 nm. Qualitatively, the NIR region is between the visible red and the highest frequency used in conventional mid-infrared. NIR spectra of foods comprise broad bands arising from overlapping absorptions corresponding mainly to overtones and combinations of vibrational modes involving C-H, N-H or O-H chemical bonds. Jensen et al. (2001) reported that Vis/NIR spectroscopy and chemometrics can be used to replace the more traditional chemical analyses such as peroxide value, headspace gas chromatography and sensory evaluations of walnut quality. The typical steps in NIR spectroscopy involve performing calibration experiment and establishing multiple linear regressions between the NIR spectra data and the analyte (Buning-Pfaue, 2003; Delwiche et al., 1991; Frake et al., 1997; Goebel and Steffens, 1998; Iwamoto et al., 1987; Luukkonen et al., 2001; Rantanen et al., 1998; Rasanen et al., 2001; White, 1994). Partial Least Squares (PLS) algorithms are generally used to set up a multivariate model to estimate the coefficient for samples of intent (Carlini et al., 2000; Shenk et al., 1992).

To the best of our knowledge, no study has demonstrated the feasibility of sorting walnuts based on MC by using NIR spectroscopy. Such information is vital to develop an automated industrial sorting device for freshly harvested in-shell walnuts just before they are introduced to the dryer. The results of this study could be applied in various ways using existing facilities for instance, sorted nuts with high MCs could be to divert to a high temperature continuous-flow dryer for initial drying and then the drying is completed in another bin dryer at conventional drying temperature. Another option would be to direct the sorted high MC nuts only to a separate bins fitted with auxiliary heaters to increase air temperature above 43°C. Air temperature would be regulated based on the average nut MCs measured by an in-bin MC sensor. These options would reduce drying times, resulting in significant increases in drying capacity, reducing energy use and complementing existing process.

The ultimate objective of this research was to investigate the feasibility of using NIR spectroscopy to sort individual freshly harvested in-shell walnuts based on their MCs, which would improve drying efficiency as well as product quality. The specific objectives of the study were as follows: (1) Establish the relationships among the walnut shell, kernel and whole nut MCs at harvest; (2) Investigate the relationship between walnut shell moisture and NIR spectral data; and (3) Conduct multivariate modeling of NIR spectral data for precise prediction of walnut (MCs).

2 Materials and methods

2.1 Materials

Freshly harvested walnuts of the Chandler variety (not treated with Ethephon) were procured from a local orchard at Cliker Orchards, CA during the 2011 harvest season and were used throughout this study. The samples were collected from the harvester and immediately transported to the laboratory where all the tests were conducted. The freshly harvested walnuts were brushed to remove dirt and visually sorted to eliminate defective ones. At least 200 walnut samples with a wide MC (11.4%-43.5% wet basis.) range were randomly selected
to conduct the NIR and MC measurements. To ensure a wide MC range, walnuts were randomly picked from three harvested walnuts categories which typically occur at harvest including samples without hulls, with hulls partially attached and with hull intact (Figure 1).

![Image of walnut hull conditions](image1.png)

**Figure 1** Different categories of walnut hull conditions at harvest: nuts without, with whole and with partially attached hulls

### 2.2 NIR treatment

All NIR measurements were made with a temperature regulated fiber optic spectrometer (S2000-TR; Ocean Optics, Dunedin, FL). The setup of the NIR spectroscopy experiments is illustrated in Figure 2. Two Teflon disks which were coated black and white were used as reference in the reflectance mode for calibration. Three points falling on the lateral circumference of walnut samples were marked. A walnut sample was placed at predetermined and fixed position within the NIR unit (Figure 2) and measurement of transmission spectra for each of the points was conducted. The procedures lasted less than 30s in order to avoid heating of the products which would result in moisture loss. There was negligible moisture loss in the sample after the NIR spectroscopy. At least 200 samples were used and triplicate measurements were done in the experiment.

