JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

# Dose Detection of Radiated Rice by Infrared Spectroscopy and Chemometrics

YONGNI SHAO,[†] YONG HE,*[,†] AND CHANGQING WU[‡]

College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310029, China, and Department of Animal & Food Science, University of Delaware, Newark, Delaware 19716

Infrared spectroscopy based on sensitive wavelengths (SWs) and chemometrics was proposed to discriminate the nine different radiation doses (0, 250, 500, 750, 1000, 1500, 2000, 2500, and 3000 Gy) of rice. Samples ($n = 16$ each dose) were selected randomly for the calibration set, and the remaining 36 samples ($n = 4$ each dose) were selected for the prediction set. Partial least-squares (PLS) analysis and least-squares-support vector machine (LS-SVM) were implemented for calibration models. PLS analysis was implemented for calibration models with different wavelength bands including near-infrared (NIR) regions and mid-infrared (MIR) regions. The best PLS models were achieved in the MIR (400–4000 $cm^{-1}$) region. Furthermore, different latent variables (5−9 LVs) were used as inputs of LS-SVM to develop the LV-LS-SVM models with a grid search technique and radial basis function (RBF) kernel. The optimal models were achieved with six LVs, and they outperformed PLS models. Moreover, independent component analysis (ICA) was executed to select several SWs based on loading weights. The optimal LS-SVM model was achieved with SWs (756, 895, 1140, and 2980 $cm^{-1}$) selected by ICA and had better performance than PLS and LV-LS-SVM with the parameters of correlation coefficient ($r$), root-mean-square error of prediction, and bias of 0.996, 80.260, and $5.172 \times 10^{-4}$, respectively. The overall results indicted that the ICA was an effective way for the selection of SWs, and infrared spectroscopy combined with LS-SVM models had the capability to predict the different radiation doses of rice.

**KEYWORDS: Infrared spectroscopy; radiated rice; sensitive wavelengths; partial least-squares analysis; least-squares-support vector machine**

## INTRODUCTION

Cereal grains contribute to over 60% of the total world food production. Cereals are predominantly composed of carbohydrates, mostly in the form of starch, with a considerable amount of protein as well as some lipids, vitamins, and minerals (*1*). Rice in China is grown mostly in the south Chin Mountains and Huai River, especially in the Changjiang River and Zhujiang River areas. The plant area of rice is one-quarter of all the grain, and the output is half of the total in grain. So, rice is a staple food, and the concerns about rice are much greater than those for other foods. However, investigations have revealed that rice would be affected by mold damage and insect damage during postharvest storage. This extensive loss of rice has stimulated a great deal of research on minimizing these damages, such as spraying insecticides consisting of pyrethrins and malathion or fumigating rice with methyl bromide, ethylene oxide, and hydrogen cyanide. However, safety about the consumption of rice contaminated with insecticides and fumigating chemicals in humans is concerning.

$\gamma$-Radiation treatment has been applied in grain storage for the control of insect infestation, microbial contamination, or the prevention of postharvested biological activities, for example, ripening, germination, and sprouting, with the merits of simplifying the whole treatment process, minimizing the processing time, leaving no detrimental residue, and achieving this for a reasonable radiation cost. Radiation-treated grain may lead to different changes in nutritional contents depending on the different radiation doses. Generally, whether the grain is radiated cannot be told by obvious appearance indications. Methods employing techniques that include gas chromatography, mass spectrometry, spectroscopy, and DNA analysis may be applied to identify certain molecular and spectroscopic characteristics of rice that are altered as a result of radiation. There have been no literature works using infrared spectroscopy to analyze the different radiation doses of rice; also, the use of mid-infrared spectroscopy (MIRS) in cereal applications has been less.

Infrared spectroscopy techniques, such as near-infrared spectroscopy (NIRS) and MIRS, have been widely applied as nondestructive analytical methods with the advantages of quick analysis, small sample preparation, and low cost. Although infrared spectroscopy techniques cannot always provide defini-

* To whom correspondence should be addressed. Tel: +86-571-86971143. Fax: +86-571-86971143. E-mail: yhe@zju.edu.cn.
[†] Zhejiang University.
[‡] University of Delaware.

