Estimating the starch-iodine blue value of residual liquid of rice by near-infrared reflectance spectroscopy

Bo-Jein Kuo¹)  Mei-Chu Hong²)

(Accepted for publication: August 22, 2001)

Abstract: Starch-iodine blue value of residual liquid (BV) could be used to predict the palatability of rice. The objectives of this study were to evaluate the potential of Near-infrared reflectance spectroscopy (NIRS) to estimate BV of rice and compare the performance of model-building and prediction in developing the calibration equation using multiple linear regression (MLR), principal component regression (PCR), and partial least squares regression (PLSR), respectively. A total of 238 rice samples, including indica, japonica, and waxy rices, were employed in analyzing the BV. The 4-wavelength MLR model gave the highest correlation coefficient and the lowest standard error of prediction than PCR and PLSR models. In other words, This MLR model had the best predictive ability in estimating the BV of rice. The predictive ability of the PCR model with 17 principal components was slightly better than that of the PLSR model with 6 components. However, The IRV value in MLR model (82.042) was much larger than that in PCR (18.331) model and PLSR (8.356) model, suggesting that more random noise was involved in MLR model than that in other models. Generally speaking, the performances of model-building and prediction demonstrate that calibration for BV is less accurate and NIRS can be used only for the first approximation of BV in the rice breeding program.

Key Words: starch-iodine blue value of residual liquid (BV), NIRS, multicollinearity, MLR, PCR, PLSR

INTRODUCTION

Rice (Oryza sativa L.) is the staple food in Taiwan. However, the consumption of rice in Taiwan is declined sharply because of the impact of western culture and industrialization. To increase the rice consumption, producing the rice with high palatability quality has become one of the most important strategies. Rice breeders in Taiwan have widely employed taste panel test to evaluate the palatability of cooked rice. Furthermore, only small quantities of sample could be available for evaluation in early generation. Hence, the exploration of the relationship

¹) Agronomy Department, National Chung Hsing University, Taichung 402, Taiwan, ROC. (Corresponding author)
²) Taichung District Agricultural Improvement Station, Changhwa, Taiwan, ROC.
between some physiochemical properties and palatability might be feasible to screen the rice lines with high palatability (Chang and Li, 1985; Kuo et al., 1987; Kuo et al., 2001a; Lii et al., 1986; Liu et al., 1988). Chikubu et al. (1985) found that protein content, maximum viscosity, minimum viscosity and breakdown viscosity in amylographic characteristics of milled rice flour, and starch-iodine blue value of residual liquid (BV) were the most important characteristics for predicting the palatability of rice. Conventional chemical analysis of BV is time-consuming and inconvenient. Therefore, establishing a fast and convenient screening technique is needed for breeding the rice with high palatability.

Near-infrared reflectance spectroscopy (NIRS) is a fast and convenient technique widely used on the analysis of several rice quality traits (Chen, 1996; Delwiche et al., 1995, 1996; Kuo et al., 2001b; Li and Shaw, 1997; Villareal et al., 1994; Windham et al., 1997; Wu et al., 1997, 2000). However, the feasibility of NIRS technique in estimating the BV is not well documented. Kuo et al. (2001b) demonstrated that NIRS was unsuccessful in estimating the BV of ground milled rice. They also found that MLR model had the best predictive ability in estimating the BV of ground milled rice. MLR showed large random noise due to the serious multicollinearity problems among the selected wavelengths. In addition, the inflated variability of the estimated regression coefficients was found in MLR model and the change of model performance was significant after deleting the outliers. PCR performed more stable than MLR, while PLSR was the most robust calibration method.

The objectives of this study were to evaluate the potential of NIRS to estimate starch-iodine blue value of residual liquid and compare the performance of model-building and prediction in developing the calibration equation using MLR, PCA, and PLSR, respectively.

**Materials and Methods**

A total of 238 rice samples were employed in this study. These samples, including cultivars or lines, consisted of indica, japonica, and waxy rices. The samples were threshed and dried overnight at 45°C to approximately 14.0±0.5% moisture content. The rough rice was hulled with a Satake testing husker and milled with a McGill #3 mill for 60 sec. Then, the milled rice was ground with a cyclone grinder (Udy, Fort Collins, CO., USA) equipped with a 0.25 mm screen.