To measure the NIR transmission spectra, light beam was directed down towards the sample. Light reflected off the sample was then picked up by a 400μm fiber optic made of schott glass. A filter was inserted at the sensor terminal to block all visible light. The detected signal was fed into a log amplifier, digitized and recorded to a computer for data processing with OOIBase32 Version software. The transmission data was converted to absorbance as log T (T = transmission) which varies linearly with the concentration of the absorber. The mathematical transformation of the data to first or second derivative was done to reduce multiplicative effects on transmission spectra, such as particle size, sample temperature, and sample compaction plus breaking inter-correlations between wavelengths. The second derivative transformation, which was used for all data reported in this paper, was approximated following the method used by Savitzky Golay (Guion José et al., 2007). A twenty one-point scale was used in the derivative calculation.

![Image of NIR setup](image2.png)

**Figure 2** Setup of near-infrared spectroscopy devices for acquiring spectral data of walnuts

### 2.3 Multivariate regression modeling

The multivariate, Partial Least Square Regression (PLS-R) of the second derivative data of the absorbance spectra and MC were performed with SAS statistical software (SAS Institute, Cary, NC) to find the linear equation and principal components of wavelengths which best correlate to the analytical data. Typically, the PLS-R is a projection method that models the relationship between the response $Y$ and the predictors $X$ (Berglund and Wold, 1997; Hoskuldsson, 1996; Martens and Naes, 1989). In the PLS-R procedure, the matrix blocks of the walnut MC and second order derivatives of absorbance were decomposed as follows:

$$X = TP^T + E$$  \hspace{1cm} (1)

$$Y = UQ^T + F$$  \hspace{1cm} (2)

where, $X$ and $Y$ are matrices of predictor and responses respectively. The matrices $T$ and $U$ are the score matrices of $X$ and $Y$ respectively. $P$ and $Q$ are the loading matrices for $X$ and $Y$ respectively while $E$ and $F$ are the residual matrices. The $X$-scores are linear combinations of the $X$-residuals or $X$ itself. The predictive formulation for $Y$ is as follows:

$$Y = TQ^T + F^*$$  \hspace{1cm} (3)

where, $F^*$ is the residual matrix. To perform the validation of the established relationships, an 8-fold
split-sample validation test was performed. Partial least square (PLS) iterations were carried out, and the number of components corresponding to the minimal value of the predicted residual sum of squares (PRESS)/(n−i−1), with i the number of latent variables and PRESS the residual sum of squares of all samples of every cross validation segment, was selected for the final model (Wold et al., 2001). PRESS statistics were established by cross validation using the comparison between real and predicted values for the calibration set. By using this approach the model was applied for the analysis of real walnut samples.

2.4 Moisture content determination

After NIR experiment, the kernel in each nut was extracted using a manual cracker with attention given to minimizing moisture loss or uptake between tests. The MCs of different components, including shell and kernel, for each nut were measured separately. Walnut shells and kernels were separately placed in pre-weighed aluminum dishes and weighed using an electronic balance (Denver Instrument, Arvada Co. USA) with an accuracy of 0.01g and dried in an air oven. Based on our preliminary tests, we found that walnut samples attained a constant dry weight after 24 h at 100°C in the air oven. Therefore, these conditions were adopted throughout the experiments in order to dry walnuts to a constant dry weight. The samples were removed from the oven after 24 h, cooled in desiccators and then weighed again. MC of the samples was determined based on the initial and final (dry) sample weights as follows:

\[ MC_{wb} = \frac{W_i - W_d}{W_i} \times 100 \]  

where, \( MC_{wb} \) is moisture content on wet basis, %, \( W_i \) and \( W_d \) are the initial and dry sample weights, g, respectively. The MC of whole nuts without hulls was determined as follows:

\[ MC_{nh} = \frac{(W_i - W_{nd}) + (W_{si} - W_{kd})}{W_{si} + W_{kd}} \times 100 \]  

where, \( W_{si} \) and \( W_{kd} \) are the initial weights of shell and kernel, g, \( W_{sd} \) and \( W_{kd} \) are the final weights of shell and kernel, g, respectively. All reported moisture contents are on wet weight basis.

3 Results and discussion

3.1 Variability of walnut moisture content at harvest

The obtained results revealed that there was a huge variability in MC among individual walnuts at harvest. The findings were consistent with previously reported studies (Butts et al., 2004; Khir et al., 2011). The MC of walnuts with hull intact at harvest was much higher than that of walnuts with hulls partially attached and without hulls. The walnuts with hulls had average MC of 35.63% ± 6.43% compared to 27.70% ± 2.17% and 7.43% ± 2.79% for walnuts with hull partially intact and without hulls, respectively. When a bulk load of walnut with kernels having high variability in MC in are subjected to the same drying conditions, over-drying of kernels with low MC occurs while kernels with high MC partly dry. In order to reduce loss of quality during post-harvest processing (Altuntas and Erkol, 2011; Sharifian and Derafshi, 2008), there is a need to sort walnuts based on individual MCs before processing. High correlation was found between shell and kernel MC for tested walnut variety as is shown in Figure 3.