10.1021/jf8000058 CCC: $40.75 
Published on Web 05/13/2008

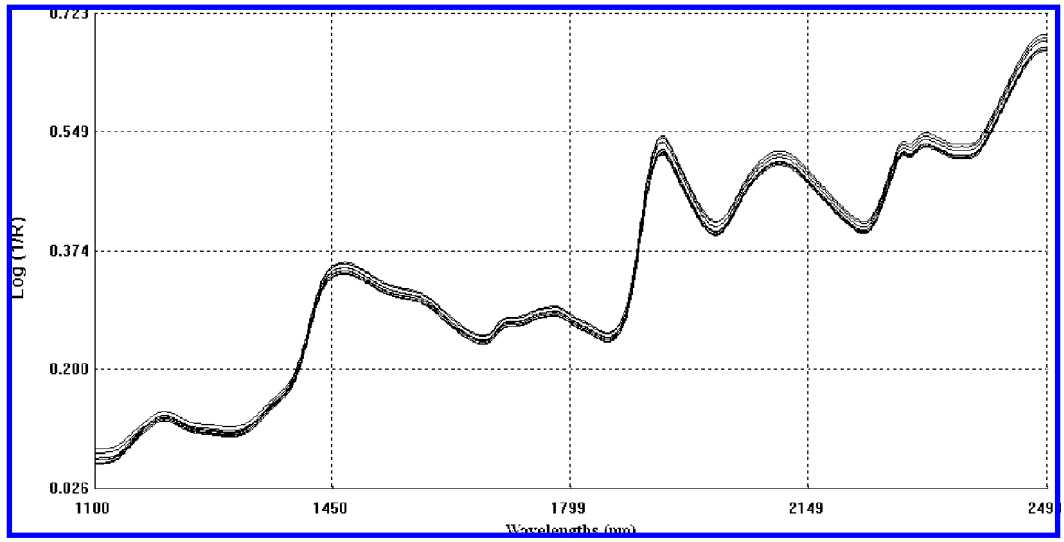


**Figure 1.** Original near-infrared spectral curves of radiated rice after nine different radiation doses.

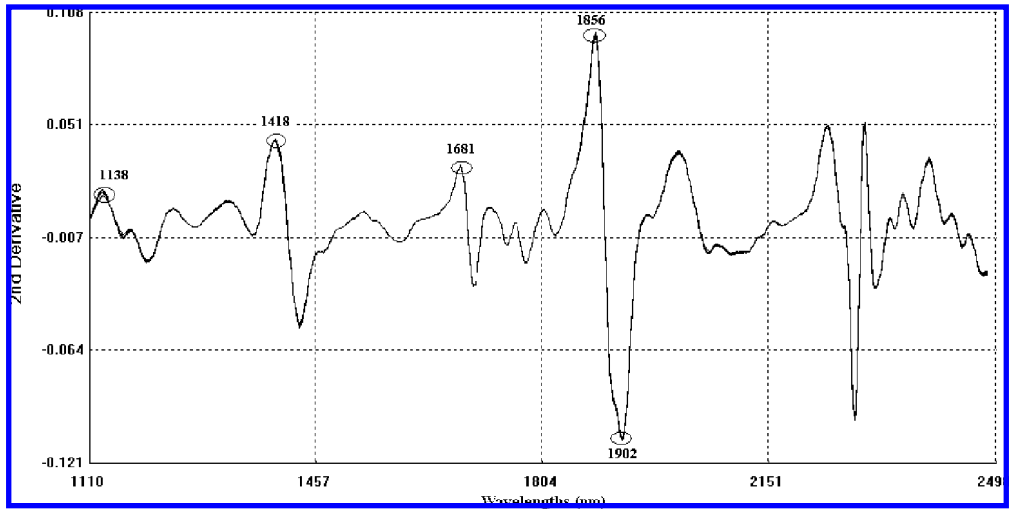


**Figure 2.** Near-infrared spectral curves of radiated rice after second derivative preprocessing.

tive compositional information about a food sample, they can often provide a means of screening food products for qualitative attributes without the involvement of time-consuming chemistry analyses in laboratories (*2*). Osborne et al. used NIRS for the authentication of Basmati rice (*3*). Delwiche and Graybosch built the identification model of waxy wheat using near-infrared reflectance spectroscopy (*4*). Wu et al. estimated the amino acid composition in the milled rice powder by near-infrared reflectance spectroscopy (*5*). Kim et al. attempted to use NIRS to authenticate Korean domestic and foreign rices (*6*). Wu and Shi studied the composition of single rice grains using near-infrared reflectance spectroscopy (*7*). Baye et al. used single kernel infrared spectroscopy to predict the maize seed composition (*1*).

The NIRS bands mainly corresponding to C−H, O−H, and N−H vibrations are originally from fundamental bands in MIR region. These bands are not directly visible since they are more severely superimposed than in the MIR spectra, resulting in low molar absorptivity of NIRS. In contrast, the MIRS bands have high molar absorptivity, and the peaks of MIRS are specific, sharp, and sensitive (*2*). MIRS are applied to detect compositional differences between samples on the basis of vibrations of various chemical groups at specific wavelengths. However, spectral reproducibility of MIRS, including the signal-to-noise ratio, is poorer as compared to NIRS (*8*). Thus, two techniques

