Determination of total quality of black tea fanning grade using near-infrared spectroscopy

A T E Prasetya, N A Wibowo, and F S Rondonuwu

1Faculty of Science and Mathematics, Universitas Kristen Satya Wacana, Diponegoro 52-60, Salatiga 50711, Indonesia
2Research Center for Near Infrared and Nanotechnology, Universitas Kristen Satya Wacana, Diponegoro 52-60, Salatiga 50711, Indonesia
ferdy@staff.uksw.edu

Abstract. Generally, the quality assessment of black tea from Camellia sinensis is determined by three important aspects including appearance, liquor, and infusion. Traditionally, the assessment is carried out by highly trained people using an organoleptic method. However, this method requires good physical and psychological health conditions as it relies on the sensitivity of their taste and flavor sensory. The method requires three to five people to avoid unfairness and to guarantee the level of consistency. In this paper, a new technique is reported for tea quality determinations using near-infrared spectroscopy followed by partial least squares (PLS) regression. The technique was developed by collecting 69 grade fanning (FANN) black tea samples of crush-tear-curl (CTC) type from 10 estates in Java and Sumatera. The samples were randomly divided into two groups consisting of 50 and 19 samples. The first group was used for developing a training set while the latter was applied for a validation set. Near-infrared spectra in the range of 4000-7500 cm⁻¹ (interval 4 cm⁻¹) were collected from the 69 samples without sample modifications or pretreatments. Then, the second derivative spectra of the training set were used to build a PLS model and validation procedures. The training set utilized results obtained from an organoleptic assessment of the same sample to calculate the predictor’s coefficient along the energy axis. The validation set then examined the prediction accuracy of the near-infrared and PLS model. The results show that the near-infrared and PLS technique successfully built a regression model based on the training set samples and properly predicted the scores obtained by the organoleptic determinations of the three quality parameters including appearance, liquor, and infusion using samples from the validation set. It can be concluded that the nondestructive measurement by near-infrared spectroscopy followed by PLS regression calculations successfully determined the black tea quality of FANN. Also, it provides an opportunity for the rapid determination of tea quality from a large number of tea samples at the market site.

1. Introduction
Quality is an important aspect of both local and international trade. Quality control is the only way that a product can be properly appreciated by traders and consumers. Unfortunately, not all of the quality can be measured easily. Determining the tea quality is important in order to justify the selling price. At present, the quality determination technique relies on a taster panel method consisting of 3-5 people who have the ability to sensory test the samples [1]. This method requires taster panels to be at one place at a certain time, physically in a good condition, and environmentally conducive. Since this method highly
relied on the human senses, the results may sometimes lead to subjective judgments. Another method is a wet chemical analysis using high-performance liquid chromatography (HPLC). However, a wet chemical analysis is relatively expensive and takes days to provide suitable results.

The parameters used in determining the quality of tea include: (1) particle size, (2) density, (3) appearance, (4) liquor, and (5) infusion. The first and second aspects can be examined easily using standard and simple equipment. However, aspects 3-5 are usually measured by using human senses that may be strongly influenced by the physical, physiological, and environmental conditions of the taster panels [1]. When any one of these conditions is not satisfied then the determination may be inaccurate. Given these conditions, it is necessary to develop an analytical method for quality determination that can be done rapidly, non-destructively, and reliably.

Fourier transform near-infrared spectroscopy (NIRS) is an emerging technique that can be used to analyze both solid and liquid samples. The advantages of this technique are the flexibility of utilizing the electromagnetic spectrum at 800-2500 nm (equivalent to the energy 12500-4000 cm⁻¹), which is related to the overtone and combination band of molecular vibration energy. NIRS has advantages such as: it is fast in analyzing the results, it does not require specific sample preparations, and it is non-destructive [2]. Commercially, this technique has been applied to specific areas of analysis, including: foodstuffs [3–5], beverages [2,6–10], and medications [11–13]. Since most of the NIR spectra are complicated due to the spectral overlap, a chemometric analysis is absolutely necessary to extract important information conveyed by the spectral data. Partial least squares regression (PLSR) is one of the chemometric methods that can be applied to the near-infrared spectroscopic data [14–16]. The aim of PLSR is to predict the dependent variable (Y-matrix) from the predictor (X-matrix) and describe the common structure of both X and Y. This is particularly important when the number of predictors is larger than the number of dependent variables as in the case of the NIRS data set.