Eight grams of milled rice were placed in a small cylinder made of 20-mesh metal wire gauze. The metal wire cylinder and contents were slowly put into a 300 ml beaker containing 160 ml distilled water. The beaker was placed into a boiling water bath for 20 min. After cooking, the cooked rice was drained for 10 min and the residual liquid was saved for later use. One ml of the cold residual liquid was made up to 100 ml with distilled water. Two ml iodine solution were added and shaken well. The absorbance of the BV was measured with a colorimeter at the wavelength of 600 nm.

A Bran + Luebbe InfraAlyzer 500 (Norderstedt, Germany) was used to collect spectrum measurement for each ground sample. Each rice flour sample in a closed standard cup was scanned from 1,100 to 2,500 at 4-nm intervals. The reflectance (R) spectrum readings were recorded and transformed to log (1/R). MLR, PCA, and PLSR were performed on the spectra and corresponding BV to develop the calibration equation (Bran + Luebbe). The stepwise search procedure was used to find the best one-wavelength
MLR model, the best two-wavelength MLR model, and so on. This procedure continued until the best 9-wavelength MLR model was found. The number of cross-validation segments (CVS) was 5 for PCR and PLSR. The standard error of prediction was calculated over all cross-validation segments for the first 20 components (SEP<sub>cv</sub>). The optimal component number of PCR and PLSR was determined by SEP<sub>cv</sub> value (Bran+Luebbe). The samples were divided into calibration and validation sets by actual BV. The actual BV of the whole samples was sorted by ascending order. In other words, the samples were arranged from lowest to highest according to the actual BV. Starting with the five samples with the lowest actual chemical values, the odd samples were assigned to the calibration set, and the even samples were assigned to the validation set. The other samples were assigned in the same way until 3 samples with the highest actual chemical value were remained. Two odd samples were assigned to the calibration set, and one even sample was assigned to the validation set. There were 145 samples in calibration set and 93 samples in validation set, respectively. The performance of model-building and prediction was evaluated using the standard error of calibration (SEC), correlation coefficient calculated from calibration set (r<sub>cal</sub>), correlation coefficient calculated from validation set (r<sub>val</sub>), the standard error of prediction (SEP), and the average difference between chemical and NIR predicted values (bias). The index of random variation (IRV) was calculated as the square root of the sums of the squares of the regression coefficients except the intercept.

**Results and Discussion**

The descriptive statistics of the calibration and validation sets are listed in Table 1. The BV in calibration set ranged from 0.009 to 0.220 with an average of 0.109. The BV in validation set ranged from 0.011 to 0.207 with an average of 0.108. All values in validation set are within the calibration set range.

Table 1. Summary of descriptive statistics for starch-iodine blue value of residual liquid of ground milled rice sample in calibration and validation sets.

<table>
<thead>
<tr>
<th>Set</th>
<th>n</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>SD</th>
<th>SD&lt;sub&gt;tot&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calibration</td>
<td>145</td>
<td>0.109</td>
<td>0.009</td>
<td>0.220</td>
<td>0.054</td>
<td>0.053</td>
</tr>
<tr>
<td>Validation</td>
<td>93</td>
<td>0.108</td>
<td>0.011</td>
<td>0.207</td>
<td>0.050</td>
<td></td>
</tr>
</tbody>
</table>

SD<sub>tot</sub> = Standard deviation of samples of calibration and validation

Fig. 1A shows the effect of standard error and correlation coefficient when MLR was used to develop the calibration equation. As the number of wavelengths was increased to 4, the value of SEP decreased to 0.025 and the value of r<sub>cal</sub> increased dramatically to 0.869. The 4-wavelength model was recommended by means of SEP and r<sub>cal</sub> values. As shown in Fig. 1B, the model performance of PCR, as evaluated by an increase in r<sub>cal</sub> value and a decrease in SEP value, increased readily up to 10 components. The
Estimating the starch-iodine blue value of residual liquid of rice by near-infrared reflectance spectroscopy

Fig 1. The plot of model-building and predictive ability performance for the blue value of residual liquid by different statistical methods.
17-component model was chosen as optimal in terms of the SEP\(_{cv}\) value. This recommended PCR model had the \(r_{val}\) and SEP of 0.841 and 0.028, respectively. Fig. 1C presents the performance of standard error and correlation coefficient using PLSR in developing the calibration equation. As the number of components was increased to 6, the value of SEP declined to 0.028 and the value of \(r_{val}\) increased to 0.828. Six components were included to develop the optimal calibration equation by PLSR method.