Regression models to predict kernel and whole walnut MC are described as follows:

\[ K_M = 0.7663 \times S_M - 0.0355 \]  

\[ W_M = 0.8831 \times S_M - 0.0177 \]  

where, \( K_M \) is the kernel moisture content; \( W_M \) is the whole in-shell walnut moisture content and \( S_M \) is the walnut shell moisture content. The relationships (Equations 6 and 7) can be used to predict MC of walnuts for sorting them into different groups. The sorted walnuts with similar moisture can then be dried separately and will significantly reduce the required drying time for each group and also minimize over- and under-drying.

Figure 3 Relationships among shell, kernel and whole nut moisture contents for walnuts of Chandler variety
3.2 NIR spectra of walnut at harvest

Results demonstrated that there were variabilities in the NIR transmission spectra for the walnuts at harvested MC. Figure 4 is a sample illustration of the mean transmission spectra of walnut with different MCs.

Figure 4 Near-infrared transmission spectra of walnuts with different moisture contents

The higher the concentration of moisture in the nut, the lower the transmission value was. It was observed that for individual walnuts, there was a spatial variation in the nut surface transmission spectra. This could be due to spatial variation in the nut shell characteristics. In practice, it would be advisable to obtain the average of multiple measurements of spectral information of the nut surface for precise prediction of the nut MC by NIR spectroscopy.

3.3 Partial least square regression and model validation

Based on study results, the variation between nut MC and nut surface transmission intensity was not linear within the studied MC range of 10% to 70% w.b. (Figure 5a). At the same time, the wavelength at which peak nut surface transmission occurred was not constant (Figure 5b). The peak transmission ranged between 820 and 910 nm. The latter phenomenon is likely due to lack of uniformity in sample matrix or variance in background composition among other factors which could contribute to substantial changes in spectra such as band broadening and peak position shifts.

Figure 5 Peak transmission intensity and wavelength at which peak transmission occurred versus walnut moisture content

Partial least square regression (PLS-R) was especially appropriate for finding a few underlying predictors that account for most of the variation in the response (MC). To accomplish the PLS-R, mathematical transformation of transmission data to second derivative was done to reduce multiplicative effects on transmission spectra, such as particle size, sample temperature, and sample compaction plus breaking inter-correlations between wavelengths among other factors. Table 1 shows how much predictor and response variations are explained by each partial least square (PLS) components. The individual and cumulative variation accounted for at least the first fifteen PLS factors. The model effects and dependent variables are also shown in Table 1. The first five PLS factors accounted for almost all of the variation in the response variable, with the fifth factor accounting for a sizable proportion. Based on the results, it was noted that with the number of latent variables, 98% of the response variation was already explained, while only 8% of the predictor variation was explained at the same level.

In the split sample validation test, data was divided into eight groups and fitted to all but one, and then the capability of the model to predict response for the omitted group was checked. PRESS statistic indicated in Table 1 is based on the residuals generated by this process. Based on the results of the cross validation procedure, a PLS model with five components (latent variables) was chosen because it was the simplest model with a PRESS statistic (0.751) that is insignificantly different from the absolute minimum PRESS value. Although the model with ten
PLS factors achieved the absolute minimum PRESS (0.744), it is not significantly different from the model with only five factors. The X- and Y-scores of a model with five PLS factors is indicated in Figure 6. From the plots, the X- and Y-scores for the five factors model was highly correlated, indicating a good model. The number on the plots refers to the data set numbers. Up to 200 data points are indicated and represent a mean of three replicates from 600 independent observations.