This paper aims to report a model of non-destructive and rapid determination of appearance, liquor, infusion, and total values using FT-NIRS combined with PLSR utilizing the spectra of 69 black tea samples from *Camellia sinensis* produced for the fanning (FANN) grade under crush-tear-curl (CTC) type.

2. Methodology

In this study, 69 samples of black tea grade FANN powder from CTC type were used. The samples were obtained from PT. Perkebunan Nusantara. Each sample was stored in a package and placed in a room with a temperature of 25°-27°C. Prior to the measurement, each sample package was poured in a container and mixed for a homogeneity so as to obtain relatively equal particle density. The homogeneous samples then had about 3-4 grams taken to be inserted into a measuring vial for spectral acquisition.

Measurements were performed using a near-infrared spectrometer (Buchi Nirfex 500 solid®) and a diffuse reflection mode was used to measure the solid tea sample. Vials were used as the sample holders during measurements. Each spectrum obtained from the measurement resulted from an average of 32 scans in an energy span of 4000-7500 cm⁻¹ with interval measurements of 4 cm⁻¹. The spectra obtained by using NIRS contained many overlapping peaks from the overtones and combination vibrations of molecules. Therefore, pre-processing including spectral smoothing, baseline corrections, and spectral resolution improvements were performed to provide better PLS calculations. Spectral smoothing aimed to reduce the noise that appeared along the spectra. Baseline corrections were done by performing first derivatives to each of the trans-reflectance spectra. Improvements to the spectral resolutions were done by performing a 2º derivative to all spectral data sets. In this case, PLSR was applied to predict the dependent variable, Y, as defined by the organoleptic method from the predictor matrix, X, constructed from a set of NIR spectral data. In PLS, X and Y were decomposed simultaneously in a way that $X = TP' + E$ with an identity matrix of $I = T^2T$ and $Y$ is estimated as $Y = TP'$. $B$ is a diagonal matrix with a regression weight, while $C$ represents the weight matrix of the dependent variables. PLSR seeks two sets of weights $w$ and $c$ to obtain a linear combination of columns $X$ and $Y$ with maximum covariance.
The specific goal is to find $t = Xw$ and $u = Yc$ with a constrain of $w^T W = 1, t^T t = 1$ and $t^T u$ is maximal [17–20]. The PLSR calculates PLS factors directly as linear combinations of the original variables [21] and it was executed by the use of a built-in SIMPLS algorithm coded in a MATLAB R2017A (Mathwork, USA) environment to the entire 2nd derivative spectra in the region of 4000-7500 cm$^{-1}$. The spectra consisted of 69 rows (from 69 samples) in which 50 spectra were used as trained sets (SDT), and the remaining 19 spectra were used as validation sets (SDV).

### 3. Results and Discussion

The spectral data obtained from the 69 black tea samples clearly show pattern similarities with various random baselines as shown in Figure 1. The baseline shift, a typical phenomenon in a reflectance mode, is generated by the variation of particle size in the solid sample leading to light scattering during spectral acquisitions by NIRS [22]. The spectral structure clearly exhibits broad absorption characteristics due to the overlapping absorption from the overtones and combination vibrations of molecular bonds as shown in Table 1. To obtain an accurate analysis, first and second derivatives were applied to all of the reflectance spectra. The first derivative removes spectrum baselines while the second one improves the spectral resolution, so that the peaks (valleys) that provide important information will become prominent, hence, significantly improve the spectral analysis [23]. Spectral smoothing was applied using the Savitzky-Golay method employing polynomial order 3 with frame size 17 before the 1st and 2nd derivation to avoid noise magnification.

**Table 1.** Molecular bonding and their corresponding vibrational energy in the NIR region.

| Description             | Energy (cm$^{-1}$) |
|-------------------------|--------------------|
| Combination C-H Stretching | 4545-4065          |
| Combination N-H and O-H Stretching | 5000-4545          |
| Overtone C-H             | 6173-5556          |
| Overtone N-H and O-H     | 7143-6250          |

**Figure 1.** Reflectance spectra of 69 samples. The spectra show various baselines with similar patterns. The energies of corresponding molecular bonding are indicated for clarity.

