Original Article

Determination of water activity, total soluble solids and moisture, sucrose, glucose and fructose contents in osmotically dehydrated papaya using near-infrared spectroscopy

Bumrungrat Rongtong, Thongchai Suwonsichon, Pitiporn Ritthiruangdej, Sumaporn Kasemsumran

A Department of Product Development, Faculty of Agro-Industry, Kasetsart University, Bangkok, 10900, Thailand
B Nondestructive Quality Evaluation Technology Lab, Kasetsart Agricultural and Agro-Industrial Product Improvement Institute, Bangkok, 10900, Thailand

Abstract

Near-infrared spectroscopy (NIRS) is a rapid analysis method that is widely used for quantitative determination of the major constituents in many food products. NIRS was applied in conjunction with a chemometric algorithm, namely the partial least squares regression (PLSR), to develop the optimum model for predicting the qualities of osmotically dehydrated papaya (ODP). Two hundred ODP samples were collected from commercial products and from different laboratory ODP processes with varying sucrose concentrations (35°Brix, 45°Brix, 55°Brix and 65°Brix) at 40 °C for 6 h and drying times at 60 °C for 2 h, 4 h, 6 h, 8 h, 10 h and 12 h. All samples were divided into a calibration set (n = 140) and a validation set (n = 60) before quality determination and NIRS analysis. Samples were scanned over the NIR spectral range of 800–2400 nm in reflectance mode and their spectra were pretreated using the second derivative method. Suitable predictive models were developed by applying full wavelength PLSR and two wavelength interval selection methods, named the moving window partial least squares regression (MWPLSR) and the searching combination moving window partial least squares regression (SCMWPLSR). The results showed that SCMWPLSR provided better performance than PLSR and MWPLSR. The root mean square error of prediction values of water activity, moisture content, total soluble solids and the sucrose, glucose and fructose contents from SCMWPLSR were 0.014, 0.69% (dry basis), 0.58, 0.014, 0.69% (dry basis) and 0.014, 0.69% (dry basis), respectively, with correlation coefficients in the range 0.981–0.994.

Copyright © 2018, Production and hosting by Elsevier B.V. on behalf of Kasetsart University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0).
conventional analytical methods such as water activity analysis, drying in an oven under pressure, refractometry and high performance liquid chromatography (HPLC) are inherently destructive, time-consuming, require laborious preparation of the sample and generate chemical waste (AOAC, 2000; Ramallo et al., 2004; Viegas et al., 2016).

Recently, near-infrared spectroscopy (NIRS) has been seen as a powerful, convenient and rapid analytical method which can be used to measure quality parameters of osmotically dehydrated fruit products (Cen and He, 2007). This technique has been reported to quantify residual moisture on osmo-air dehydrated apple rings (Barzaghi et al., 2008), to monitor osmotic exchanges by determining mass balances (water loss, solid gain and sugar intake) of a blueberry osmo-air dehydration process (Sinelli et al., 2011) and to determine the total sugars to sucrose ratio of osmotic dehydrated mango and pineapple prior to drying (Mahayothee et al., 2017). In the construction of calibration models for quantitative determination using NIRS, the most commonly used multivariate analysis method is partial least squares regression (PLSR). This paper considered NIRS combined with both chemometric algorithms, named the moving window partial least squares regression (MWPLSR) and the searching combination moving window partial least squares regression (SCMWPLSR) to improve the accuracy of the predictive PLSR model. MWPLSR works as a spectral selection method to locate informative regions through spectra while SCMWPLSR is applied to these informative regions for optimizing and seeking the best optimized informative regions as either an individual region or combination of the informative regions (Kasemsumran et al., 2006). Within this context, the aim of this study was to apply NIRS combined with chemometric algorithms to identify the optimum predictive model for quantifying osmotically dehydrated papaya (ODP) quality.

**Materials and methods**

**Samples**

In total, 200 ODP samples were used consisting of commercial ODP products ($n = 60$) and ODP laboratory samples ($n = 140$). The commercial ODP products were purchased from food fairs, souvenir shops, convenience stores and supermarkets around Kasetsart University, Bangkok, Thailand. The samples were stored at room temperature ($25 \pm 1 ^\circ C$) until they were analyzed.

