

Nondestructive Estimation of Total Free Amino Acid in Green Tea by Near Infrared Spectroscopy and Artificial Neural Networks*

Zhiming Guo¹, Liping Chen^{1,**}, Chunjiang Zhao¹,
Wenqian Huang¹, and Quansheng Chen²

¹ National Engineering Research Center of Intelligent Equipment for Agriculture,
Beijing 100097, China

² School of Food & Biological Engineering,
Jiangsu University, Zhenjiang 212013, China
chenlp@nercita.org.cn

Abstract. The total free amino acid of green tea was nondestructive estimated by near-infrared (NIR) spectroscopy combined with multivariate calibration, compared the performance of back propagation neural networks (BP-NN) and partial least squares (PLS) regression analysis. The original spectra of tea samples in wavelength range of 10000-4000cm⁻¹ were acquired. Spectral pretreatment methods were applied to reduce the systematic noise, and enhance the contribution of the chemical composition. The model was optimized by cross validation, and its performance was evaluated according to root mean square error of prediction (RMSEP) and correlation coefficient (R) in prediction set. Experimental results showed that the performance of BP-NN model was superior to the performances of PLS model, from the point of view of the predictive ability. The optimal results of the BP-NN model with multiplicative scatter correction spectral pretreatment were achieved as follow: RMSEP=0.246 and R_p=0.958 in the prediction set, respectively. It can be concluded that NIR spectroscopy combined with BP-NN has significant potential in quantitative analysis and monitor of free amino acid content in green tea.

Keywords: Near infrared spectroscopy, back propagation neural network, nondestructive estimation, free amino acid, green tea.

1 Introduction

Tea is considered one of the most popular beverages consumed worldwide. It is made from the leaves of the plant *Camellia sinensis* (L.). There are different tea varieties with different external qualities and inner qualities due to morphologic and chemical

* Financial supported by National Natural Science Foundation (No.31071324) and National High Technology Research and Development Program of China (No. 2011AA100703).

** Corresponding author.

diversities [1, 2]. Tea has important physiological properties and potential health benefits due to the presence of compounds such as polyphenols, amino acids, vitamins, carbohydrates, caffeine, and purine alkaloids, and those chemical constituents influence the taste and flavor of tea infusions [3, 4]. Tea flavor is enhanced by the umami taste of amino acids. The predominant amino acid in tea infusion is a unique amino acid, theanine, which exists only in the free form and accounts for about 50% of the total free amino acids (TFAA). It has been reported that theanine can decrease the level of norepinephrine and serotonin in brain [5], and intake of theanine by hypertensive rats results in decreased blood pressure [6]. The total content of amino acids in green tea is the greatest in all kinds of tea. It has been demonstrated that there is a relationship between the quality of green tea and the amino acid contents [7-9]. However, amino acid levels in tea vary during each manufacturing stage. Young two-leaf flushes are typically chosen for manufacturing the finest tea because theanine accumulates in growing shoots [10]. Therefore, a rapid assay for TFAA in tea is essential for quality control, evaluation and classification in tea processing.

There many descriptions of procedures for analysis of amino acids in green tea based on high-performance liquid chromatography (HPLC) [11,12], capillary electrophoresis (CE) [13], fluorometric flow-injection analytical [14] and ninhydrin colorimetric assay. Although these assays mentioned above are accurate to quantify individual compounds, they are labor intensive, time-consuming and requires large quantities of reagents for quality control purposes.

Near infrared reflectance (NIR) spectroscopy has been proved to be a powerful analytical tool that can be acted as a replacement of time-consuming chemical assay and there is no need for sample preparation or trained staff to perform the analysis routinely. NIR spectroscopy can be used online as well as in the field using the appropriate instrument. It has been introduced as a rapid and non-destructive method for evaluating tea quality parameters (such as polyphenols, catechins, gallic acid, caffeine and theobromine) [15-18]. The few published studies on green tea focused mainly on the prediction of bioactive components. In the literature no studies were found testing NIR spectroscopy for the prediction of free amino acid composition of green tea. Although the linear partial least squares (PLS) analyses is the most common techniques for spectral calibration and prediction [19], with PLSR being the most accurate, other nonlinear techniques e.g. artificial neural network (ANN) have got much less attention and were rarely explored for the spectral analysis in tea sciences. There has been no report on using artificial neural networks for studying the TFAA dependence on the main components of the tea samples.