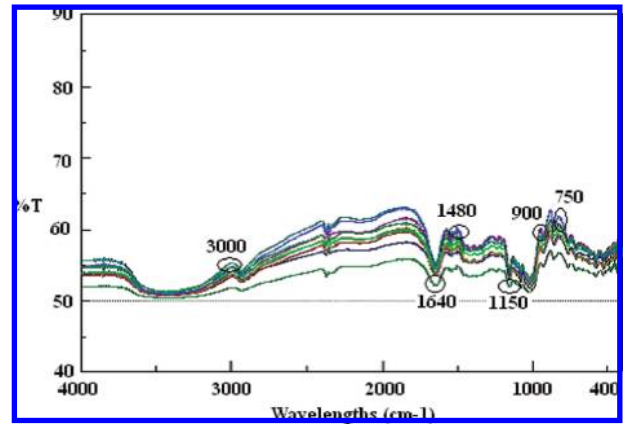


**Figure 3.** Mid-infrared spectral curves of radiated rice after nine different radiation doses.

have independent advantages and disadvantages and need to be considered according to different situations.

The objectives of this paper were (i) to study the feasibility of using infrared spectroscopy to predict the different radiation doses of rice; (ii) to compare the performance of different wavelength bands including the NIR region (1100–2500 nm) and the MIR region (400–4000 $cm^{-1}$) by partial least-squares (PLS) analysis; (iii) to compare the prediction precision using

**Table 1.** Prediction Results of Different Wavelength Regions by PLS Models

| data set | region | LVs | correlation coefficients | RMSEC (RMSEP) | bias | slope | offset |
|---|---|---|---|---|---|---|---|
| calibration | NIR | 7 | 0.976 | 116.428 | $5.631 \times 10^{-2}$ | 0.953 | 65.346 |
| | MIR | 6 | 0.985 | 97.125 | $2.507 \times 10^{-3}$ | 0.970 | 56.412 |
| prediction | NIR | 7 | 0.964 | 122.308 | $8.340 \times 10^{-2}$ | 0.929 | 68.124 |
| | MIR | 6 | 0.978 | 114.902 | $-6.032 \times 10^{-3}$ | 0.956 | 58.302 |

**Table 2.** Explained Variance of LVs in the MIR Region

| parameters | | | | | |
|---|---|---|---|---|---|
| LVs | 5 | 6 | 7 | 8 | 9 |
| EV[a] (%) | 89.203 | 95.120 | 96.245 | 97.862 | 98.493 |

[a] EV, explained variance.

different latent variables (5−9 LVs) for least-squares-support vector machine (LS-SVM); and (iv) to select the optimal sensitive wavelengths (SWs) for the development of portable instruments and online monitoring for commercial applications of different radiated doses of rice.

## MATERIALS AND METHODS

**Rice Samples.** The origins of the rice samples were unpolished (a unique rice type), and they were polished before they were radiated with $\gamma$-radiation. The treatment was performed in a $^{60}$Co irradiator at Zhejiang University with a dosage rate of 2.5 kGy/h, and nine different radiation doses dealt with the following doses: 0, 250, 500, 750, 1000, 1500, 2000, 2500, and 3000 Gy. After radiation, a total 180 rice samples were prepared as rice flours by milling using a flour mill, with 20 samples from each radiation dose; all flours were passed through a 100 mesh sieve. All samples were divided into calibration sets of 144 samples (16 samples for each dose) and prediction sets of 36 samples (four samples for each dose). No single sample was used in calibration and prediction sets at the same time. To compare the performance of different calibration models, the samples in the calibration and prediction sets were kept unchanged for all calibration models.

**Spectral Acquisition and Preprocessing.** Two spectrometers were used to measure the rice samples: a Foss NIRSystems 6500 (NIRSystems, Inc., Silver Spring, MD), with a spectral range of 1100–2500 nm, and a FT-IR spectrometer (FT/IR-4100 type A FT-IR Spectrometer, JASCO International Co., Ltd., Japan), with a spectral range of 7800–350 $cm^{-1}$. In NIR measurements, about 4 g of rice flour for each sample was scanned in duplicate in a small ring cup (NR-7073; internal diameter, 35 mm; depth, 9 mm). Each spectrum represented the average of 32 scans and was recorded as log (1/$R$) at 2 nm increments. For MIR measurements, each sample was mixed with potassium bromide (KBr) at a ratio of 1:49, and then, the mixture was compressed into slices and loaded on a slide holder. The spectral curves were obtained in transmission mode. Duplicates of each sample were scanned twice (rotating the ring cup to a different position). The average spectrum of each sample was used for further analysis.

Before the calibration stage, both NIR and MIR spectra data were preprocessed. The Savitzky−Golay smoothing was used to reduce the noise (*9, 10*), with a window width of 7 (3–1–3) points. The multiplicative scatter correction (MSC) was used to correct additive and multiplicative effects in the spectra (*11*). To analyze the influence of different radiation doses on rice, the internal quality parameters of radiated rice were used, including starch and protein.