**Papaya sample preparation and osmotic dehydration treatment**

Papaya fruits (Carica papaya L. cv. Khaek dam) harvested 92 d after anthesis were bought from a local market, Talaadthai, Thailand. The ODPs were prepared according to the method described by Thimthong et al. (2012), with slight modifications. Each papaya was peeled, cut into a cube (1 cm $\times$ 1 cm $\times$ 1 cm) and kept in 0.3% Na$_2$S$_2$O$_5$ (Sigma-Aldrich; Germany) before further use. The prepared papaya cubes were washed before soaking in a solution containing 0.1% Na$_2$S$_2$O$_5$ (Merck; Germany) and 2% CaCl$_2$ (Merck; Germany) with the ratio of fruit and solution being 1:3 for 30 min, then removed from the solution and washed with water. The samples were blanched in boiled water for 5 min at 100 °C. Commercial food grade sucrose was used to prepare the osmotic solutions. The five concentration levels of sucrose solution were 35°Brix, 45°Brix, 55°Brix, 65°Brix and 75°Brix, respectively. The cubes were osmotically dehydrated using a ratio of fruit to syrup at 1:2 (weight per weight). The temperature in the osmotic dehydrated process was maintained at 40 °C in a controlled temperature shaking water bath (BS-11; Lab Companion; Korea). The osmotic dehydration process was performed for 6 h. The osmotically dehydrated fruits were drained for 1–2 min and then gently blotted with absorbent paper to remove any excess and adhering solution on the fruit surface. These samples were dried in a hot-air oven (UF110; Memmert; Germany) at 60 °C and sampled at 2 h, 4 h, 6 h, 8 h, 10 h and 12 h for quality determination. The ODP laboratory samples were stored in aluminum bags and kept at room temperature ($25 \pm 1 ^\circ C$) until they were analyzed.

**Near-infrared spectral collection**

NIR spectrometry (SpectraStar™ 2500; Unity Scientific; USA) was used to acquire reflectance spectra (log 1/R) of the ODP samples. Samples were filled and scanned in a glass Petri dish with a diameter of 9 cm using a sample thickness greater than 1 cm (Fig. 1). The spectral acquisition and instrument control were controlled using the Info star software version 3.10.0 (Unity Scientific; USA). NIR spectra were recorded in a wavelength range of 800–2400 nm with an average resolution every 2 nm. The 1202–1328 nm region represented a baseline shifting due to the different responses of two detector types, so this region was not used to avoid heavily overlapping absorption bands (Rongtong and Suwonsichon, 2014). Three replicates were acquired per sample and the spectra were averaged before further calculations.

**Quality determination of osmotically dehydrated papaya samples**

**Water activity**

ODP samples cut into smaller pieces were measured at 25 °C using a water activity instrument (Aqualab CX3-TE; Labo-Scientifica; Italy). After equilibration, the water activity value was recorded.

**Moisture content**

ODP samples (3 g) were weighed into aluminum cans for moisture content determination (AOAC, 2000) in a vacuum oven (VD 53; Binder; Germany) at 70 °C under 100 mBar pressure until constant weight was obtained (Chenlo et al., 2007).

**Total soluble solids**

ODP samples (approximately 50 g) were blended with 200 mL of deionized water using a blender. The total soluble solids were...
determined using a digital refractometer (PAL-a; ATAGO; Japan) and the °Brix value calculated using a dilution factor (adapted from AOAC, 2000).

Sucrose, fructose and glucose contents

The sucrose, fructose and glucose contents in ODP were determined using an HPLC method of Liu et al. (2013) with some modifications. A sample of 50 g of ODP samples was mixed and homogenized with 200 mL deionized water and heated in a shaking water bath (T22LAS; Thomas Kagaku Co., Ltd.; Japan) using a shaking frequency of 60 revolutions per minute (rpm) for 30 min at 50 °C, then centrifuged at 9000 rpm/min for 15 min at 4 °C in a refrigerated tabletop centrifuge (Rotina 35R; Andreas Hettich GmbH & Co.; Germany), the supernatant was collected and used for the HPLC determination of sucrose, fructose and glucose in samples. Samples of a 1/10 dilution of supernatant with deionized water were filtered through VertiClean™ CA syringe filters (0.45 µm; Vertical Chromatography Co. Ltd.; Thailand). The HPLC system (Shimadzu Co.; Japan) was composed of a micro-plunger pump (LC-10ADVP; Shimadzu Co.; Japan) connected to a refractive index detector (RID-10A; Shimadzu; Japan). The column was equipped with a VertiSep™ SUGAR CMP HPLC column (7.8 x 300 mm internal diameter, 9 µm particle size; Vertical Chromatography Co. Ltd.; Thailand) as a guard column. The column was held at 80 °C. The mobile phase was deionized water and the flow rate was 0.4 mL/min for 35 min. A 20 µL sample loop was used for injection. Quantification was carried out using an external standard with HPLC-grade sucrose, glucose and fructose (Sigma-Aldrich; Germany). The results were expressed as grams per 100 g of sample.