Table 2 summarizes the model performance of BV by different statistical methods. The IRV value in MLR model (82.042) was much larger than that in PCR (18.331) model and PLSR (8.356) model, suggesting that the more random noise was involved in MLR model than that in other models. The 4-wavelength MLR model gave the highest \(r_{val}\) value and the lowest SEP value than the other models. The predictive ability of PCR model with 17 principal components was slightly better than that of PLSR model with 6 components in terms of \(r_{val}\) and SEP values. These results are consistent with the findings of Kuo et al. (2001b), who showed that the best calibration and prediction of BV were obtained by the MLR model with 9 wavelengths, which produced a \(r_{val}\) of 0.754 and a SEP of 0.0437, respectively. The samples had the readings of BV between 0.009 and 0.303. Although the calibration equation developed in this study had better predictive ability, the equation could only apply to a narrow range (0.009-0.220). Kuo et al. (2001b) also found that the change of performance in model-building and prediction by PLSR was not significant after removing the outlying observations. Therefore, PLSR was a robust approach. Moreover, PLSR might achieve the optimal model with a small number of components than PCR. Generally speaking, the performance in model-building and prediction for BV was still less accurate.

**Table 2. The model performance for starch-iodine blue value of residual liquid by different statistical methods.**

<table>
<thead>
<tr>
<th>Method</th>
<th>Term (^1)</th>
<th>(r_{cal}) (^2)</th>
<th>SEC (^3)</th>
<th>IRV (^4)</th>
<th>SEP(_{cv}) (^5)</th>
<th>SEP (^6)</th>
<th>(r_{val}) (^7)</th>
<th>Bias (^8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLR</td>
<td>4</td>
<td>0.900</td>
<td>0.024</td>
<td>82.042</td>
<td>0.0250</td>
<td>0.869</td>
<td>0.0002</td>
<td></td>
</tr>
<tr>
<td>PCR</td>
<td>17</td>
<td>0.932</td>
<td>0.021</td>
<td>18.331</td>
<td>0.023</td>
<td>0.841</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>PLSR</td>
<td>6</td>
<td>0.891</td>
<td>0.025</td>
<td>8.356</td>
<td>0.028</td>
<td>0.828</td>
<td>0.0006</td>
<td></td>
</tr>
</tbody>
</table>

\(^1\) Number of wavelengths or components.

\(^2\) \(r_{cal}\) = correlation coefficient of calibration set.

\(^3\) SEC = standard error of calibration.

\(^4\) IRV = index of random variation.

\(^5\) SEP\(_{cv}\) = standard error of cross-validation.

\(^6\) SEP = standard error of prediction.

\(^7\) \(r_{val}\) = correlation coefficient of validation set.

\(^8\) Bias = the average difference between predicted and chemical values.
The scatter plots of chemical value and NIR predicted value for both data sets are shown in Fig. 2. The number of wavelengths and components selected in calibration set was also used in validation set. When the validation set was used, the NIR values predicted by MLR model lie close to the ideal regression line, i.e. a line through the origin and with a slope of 1, than those predicted by PCR and PLSR models (Fig. 2B, D, and F). This finding also confirmed the best predictive ability of MLR model.

The residual plots for calibration and validation sets are shown in Figure 3. These residuals fluctuated in a random pattern around the base line 0, displaying no systematic tendency. This also indicates regression models are appropriate. All residuals calculated from MLR model based on the validation set randomly distributed between -0.05 and 0.05 (Fig. 3B). However, some residuals calculated from PCR and PLSR models fell outside the range between -0.05 and 0.05 (Fig. 3D and F).