**Figure 2.** The second derivative spectra of 69 samples.
PLS calculations were applied to the 2nd derivative spectral calibration data set. One important parameter in the PLS model is the number of components, $n$, used in calculations. The optimum value of $n$ provides a good regression coefficient, and this optimum value is achieved when the mean square error of the validation set (MSE$_{\text{Val}}$) is minimum. Please note that the mean square error of the training set (MSE$_{\text{Train}}$) will approach zero as $n$ increases. This result can easily be seen from Figure 3 (a-d), in that as the number of $n$ increases the MSE$_{\text{Train}}$ ($\bullet$) approaches 0, while the MSE$_{\text{Val}}$ ($\square$) decreases to a certain value of $n$ and then increases again. This means that when $n$ is larger than the $n$ optimum, the predictive ability becomes weak. Table 2 shows a summary of the statistical results of calibration models constructed from the analysis using the PLSR method. The optimal $n$ value for quality of appearance, liquor, infusion, and the total is 12, 9, 10, and 14, respectively. Each tea quality has a different $n$ for the construction of its calibration model. The calibration model is said to be good if the coefficient determination, $R^2$, approaches 1 and the MSE$_{\text{Val}}$ value approaches 0. The results show that the calibration model is able to predict the four tea qualities in the validation set. It can be seen from the coefficient determination of the validation set that the $R^2$ for the quality of appearance, liquor, infusion, and the total are 0.98, 0.76, 0.96, 0.99, respectively. These results indicate that the training data set provides good information and can be used for the calibration model.

Table 2. Choice of the $n$ for minimum MSE$_{\text{Val}}$ and its corresponding MSE$_{\text{Train}}$ and ($R^2$).

| Quality  | $n$ | MSE$_{\text{Train}}$ | MSE$_{\text{Val}}$ | $R^2_{\text{Val}}$ |
|---------|----|----------------------|-------------------|-------------------|
| Appearance | 12 | 0.06                 | 0.87              | 0.98              |
| Liquor   | 9  | 0.08                 | 0.37              | 0.76              |
| Infusion | 10 | 0.04                 | 0.18              | 0.96              |
| Total    | 14 | 0.03                 | 1.46              | 0.99              |

(a) (b)
Figure 3. MSE$_a$ (●) and MSE$_v$ (□) of appearance (a), liquor (b), infusion (c), and total (d). The trend line (−) in MSE$_v$ assists in determining $n$ for optimum accuracy of predictions.

The correlations between the NIR predictions and the observed values are shown in Figure 4 (a-d). It can be seen that two sets, i.e. SDT (◊) and SDV (◆), reside in a linear line. SDT is a data set used to construct calibration models, while SDV is a data set used to validate the calibration model. In these figures, most of the validation data points reside around a linear line with. These results show that the prediction accuracy for appearance, liquor, infusion, and total is similar to that obtained from calibration data set.
Figure 4. Calibration curve showing comparisons between the NIR predictions vs. the organoleptic test scores (observed) among the validation data set (SDV, ◆) and the training data set (SDT, ◂) for 4 qualities of appearance (a), liquor (b), infusion (c), and total (d).

4. Conclusion
Determining the quality scores for appearance, liquor, infusion, and total can be done well using NIRS combined with PLSR methods. This analysis successfully built a calibration model and is able to predict the quality scores of the four black tea (Camellia sinensis) FANN grade in the curl, tear, and cut (CTC) type. Overall, this method can be an alternative for determining the quality of tea rapidly, non-destructively, and potentially with low costs.

5. Acknowledgments
FSR would like to thank Direktorat Riset dan Pengabdian Masyarakat, Direktorat Jenderal Penguatan Riset dan Pengembangan, Kementrian Riset, Teknologi dan Pendidikan Tinggi Republik Indonesia for his research funding in 2016 and the Satya Wacana Research Funding in 2017. An acknowledgment is also given to Mr. Tafip Haryanto who was very helpful in preparing the laboratory references.