**Partial Least Squares Analysis.** In the development of the PLS model, calibration models were built between the spectra and the radiation doses; full cross-validation was used to evaluate the quality and to prevent overfitting of the calibration models. The optimal number of LVs was determined by the lowest value of predicted residual error sum of squares (PRESS). The prediction performance was evaluated by the correlation coefficient ($r$) and root-mean-square error of calibration (RMSEC) or prediction (RMSEP). The ideal model should have a higher $r$ value and lower RMSEC and RMSEP values. The prediction set was applied to evaluate the accuracy of the models to classify rice samples according to different doses.

**Independent Component Analysis (ICA).** ICA was originally developed to deal with problems closely related to the cocktail party problem (*12*). As an effective approach to the separation of blind signals, ICA has recently attracted broad attention and has been successfully used in many fields, for example, medical signal analysis, image processing, dimension reduction, fault detection, and near-infrared spectral data analysis (*13–18*).

ICA is a well-established statistical signal processing technique that aims to decompose a set of multivariate signals into a base of statistically independent components with minimal loss of information content. The independent components are LVs, meaning that they cannot be directly observed, and the independent component must have non-Gaussian distributions. A chief explanation of noise-free ICA model could be written as the following expression:

$$\mathbf{x} = \mathbf{As} \tag{1}$$

where $\mathbf{x}$ denotes the recorded data matrix and $\mathbf{s}$ and $\mathbf{A}$ represent the independent components and the coefficient matrix, respectively. The goal of ICA is to find a proper linear representation of non-Gaussian vectors so that the estimated vectors are as independent as possible and the mixed signals can be denoted by the linear combinations of these independent components. The ICs were obtained by a high-order statistic, which is a much stronger condition than orthogonality. This goal is equivalent to finding a separating matrix $\mathbf{W}$ that satisfies

$$\mathbf{s} = \mathrm{WX} \tag{2}$$

where $\hat{s}$ is the estimation of $\mathbf{s}$.

The separating matrix $\mathbf{W}$ can be trained as the weight matrix of a two-layer feed-forward neural network in which $\mathbf{x}$ is the input and $\hat{s}$ is the output.

There are lots of algorithms for performing ICA (*19, 20*). Among these algorithms, the fast fixed-point algorithm (FastICA) is highly efficient for performing the estimation of ICA, which was developed by Hyvärinen and Oja (*21*).

FastICA was chosen for ICA and carried out in Matlab 7.0 (The Math Works, Natick, United States) according to the following steps (*16*):

(1) Choose an initial random weight vector $\mathbf{w}$ (0) and let $k = 1$, where $\mathbf{w}$ is an $l$-dimensional (weight) vector in the weight matrix $\mathbf{W}$, and $k$ is an irrelevant constant.
(2) Let $w(k) = E\{xg[w(k-1)^T x]\} - E\{g'[w(k-1)^T x]\}w(k-1)$, where $g$ is the first derivative of the function $G$, and $G$ is practically any nonquadratic function.
(3) Let $w(k) = w(k)/\|w(k)\|$.
(4) If $|w(k)^T w(k-1)|$ is not close enough to 1, let $k = k + 1$ and go back to step 2. Otherwise, output the vector $\mathbf{w}$ ($k$).

**LS-SVM.** LS-SVM can work with linear or nonlinear regression or multivariate function estimations in a relatively fast way (*22–24*). It uses a linear set of equations instead of a quadratic programming (QP) problem to obtain the support vectors (SVs). The details of the LS-SVM algorithm can be found in the literature (*25, 26*). The LS-SVM model can be expressed as:

$$y(x) = \sum_{k=1}^{N} \alpha_k K(x, x_k) + b \tag{3}$$

where $K(x, x_i)$ is the kernel function, $x_i$ is the input vector, $\alpha_i$ is a Lagrange multiplier called the support value, and $b$ is the bias term.

In the model development using LS-SVM and the radial basis function (RBF) kernel, the optimal combination of $gam(\gamma)$ and $sig2(\sigma^2)$ parameters was selected when resulting in smaller root-mean-square errors of cross-validation (RMSECV). In this study, $gam(\gamma)$ was