Table 1  Percentage of variation explained by model and cross validation results

<table>
<thead>
<tr>
<th>Number of extracted factors</th>
<th>Model effects</th>
<th>Dependent variables</th>
<th>Split-sample Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td>Total</td>
<td>Current</td>
<td>Total</td>
</tr>
<tr>
<td>0</td>
<td>1.915</td>
<td>1.915</td>
<td>70.693</td>
</tr>
<tr>
<td>1</td>
<td>1.832</td>
<td>3.747</td>
<td>17.346</td>
</tr>
<tr>
<td>2</td>
<td>1.626</td>
<td>5.372</td>
<td>5.353</td>
</tr>
<tr>
<td>3</td>
<td>1.583</td>
<td>6.955</td>
<td>2.749</td>
</tr>
<tr>
<td>4</td>
<td>1.493</td>
<td>8.449</td>
<td>1.702</td>
</tr>
<tr>
<td>5</td>
<td>1.436</td>
<td>9.884</td>
<td>0.821</td>
</tr>
<tr>
<td>6</td>
<td>1.342</td>
<td>11.226</td>
<td>0.491</td>
</tr>
<tr>
<td>7</td>
<td>1.321</td>
<td>12.547</td>
<td>0.339</td>
</tr>
<tr>
<td>8</td>
<td>1.353</td>
<td>13.906</td>
<td>0.208</td>
</tr>
<tr>
<td>9</td>
<td>1.270</td>
<td>15.170</td>
<td>0.124</td>
</tr>
<tr>
<td>10</td>
<td>1.538</td>
<td>16.707</td>
<td>0.059</td>
</tr>
<tr>
<td>11</td>
<td>1.199</td>
<td>17.906</td>
<td>0.045</td>
</tr>
<tr>
<td>12</td>
<td>1.106</td>
<td>19.012</td>
<td>0.028</td>
</tr>
<tr>
<td>13</td>
<td>1.096</td>
<td>20.108</td>
<td>0.016</td>
</tr>
<tr>
<td>14</td>
<td>1.136</td>
<td>21.244</td>
<td>0.010</td>
</tr>
</tbody>
</table>

Minimum root mean PRESS 0.744
Minimizing number of factors 10
Smallest number of factors with p > 0.1 5

Figure 6  The fifth X- and Y-scores for PLS models

To determine the accuracy of the NIR calibration, a linear fitting of predicted and observed walnut MC was performed. Figure 7 shows measured and predicted values of walnut MC as well as the residuals associated with the prediction. It was observed that NIR spectroscopy could be used to predict the MC of walnuts shell \(R^2=0.978\). It is easy to determine the MC of either the kernel or the whole nut because of the strong correlation to the walnut shell MC. It is important to emphasize that nonlinearities are not necessarily obvious from the residuals of the PLS-model (Figure 7b) (Isaksson and Naes, 1988). Based on NIR spectroscopy information and PLS-R of the data, it is feasible to sort walnuts to achieve desired moisture difference among samples within each batch introduced into the dryers. This consequently will reduce drying time and lead to over-drying, but improve energy efficiency.

4 Conclusions

The study investigated correlations among near-infrared (NIR) spectral data and MC of freshly harvested in-shell walnuts and determine feasibility of using NIR spectroscopic sorting of walnuts to minimize moisture variability in individual nuts within each batch introduced into dryers and concluded that:
(1) Moisture content among freshly harvest walnut varied significantly ($p<0.05$);

(2) The higher the concentration of moisture in the nut, the lower the transmission value ($p<0.05$) was;

(3) The peak NIR transmission intensity occurred at wavelengths between 820 and 910 nm for in-shell walnuts of MC of 10% to 70%;

(4) The five partial least square regression latent variables were adequate to explain 98% of response variation (MC);

(5) The study established the feasibility of using NIR spectroscopy to sort walnuts of different shell moisture contents (MC).

Acknowledgements

The authors wish to express appreciation to Ruipeng Fu, Hui Ean Teh, Tianxin Wang, Brian De La Cruz and Guang Shi of the University of California Davis for their strong support in sample collection and preparation. The authors also appreciate the advice from Dr. Leroy Garcia, and Dr. Gopal Tiwari regarding NIR spectroscopy research. In addition, the authors wish to communicate lots of gratitude to industry partners, especially Cilker Orchards for supplying walnuts samples and the Energy Innovations Small Grants (EISG) program for providing funding to conduct this research. The research was conducted at the Western Regional Research Center of USDA-ARS and Department of Biological and Agricultural Engineering, University of California, Davis, USA.

[References]