Data analysis

The 200 samples were divided into a calibration set (n = 140) and a validation set (n = 60). The second derivative spectral pretreatment method was applied to process all spectra before developing partial least squares regression (PLSR) models using the Unscrambler software version 8.0 (CAMO AS; Norway). Additionally, two wavelength interval selection methods named the moving window partial least squares regression (MWPLSR) and the searching combination moving window partial least squares regression (SCMWPLSR) were also applied to determine the suitable input wavelength variables using an in-house-written program with the MATLAB software (version 6.5; The MathWorks; USA). The prediction performance was evaluated using the correlation coefficient (R) and root mean square error of prediction (RMSEP). The ratio of the standard error of prediction to deviation (RPD) value is an index used to check robustness of a model, and relatively high RPD values indicate models having greater power to predict the chemical composition (Cozzolino et al., 2005). Ideally, the optimum model should have higher R and RPD values, and lower RMSEP values.

Results and discussion

Quality determination of ODP samples

The statistical characteristics of the calibration and prediction sets are summarized in Table 1. A wide variability in composition was observed as a result of the quality parameters: water activity values ranged between 0.472 and 0.870, moisture contents between 7.72% and 28.69% (dry basis), total soluble solids from 58.0 °Brix to 73.6 °Brix, sucrose contents from 38.22 g/100 g of sample to 406.56 g/100 g of sample, glucose contents from 36.87 g/100 g of sample to 356.58 g/100 g of sample and fructose contents between 31.48 g/100 g of sample and 308.85 g/100 g of sample.

Near-infra red spectra

Fig. 2A shows the original NIR reflectance spectra of the 200 ODP samples in the spectral region 800–2400 nm with the major
Fig. 3. Residue lines obtained by moving window partial least squares regression for quality parameters of osmotically dehydrated papaya: (A) water activity; (B) moisture content; (C) total soluble solid; (D) sucrose; (E) glucose; (F) fructose, where gray-shaded blocks indicate final informative regions.
Fig. 3. (Continued).
absorption bands of water at 970 nm, 1450 nm and 1940 nm. The positions of these bands can be shifted by temperature changes or hydrogen bonding interactions with sample components. At 970 nm, this could be attributed to the second overtone bands of O–H groups present in the water. The band at 1450 nm corresponds to the first O–H stretching overtone; the band at 1940 nm is the O–H stretching and deformation combination band (Osborne et al., 1993; Maeda et al., 1995; Blanco et al., 2000). As shown in Fig. 2A, the raw spectral data needed to undergo spectral preprocessing. There were undesirable systematic variations in the data, such as peak overlap, baseline drift and random noise. In this study, the second derivative spectral preprocessing method was applied. The resulting second derivative spectra are shown in Fig. 2B.

In informative regions obtained using moving window partial least squares regression

The residual lines obtained using MWPLSR for six components (water activity, moisture content, total soluble solids and sucrose, glucose, and fructose contents) are plotted in Fig. 3A–F. In Fig. 3A and B, the three informative regions selected using MWPLSR for water activity and moisture content corresponded to the region giving small values of log of the sums of squared residuals (log (SSR)). The NIR region around 970 nm related to O–H stretch second overtones, at 1450 nm to O–H stretch first overtones and at 1940 nm to O–H stretch and O–H deformation (Osborne et al., 1993).

Fig. 3C shows three informative spectra regions of 866–1174, 1434–1938 and 2028–2386 nm, which had small values of SSR. These informative regions contained useful information for PLSR model building of the total soluble solids content due to the C–H stretch third overtones, O–H stretch second overtones, combinations of –CH2 stretching and C–H deformation (Williams and Norris, 2001).