**Table 3.** Prediction Results of Different Radiation Dose with Different LVs by LV-LS-SVM Models in the MIR Region

| data set | LVs | ($\gamma$, $\sigma^2$) | correlation coefficients | RMSEC (RMSEP) | bias | slope | offset |
|---|---|---|---|---|---|---|---|
| calibration | 5 | (42.1, 12.4) | 0.986 | 98.635 | $-1.145 \times 10^{-3}$ | 0.972 | 56.075 |
| | 6 | (116.6, 24.6) | 0.992 | 88.421 | $8.125 \times 10^{-4}$ | 0.984 | 50.467 |
| | 7 | (54.5, 27.8) | 0.984 | 101.427 | $4.204 \times 10^{-3}$ | 0.968 | 55.346 |
| | 8 | (188.7, 21.8) | 0.979 | 108.604 | $8.456 \times 10^{-3}$ | 0.958 | 58.024 |
| | 9 | (254.9, 20.4) | 0.972 | 112.410 | $5.301 \times 10^{-2}$ | 0.945 | 58.697 |
| prediction | 5 | (42.1, 12.4) | 0.978 | 109.127 | $-5.064 \times 10^{-3}$ | 0.956 | 58.945 |
| | 6 | (116.6, 24.6) | 0.989 | 95.763 | $-3.621 \times 10^{-3}$ | 0.978 | 54.568 |
| | 7 | (54.5, 27.8) | 0.975 | 111.408 | $-7.042 \times 10^{-2}$ | 0.951 | 60.042 |
| | 8 | (188.7, 21.8) | 0.969 | 117.042 | $8.362 \times 10^{-2}$ | 0.939 | 61.347 |
| | 9 | (254.9, 20.4) | 0.964 | 121.487 | $-9.139 \times 10^{-2}$ | 0.929 | 61.842 |

optimized in the range of $2^{-1}$–$2^{10}$ and 2–$2^{15}$ for $sig2(\sigma^2)$ with adequate increments. These ranges were chosen from previous studies where the magnitude of parameters was optimized. The grid search had two steps: The first step was for a crude search with a large step size, and the second step was for the specified search with a small step size. The free LS-SVM toolbox (LS-SVM v 1.5, Suykens, Leuven, Belgium) was applied with MATLAB 7.0 to develop the calibration models.

## RESULTS AND DISCUSSION

**Features of Spectra and Statistics of Starch and Protein.** **Figure 1** shows the NIR spectral curves of radiated rice with nine different radiation doses. The range of starch was 20.963−24.124, and protein was 7.669−7.303, verified with different radiation doses. The trend of different doses in NIR region is similar, so we treated them with a second derivative (**Figure 2**). The prominent features are that the absorption peaks were associated with the first overtone of C−H stretching of the starch around 1131–1155 nm, the first overtone of O−H symmetric and asymmetric stretching of water around 1418 nm, the first overtone of the C−H asymmetric stretching of lipid around 1681 nm, and the O−H bending and asymmetric stretching combination band of water around 1852–1906 nm (*27*). As compared to the NIR region, the curves are complex in the MIR region (**Figure 3**). The MIR range covered two large water bands, $\nu_2$ and $\nu_L$, centered around 1640 and 750 $cm^{-1}$, the amide band I around 1558–1705 $cm^{-1}$, the amide bands II around 1480–1613 $cm^{-1}$, and the amide bands III around 1200–1280 $cm^{-1}$, and the phosphate groups covalently bind to casein proteins around 1060–1100 $cm^{-1}$. There also exist some peaks around 3000, 1480, 1150, and 900 $cm^{-1}$. In both NIR and MIR regions, the spectral curves did not show an obvious rule, and a calibration procedure was required to analyze the properties of radiated rice using sophisticated statistical techniques.

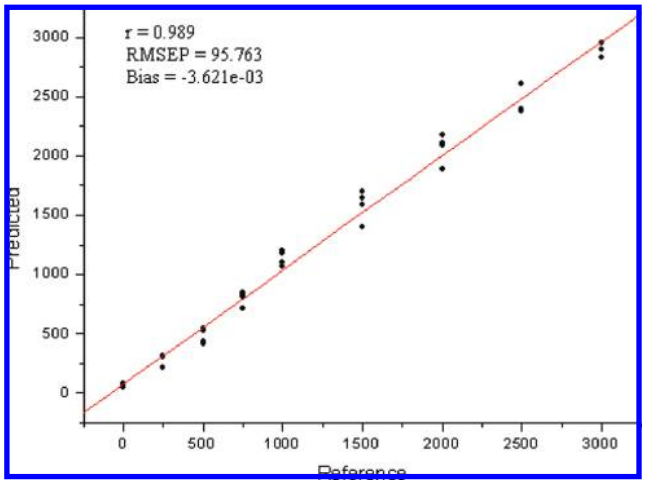


**Figure 4.** Predicted vs reference values for different radiated doses of rice by LV-LS-SVM models in the mid-infrared region.