In MLR model, the wavelengths at 2036, 2192, 2296, and 2316 nm were selected, indicating that principal spectral region of BV was within the longer wavelength region (Fig. 4). This was congruent with the finding of Su et al. (2001), who developed an optimal MLR model with wavelengths at 2264, 2280, 2296, 2332, 2356, 2380, 2416, 2424, and 2472 nm. However, the inflated variability of the estimated regression coefficients was found in MLR model, suggesting that the multicollinearity among the wavelengths reduced the precision of the estimated regression coefficients. In other words, the estimated regression coefficients tend to become erratic from sample to sample. PLSR developed a more stable calibration model. Consequently, these estimated regression coefficients calculated by PLSR were more close to the true parameter values (Kuo et al., 2001b).

Several authors have indicated that multicollinearity among the independent variables sometimes cause severe problems in multiple regression. Especially, the estimated regression coefficients could be very unstable (Montgomery and Peck, 1982; Neter et al., 1996). In this study, MLR exhibited extremely large value of IRV than the value in PCR and PLSR (Table 2), suggesting that the multicollinearity problems existed among the selected wavelengths when the MLR model was used. Although PCR solves the multicollinearity problem and eliminates the lesser components, some useful information may be lost through the discarded components (Geladi and Kowalski, 1986). However, PLSR may use the first fewer components to achieve the optimal model performance (Helland, 1988).

Conclusions

In this study, 4-wavelength MLR model gave the highest correlation coefficient and the lowest standard error of prediction than PCR and PLSR models. In other words, this MLR model had the best predictive ability in estimating the BV of rice. The predictive ability of the PCR model with 17 principal components was slightly better than that of the PLSR model with 6 components. However, The IRV value in MLR model was much larger than that in PCR model and PLSR model, indicating more random noise was involved in MLR model than that in other models. The performances of model-building and prediction demonstrate that calibration for BV is less accurate and NIRS can be used only for the first screening of BV in the rice breeding program.
Fig 2. The relationship between chemical and predicted values of starch-iodine blue value of residual liquid analyzed by MLR, PCR, and PLSR for the calibration set and validation set, respectively. 
Cal = calibration set, Val = validation set
Estimating the starch-iodine blue value of residual liquid of rice by near-infrared reflectance spectroscopy

Fig 3. The residual plot of starch-iodine blue value of residual liquid analyzed by MLR, PCR, and PLSR using calibration set and validation set, respectively.

Cal = calibration set, Val = validation set
Fig 4. The regression coefficients of principal wavelengths for starch-iodine blue value of residual liquid using MLR method.

References

Estimating the starch-iodine blue value of residual liquid of rice by near-infrared reflectance spectroscopy

581-607.


利用近紅外線反射光譜估計稻米炊飯液碘呈色度

郭寶錚 1) 洪梅珠 2)

(接受刊載日期：中華民國 90 年 8 月 22 日)

摘要：炊飯液碘呈色度此一特性可被用來預測稻米的嗜口性。本研究的目的在於評估利用近紅外線反射光譜研究估計稻米炊飯液碘呈色度的可能性並比較複線性迴歸，主成份迴歸及淨最小平方法在建立模式及預測能力上的表現。228 個水稻樣品包括秈、粳及糯稻被用來分析炊飯液碘呈色度，並建立檢量線，結果發現利用 4 個波長的複線性迴歸模式可產生最高的相關係數及最低的預測機差，也就是說其在建立模式及預測能力上表現最佳。利用 17 個主成份所構成的主成份迴歸模式又較 6 個成份所構成的淨最小平方迴歸模式為佳，但是複線性迴歸中 IRV 值(82.042)遠大於主成份迴歸(18.331)及淨最小平方迴歸(8.356)的 IRV 值，顯示複線性迴歸模式較其他模式包含更多的隨機干擾。綜合言之，就本研究得知，稻米炊飯液碘呈色度的檢量線並不夠準確，在水稻育種上近紅外線反射光譜對稻米炊飯液中碘呈色度僅能做初步估計之用。

關鍵詞：炊飯液碘呈色度，近紅外線反射光譜，共線性，複線性迴歸，主成份迴歸，淨最小平方迴歸

1) 國立中興大學農藝系副教授 (通訊作者)
2) 台中區農業改良場副研究員