The residue lines for the other three components (sucrose, glucose and fructose contents) obtained using MWPLSR are shown in Fig. 3D–F. The three spectral regions of 910–1162, 1350–1792 and 1904–2210 nm are considered as informative regions for sucrose (Fig. 3D). These regions contained the regions for the C–H stretch third overtones, O–H stretch first overtones and the combination of O–H stretching and deformation modes of sucrose (Osborne et al., 1993; Williams and Norris, 2001). It can be seen in Fig. 3E that glucose has two informative regions of 840–1062 and 1332–1790 nm where bands due to the first overtones of the O–H and C–H stretching modes of glucose exist (Workman and Weyer, 2007). Two informative regions of 816–1050 and 1332–1772 were obtained for fructose (Fig. 3F). These regions corresponded to the regions for bands due to the combinations of C–H stretching and O–H deformation modes (Osborne et al., 1993). These results were in agreement with Sinelli et al. (2011) who reported the performance of NIRS as a tool for monitoring blueberry osmo and air dehydration processes infused in sucrose and glucose/fructose solutions.

Comparison of partial least squares regression calibration models

The calibration models for each component were developed for the whole range, informative regions selected using MWPLSR, the direct combination of the informative regions obtained using MWPLSR and the combination regions selected using SCMWPLSR. It can be seen from Table 2 that the PLSR models based on the informative regions selected using MWPLSR had better performance than the one based on the whole spectral region. With the optimized combination of informative regions obtained using SCMWPLSR, the developed model substantially improved the prediction ability and produced the best prediction results. SCMWPLSR is a powerful method for finding a suitable input wavelength variable, which can be used to improve the performance of a PLSR model with low RMSEP, a small number of PLSR factors and high correlation coefficients. The lowest RMSEP values for water activity, moisture content, total soluble solids and the sucrose, glucose and fructose contents were 0.014, 0.69% (dry basis), 0.58 Brix, 14.44 g/100 g of sample, 6.72 g/100 g of sample and 4.89 g/100 g of sample, respectively. The wavelength interval selection methods had the advantages of being able to handle collinear X-variables in the NIR data which contained many redundant variables and were highly correlated as well as the small number of data obtained using wavelength interval selection methods being able to be used to avoid causing over-fitting due to selecting too many wavelengths (Du et al., 2004). The results

---

**Table 2** Prediction results of partial least squares regression calibration models for water activity, moisture content, total soluble solids and sucrose, glucose and fructose contents of osmotically dehydrated papaya samples.

<table>
<thead>
<tr>
<th>Analyte</th>
<th>Method</th>
<th>Wavelength region (nm)</th>
<th>F</th>
<th>Calibration set</th>
<th>Validation set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Rcal</td>
<td>RMSEC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RMSEP</td>
<td>RPD</td>
</tr>
<tr>
<td>Water activity</td>
<td>PLSR</td>
<td>800–1200, 1330–2400</td>
<td>8</td>
<td>0.978</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>MWPLSR</td>
<td>890–1190, 1400–1700, 1900–2150</td>
<td>7</td>
<td>0.995</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>SCMWPLSR</td>
<td>916–1100, 1412–1620, 1910–2138</td>
<td>7</td>
<td>0.999</td>
<td>0.008</td>
</tr>
<tr>
<td>Moisture content (%)dry basis</td>
<td>PLSR</td>
<td>800–1200, 1330–2400</td>
<td>5</td>
<td>0.982</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>MWPLSR</td>
<td>884–1160, 1338–1638, 1732–2070</td>
<td>5</td>
<td>0.996</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>SCMWPLSR</td>
<td>916–1132, 1382–1636, 1880–2056</td>
<td>5</td>
<td>0.999</td>
<td>0.68</td>
</tr>
<tr>
<td>Total soluble solid (%)Brix</td>
<td>PLSR</td>
<td>800–1200, 1330–2400</td>
<td>8</td>
<td>0.979</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>MWPLSR</td>
<td>866–1174, 1434–1938, 2028–2386</td>
<td>6</td>
<td>0.991</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>SCMWPLSR</td>
<td>910–1174, 1446–1918, 2048–2348</td>
<td>6</td>
<td>0.996</td>
<td>0.51</td>
</tr>
<tr>
<td>Sucrose (g/100 g of sample)</td>
<td>PLSR</td>
<td>800–1200, 1330–2400</td>
<td>8</td>
<td>0.972</td>
<td>16.90</td>
</tr>
<tr>
<td></td>
<td>MWPLSR</td>
<td>910–1162, 1350–1792, 1904–2210</td>
<td>6</td>
<td>0.987</td>
<td>15.59</td>
</tr>
<tr>
<td></td>
<td>SCMWPLSR</td>
<td>904–1122, 1428–1790, 1972–2104</td>
<td>6</td>
<td>0.990</td>
<td>11.74</td>
</tr>
<tr>
<td>Glucose (g/100 g of sample)</td>
<td>PLSR</td>
<td>800–1200, 1330–2400</td>
<td>8</td>
<td>0.922</td>
<td>19.49</td>
</tr>
<tr>
<td></td>
<td>MWPLSR</td>
<td>840–1062, 1312–1790</td>
<td>6</td>
<td>0.982</td>
<td>7.18</td>
</tr>
<tr>
<td></td>
<td>SCMWPLSR</td>
<td>856–1060, 1340–1788</td>
<td>6</td>
<td>0.989</td>
<td>6.54</td>
</tr>
<tr>
<td>Fructose (g/100 g of sample)</td>
<td>PLSR</td>
<td>800–1200, 1330–2400</td>
<td>8</td>
<td>0.949</td>
<td>11.02</td>
</tr>
<tr>
<td></td>
<td>MWPLSR</td>
<td>816–1050, 1332–1772</td>
<td>6</td>
<td>0.983</td>
<td>5.04</td>
</tr>
<tr>
<td></td>
<td>SCMWPLSR</td>
<td>880–1058, 1410–1768</td>
<td>6</td>
<td>0.986</td>
<td>4.62</td>
</tr>
</tbody>
</table>