**PLS Models.** A PLS model was developed using the preprocessed spectra data by Savitzky−Golay smoothing and MSC, and calibration models were built between the spectra and the different radiation doses. Considering the spectroscopy categories related to the wavelength bands, the NIR region (1100–2500 nm) and MIR region (400–4000 $cm^{-1}$) were separated to establish two models. Different LVs were applied to build the calibration models, and no outliers were detected in the calibration set during the development of PLS models. The results of calibration and prediction sets are shown in **Table 1**. The models built with MIR region 400–4000 $cm^{-1}$ turned out to be the best for the prediction of different radiated doses. The $r$, RMSEP, and bias for the MIR model were 0.978, 114.902, and $-6.032 \times 10^{-3}$, respectively. The original spectral transmission plots (**Figures 1** and **2**) show that a high amount of absorption bands is present in NIR spectra, but it is not directly visible since bands are more severely superimposed than in the MIR spectra. In the MIRS region, the curves are smoother, and there are great differences among each dose sample because the MIRS provide more information about frequencies and intensities, which are stronger, than NIRS does (*8*). Chemometrics techniques such as LV-LS-SVM and SW-LS-SVM from ICA were used to separate useful information from irrelevant contributions. The result based on the MIR region is better than NIR, even if calibration models with NIR could lead to less satisfactory results; however, their use could be preferable, due to the easier sample preparation in NIR spectroscopy than in MIR spectroscopy.

**LV-LS-SVM Models.** LVs obtained from PLS were applied as inputs of LS-SVM models to improve the training speed and reduce the training error of the MIR model because the training time increased with the square of the number of training samples and linearly with the number of variables. From the aforementioned analysis of the performance of PLS models, the LVs from the MIR region were used as new eigenvectors to enhance the features of spectra and reduce the dimensionality of the spectra data matrix. Several LVs were extracted from the spectra of 180 samples. **Table 2** shows the explained variance of $Y$ (different radiation dose) of the first 5−9 LVs. The variance of the first five LVs could explain more than 89% of the total variance, and the ninth LV only interpreted an additional 0.631%, which contributed not so much as the other aforementioned LVs. So, it was not necessary for the consideration of less than five LVs or more than nine LVs. The LS-SVM models with 5−9 LVs were developed separately to find the best number of LVs.

Before the LS-SVM calibration model can be built, three steps are crucial for the optimal input feature subset, proper kernel function, and the optimal kernel parameters. First, the 5−9 LVs obtained from PLS analysis must be used as the input data set. Second, RBF could handle the nonlinear relationships between the spectra and the target attributes. Finally, two important

**Table 4.** Prediction Results of Different Radiation Dose by SW-LS-SVM Models in the MIR Region

| data set | ($\gamma$, $\sigma^2$) | correlation coefficients | RMSEC (RMSEP) | bias | slope | offset |
|---|---|---|---|---|---|---|
| calibration | (46.7, 12.8) | 0.998 | 78.056 | $5.564 \times 10^{-5}$ | 0.996 | 48.259 |
| prediction | (46.7, 12.8) | 0.996 | 80.260 | $5.172 \times 10^{-4}$ | 0.992 | 47.352 |

parameters *gam*($\gamma$) and *sig2*($\sigma^2$) should be optimal for RBF kernel function as aforementioned in multivariate analysis.

The performance of these MIR models was evaluated by 36 samples in the prediction set, and the results are shown in **Table 3**. With a comparison of the results for calibration and prediction sets, the best performance was achieved with six LVs. The *r*, RMSEP, and bias for prediction sets were 0.989, 95.763, and $-3.621 \times 10^{-3}$, respectively. The results for calibration and prediction sets showed that LS-SVM models outperformed PLS models (**Tables 1** and **3**). **Figure 4** shows the predicted vs reference charts. The solid line is the regression line corresponding to the correlation between the prediction and the reference values.

**SW-LS-SVM Models.** ICA was applied for the selection of SWs, which could reflect the main features of the raw absorbance spectra. FastICA was used to the preprocessed spectra data, and the main absorbance peaks and valleys were indicated by the spectra of ICs. The SWs were selected by the weights of the first four ICs, whose wavelengths with the highest weights of each IC were selected as the SWs. Four SWs were selected corresponding to four ICs. They were 756, 895, 1140, and 2980 $cm^{-1}$. To evaluate the performance of SWs, they were applied as the input data matrix to develop the SW-LS-SVM models. The prediction results for calibration and prediction sets are shown in **Table 4**, and the *r*, RMSEP, and bias were 0.996, 80.260, and $5.172 \times 10^{-4}$, respectively. **Figure 5** shows the predicted vs reference charts. The SW-LS-SVM models achieved a better performance as compared to the best LV-LS-SVM models in both calibration and prediction sets. Wavelengths at 756 $cm^{-1}$ were close to the absorbance peaks, 750 $cm^{-1}$ due to the water bands $\nu_L$. The wavelength at 1140 $cm^{-1}$ was close to the amide bands III around 1200–1280 $cm^{-1}$, and phosphate groups covalently bind to casein proteins around 1060–1100 $cm^{-1}$. Therefore, the selection of SWs was suitable for such a situation in the present study, and the effectiveness of SWs was also validated. The SWs represented most of the features of the original spectra and could replace the whole wavelength region to predict different radiation doses of rice. Furthermore, the SWs might be important for the development of portable instruments and online monitoring for commercial applications of different radiated doses of rice.