6. References
[1] Yan S H 2007 NIR evaluation of the quality of tea and its market price Spectrosc. Eur. 19 16–9
[2] Tan S M, Luo R, Zhou Y, Gong H and Tan Z 2012 Rapid and non-destructive discrimination of tea varieties by near infrared diffuse reflection spectroscopy coupled with classification and regression trees Afr. J. Biotechnol. 11 2303–12
[3] Ding X, Ni Y and Kokot S 2015 NIR spectroscopy and chemometrics for the discrimination of pure, powdered, purple sweet potatoes and their samples adulterated with the white sweet potato flour Chemom. Intell. Lab. Syst 144 17–23
[4] López A, Arazuri S, García I, Mangado J and Jarén C 2013 A review of the application of near-infrared spectroscopy for the analysis of potatoes J. Agric. Food Chem. 61 5413–24
[5] Meng Y, Wang S, Cai R, Jiang B and Zhao W 2015 Discrimination and content analysis of fritillaria using near infrared spectroscopy J. Anal. Methods Chem. 2015 1–8
[6] Arai K 2013 Method for Tealeaves Quality Estimation Through Measurements of Degree of Polarization, Leaf Area Index, Photosynthesis Available Radiance and Normalized Difference Vegetation Index for Characterization of Tealeaves IJARAI 2 17–24
[7] Ligor M, Kornysöva O, Maruška A and Buszewski B 2008 Determination of Flavonoids in Tea and Rooibos Extracts by TLC and HPLC JPC–J. Planar Chromat. 21 5 355–60
[8] Fu X S, Xu L, Yu X P, Ye Z H and Cui H F 2013 Robust and automated internal quality grading of a Chinese green tea (Longjing) by near-infrared spectroscopy and chemometrics J. Spectrosc. 1 1–7
[9] Kumar S, Panchariya P C, Prasad B and Sharma A L 2013 Discrimination of Indian tea varieties using UV-VIS-NIR spectrophotometer and pattern recognition techniques Int. J. of Comp. Sci. and Comm. 2 15–9
[10] Xie L, Ye X, Liu D and Ying Y 2009 Quantification of glucose, fructose and sucrose in bayberry juice by NIR and PLS Food Chem. 114 1135–40
[11] Khorasani M, Amigo J M M, Sonnergard J, Olsen P, Bertelsen P and Rantanen J 2015 Visualization and prediction of porosity in roller compacted ribbons with near-infrared chemical imaging (NIR-CI).J. Pharm. Biomed. Anal. 109 11–7
[12] Moffat A C, Assi S and Watt R A 2010 Identifying Counterfeit Medicines Using near Infrared Spectroscopy J. Near Infrared Spec. 18 1–15
[13] Zhang C and Su J 2014 Application of near infrared spectroscopy to the analysis and fast quality assessment of traditional Chinese medicinal products Acta Pharm. Sin. B 4 182–92
[14] Næs T and Martens H 1988 Principal component regression in NIR analysis: Viewpoints, background details and selection of components J. Chemom. 2 155–67
[15] Martens H and Næs T 1989 Multivariate calibration vol 4 (Chichester: John Wiley & Sons)
[16] Miranda A A, Le Borgne Y A and Bontempi G 2008 New routes from minimal approximation error to principal components Neural Process. Lett. 27 197–207
[17] Wold S, Sjöström M and Eriksson L 2001 PLS-regression: a basic tool of chemometrics Chemom. Intell. Lab. Syst. 58 109–30
[18] Tobias R D 1995 An Introduction to Partial Least Squares Regression Proceedings of the Twentieth Annual SAS Users Group International Conference, Cary, NC: SAS Institute Inc 1250–1257
[19] Abdi H 2003 Partial Least Squares (PLS) Regression Encyclopedia for research methods for the social sciences pp 792–5
[20] Nadler B and Coifman R R 2005 Partial least squares, Beer’s law and the net analyte signal: statistical modeling and analysis J. Chemom. 19 45–54
[21] De Jong S 1993 SIMPLS: An alternative approach to partial least squares regression Chemom. Intell. Lab. Syst. 18 251–63
[22] Norris K H and Davies A M C 2012 Examining diffuse reflection and transmission spectra more thoroughly: Part 1. Instrument noise Spectroscopy Asia 8 12–7
[23] Fahri Y A, Budiastra I W N and Purwanto 2016 Penggolongan Mangga Gedong Gincu Berdasarkan Rasio Kandungan Gula Asam Menggunakan Prediksi Near Infrared Spectroscopy JTEP 4 31–6