F = factors; Rcal = correlation coefficient of calibration; RMSEC = root mean square error of calibration; RPD = correlation coefficient of validation; RMSEP = root mean square error of prediction; RPD = ratio of standard error of prediction to deviation; PLSR = partial least squares regression; MWPLSR = moving window partial least squares regression; SCMWPLSR = searching combination moving window partial least squares.
suggested that SCMWPLSR was better than PLSR for NIRS and this conclusion was also supported by other studies. For example, Ritthiruangdej et al. (2005) compared the prediction capabilities of different spectral analysis models, including the PLSR, MWPLSR and SCMWPLSR methods using a visible-NIR spectrometric technique to quantitatively analyze the total nitrogen content, pH, refractive index, density and °Brix in Thai fish sauces. The result demonstrated that NIR spectroscopy combined with SCMWPLSR was powerful for both the quantitative and qualitative analyses of Thai fish sauces. Kang et al. (2006) developed PLSR calibration models for cholesterol, glucose and urea in control serum solutions using NIRS and SCMWPLSR and demonstrated the use of MWPLSR and SCMWPLSR in low concentration ranges for the three blood components and concluded that SCMWPLSR could improve the

Fig. 4. Scatter plots between actual and predicted quality parameters of ODP samples: (A) water activity; (B) moisture content; (C) total soluble solid; (D) sucrose; (E) glucose; (F) fructose, where the partial least squares regression model was developed using the optimized informative regions obtained using searching combination moving window partial least squares regression.
performance of PLSR models by optimizing the informative spectral regions. The performance of the PLSR model obtained using SCMWPLSR was evaluated based on conformity of the observed and predicted quality parameters. The relationship between the observed and predicted water activity, moisture content, total soluble solids and the sucrose, glucose and fructose contents for the test datasets in SCMWPLSR are presented in Fig. 4A–F, respectively. The R values (0.981–0.994) in the test sets models were larger than 0.9, and therefore the models can be considered to be reasonably acceptable. The results from this comparison suggested that these models will be able to reasonably predict the quality of new ODP samples using their NIR spectra.

NIRS combined with the chemometric algorithms such as in the MWPLSR and SCMWPLSR methods could be used to develop predictive models for determining the water activity, moisture content, total soluble solids and the sucrose, glucose and fructose contents of ODP. The SCMWPLSR models performed well with their statistical results providing the lowest RMSEP values and high correlation coefficients. Therefore, NIRS could be considered as a nondestructive tool for ODP quality control. Furthermore, it could be designed as an on-line measurement tool during ODP production.

Conflict of interest

The authors declare that there are no conflicts of interest.

Acknowledgements

The authors acknowledge funding received from the Kasetsart University Research and Development Institute (KURDI), Bangkok, Thailand and the Non-Destructive Quality Evaluation Special Research Center and Cassava and Starch Technology Research Unit for laboratory facilities in the Kasetsart Agricultural and Agro-Industrial Product Improvement Institute (KAPI), Bangkok, Thailand.

References