Many factors affect the precision and reliability of calibration and prediction, such as sample preparation, accuracy of the reference data, etc. The differences in rice sample preparations may influence the results, where the rice sample was milled to flour (*28*). Delwiche et al. used milled whole grain samples and showed that flour spectra were superior to those whole ones (*28, 29*). Shu et al. also reported that the milled rice flour was superior to brown rice flour in calibration for alkali spreading value (*30*). In addition, the grain sizes of the milled rice still vary with different rice genotypes, so all flours pass through a fixed mesh sieve, and in homogeneous grain, size may result in better calibration performance.

The prediction of different radiated doses of rice was successfully performed based on infrared spectroscopy and chemometric methods of PLS and LS-SVM models. Savitzky–Golay smoothing and MSC were used as preprocessing methods. The PLS model with the value of *r*, RMSEP, and bias in prediction set, 0.978, 114.902, and −6.032e-03, respectively, was better in the MIR region than the NIR region. The numbers of LVs of LS-SVM were selected by the PLS from the MIR region, and the optimal LS-SVM model was achieved with six LVs; the *r*, RMSEP, and bias in prediction set were 0.989, 95.763, and $-3.621 \times 10^{-3}$, respectively, which outperformed PLS model. ICA was executed to select several SWs based on loading weights, and the optimal LS-SVM model was achieved with SWs (756, 895, 1140, and 2980 $cm^{-1}$) selected by ICA and had better performance than PLS and LV-LS-SVM with the parameters of *r*, RMSEP, and bias in the prediction set equal to 0.996, 80.260, and $5.172 \times 10^{-4}$, respectively. The overall results indicted that the ICA was a powerful way for the selection of SWs, and infrared spectroscopy combined with LS-SVM models had powerful capability to predict the radiation dose of rice. Further optimization and interpretation of the SWs selection method will be needed to improve the calibration generalization and stability for practical applications.

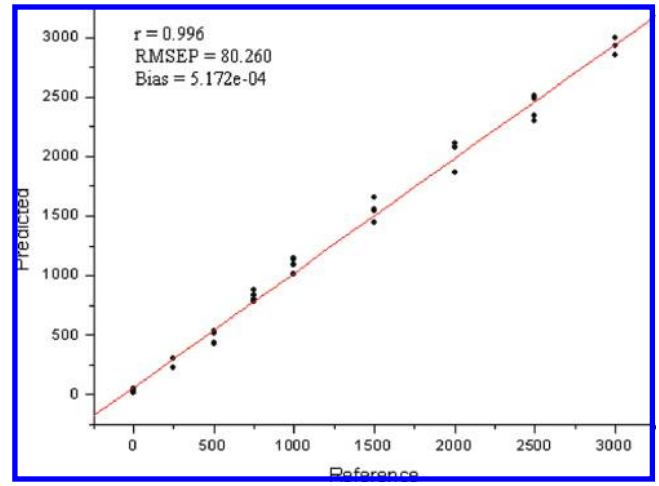


**Figure 5.** Predicted vs reference values for different radiated doses of rice by SW-LS-SVM models in the mid-infrared region.

## LITERATURE CITED

(1) Baye, T. M.; Pearson, T. C.; Settles, A. M. Development of a calibration to predict maize seed composition using single kernel near infrared spectroscopy. *J. Cereal Sci.* **2006**, *43*, 236–243.

(2) Reid, L. M.; Woodcock, T.; O'Donnell, C. P.; Kelly, J. D.; Downey, G. Differentiation of apple juice samples on the basis of heat treatment and variety using chemometric analysis of MIR and NIR data. *Food Res. Int.* **2005**, *38*, 1109–1115.

(3) Osborne, B. G.; Mertens, B.; Thompson, M.; Fearn, T. Authentication of Basmati rice using near infrared spectroscopy. *J. Near Infrared Spectrosc.* **1993**, *1*, 77–83.

(4) Delwiche, S. R.; Grayboschf, R. A. Identification of waxy wheat by near-infrared reflectance spectroscopy. *J. Cereal Sci.* **2002**, *35*, 29–38.

(5) Wu, J. G.; Shi, C. H.; Zhang, X. M. Estimating the amino acid composition in the milled rice powder by near-infrared reflectance spectroscopy. *Field Crops Res.* **2002**, *75*, 1–7.

(6) Kim, S. S.; Rhyu, M. R.; Kim, J. M.; Lee, S. H. Authentication of rice using near-infrared reflectance spectroscopy. *Cereal Chem.* **2003**, *80*, 346–349.

(7) Wu, J. G.; Shi, C. H. Prediction of grain weight, brown rice weight and amylase content in single rice grains using near-infrared reflectance spectroscopy. *Field Crops Res.* **2004**, *87*, 13–21.

(8) Chung, H.; Ku, M. S.; Lee, J. S. Comparison of near-infrared and mid-infrared spectroscopy for the determination of distillation property of kerosene. *Vib. Spectrosc.* **1999**, *20*, 155–163.

(9) Savitzky, A.; Golay, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627–1639.
(10) Gorry, P. A. General least-squares smoothing and differentiation by the convolution (Savitzky-Golay) method. *Anal. Chem.* **1990**, *62*, 570–573.
(11) Helland, I. S.; Naes, T.; Isaksson, T. Related versions of multiple scatter correction. *Chemom. Intell. Lab. Syst.* **1995**, *29*, 233–241.
(12) Amari, S.; Cichocki, A.; Yang, H. H. A new learning algorithm for blind signal separation. *Adv. Neural Inf. Process. Syst.* **1996**, *8*, 757–763.
(13) Hyvarinen, A. Sparse code shrinkage: Denoising of nongaussian data by maximum likelihood estimation. *Neural Comput.* **1999**, *11*, 1739–1768.
(14) Hoyer, P. O.; Hyvarinen, A. Independent component analysis applied to feature extraction from colour and stereo images. *Network-Comput. Neural Syst.* **2000**, *11*, 191–210.
(15) Hyvarinen, A.; Hoyer, P. O. Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural Comput.* **2000**, *12*, 1705–1720.
(16) Chen, J.; Wang, X. Z. A new approach to near-infrared spectral data analysis using independent component analysis. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 992–1001.
(17) Bi, X.; Li, T. H.; Wu, L. Application of independent component analysis to the IR spectra analysis. *Chem. J. Chin. Univ.* **2004**, *25*, 1023–1027.
(18) Shao, X. G.; Wang, G. Q.; Wang, S. F.; Su, Q. D. Extraction of mass spectra and chromatographic profiles from overlapping GC/MS signal with background. *Anal. Chem.* **2004**, *76*, 5143–5148.
(19) Hyvarinen, A.; Karhunen, J.; Oja, E. *Independent Component Analysis*; Wiley: New York, 2001.
(20) Lee, T. W. *Independent Component Analysis: Theory and Application*; Kluwer: Boston, MA, 1998.
(21) Hyvarinen, A.; Oja, E. Independent component analysis: Algorithms and applications. *Neural Networks* **2000**, *13*, 411–430.
(22) Suykens, J. A. K.; Vanderwalle, J. Least squares support vector machine classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300.
(23) Borin, A.; Ferrao, M. F.; Mello, C.; Maretto, D. A.; Poppi, R. J. Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk. *Anal. Chim. Acta* **2006**, *579*, 25–32.
(24) Chen, Q. S.; Zhao, J. W.; Fang, C. H.; Wang, D. M. Feasibility study on identification of green, black and Oolong teas using near-infrared reflectance spectroscopy based on support vector machine (SVM). *Spectrochim. Acta, Part A* **2007**, *66*, 568–574.
(25) Guo, H.; Liu, H. P.; Wang, L. Method for selecting parameters of least squares support vector machines and application. *J. Syst. Simul.* **2006**, *18*, 2033–2036–2051.
(26) Chen, Q. S.; Zhao, J. W.; Fang, C. H.; Wang, D. M. Feasibility study on identification of green, black and Oolong teas using near-infrared reflectance spectroscopy based on support vector machine (SVM). *Spectrochim. Acta, Part A* **2007**, *66*, 568–574.
(27) Barton, F. E.; Himmelsbach, D. S.; McClung, A. M.; Champagne, E. T. Two-dimensional vibration spectroscopy of rice quality and cooking. *Cereal Chem.* **2002**, *79*, 143–147.
(28) Bao, J. S.; Shen, Y.; Jin, L. Determination of thermal and retrogradation properties of rice starch using near-infrared spectroscopy. *J. Cereal Sci.* **2007**, *46*, 75–81.
(29) Delwiche, S. R.; McKenzie, K. S.; Webb, B. D. Quality characteristics in rice by near-infrared reflectance analysis of whole-grain milled samples. *Cereal Chem.* **1996**, *73*, 257–263.
(30) Shu, Q. Y.; Wu, D. X.; Xia, Y. W.; Gao, M. W.; McClung, A. Analysis of grain quality characters in small ground brown rice samples by near infrared reflectance spectroscopy. *Sci. Agric. Sin.* **1999**, *19*, 92–97.

**Received for review January 1, 2008. Revised manuscript received March 8, 2008. Accepted March 10, 2008. This study was supported by the National Science and Technology Support Program (2006BAD10A09), the Teaching and Research Award Program for Outstanding Young Teachers in Higher Education Institutions of MOE, People's Republic of China, the Natural Science Foundation of China (Project 30671213), the Specialized Research Fund for the Doctoral Program of Higher Education (Project 20040335034), and the Science and Technology Department of Zhejiang Province (Projects 2005C21094 and 2005C12029).**

JF8000058