The objective of this work was to evaluate the ability of NIR spectra to predict the development of TFAA in whole green tea leaves. In addition, the prediction capabilities of different calibration models, including PLS and ANN were investigated and compared for their use in modeling the TFAA. The results of this study may open a wide range of applications in process control or in laboratory routine for the quantification of TFAA in tea.

2 Materials and Methods

2.1 Preparation of Samples

A total of 90 green tea samples were collected from different geographical regions of PR China (i.e. Jiangsu, Henan, Zhejiang, Hunan, Fujian, Anhui, Jiangxi, Sichuan, Guangdong and Yunnan) for experiments. Based on the results of the pre-test, the moisture content of each commercial green tea was $4.0 \pm 0.5\%$. The chemical compositions of these tea samples varied with their species, cultivation conditions such as weather, temperature, moisture, latitude, and season, and process of manufacturing [20]. Samples were stored under controlled circumstance (20°C and 45% relative humidity) before measurements.

2.2 Reference Measurement

The content of TFAA in tealeaves was measured by a ninhydrin colorimetric assay, which can refer to GB/T 8314-2002, a national standard method used to determine free amino acids content of tea infusions in China. The calibration standard is L-theanine, which was obtained from Sigma (St. Louis, MO, USA). In a 25-ml volumetric flask, 1.0 ml of each tea infusions was added. This was followed by addition of 0.5 ml of 1/15 mol/L phosphate buffer solution (pH 8.0) and 0.5 ml of 2% ninhydrin solution containing 0.8 mg/ml of $\text{SnCl}_2 \cdot 2\text{H}_2\text{O}$. The mixtures in the volumetric flasks were then placed on a boiling water bath for 15 min. The probes were quickly cooled with cold water, and adjusted to 25 ml with water. After they were left to stand still for 10 min, a UV-VIS spectrophotometer (UV-1601, Beijing Rayleigh Analytical Instrument Co., Ltd, China) used to detect the absorbance of the reaction solution in a 10mm light-path cell at 570 nm.

2.3 NIR Spectra Acquisition

An AntarisTM II Method Development Sampling (MDS) system (Thermo Fisher Scientific Inc., Madison, WI, USA) spectrometer was used to measure diffuse reflectance spectra from $10,000$ to 4000 cm^{-1} at 4 cm^{-1} resolution by co-adding 32 scans. Spectra were using an integrating sphere and an empty cell as a reference. FT-NIR measurements were performed with a high sensitivity InGaAs detector with a tungsten lamp as the NIR source. The reflectance values (R_v) were converted into absorbance (A) values using the formula $A = \log(1/R_v)$ for data analysis.

The standard sample accessory holder was used for performing the tea spectra collection. The sample accessory holder is sample cup specifically designed by Thermo Fisher Scientific. For each tea sample, $10 \pm 0.1\text{g}$ of dry tealeaves was filled into the sample cup in the standard procedure depending upon the bulk density of materials. When spectra collecting, tea sample was collected one times every rotating the cup 120 degree angle, thus, each sample was collected three times. The average of these three spectra was used in the next analysis.

2.4 Spectral Data Processing

NIR spectra could be affected by the physical properties of the samples and other environmental noises. Thus, it is necessary to perform mathematical processing to reduce the systematic noise, and enhance the contribution of the chemical composition [21]. Pretreatment is usually performed in order to furnish more parsimonious PLS models, by avoiding the presence of certain spectral trends before regression. In the present work, it was only applied to the experimental NIR data set, where scattering corrections are frequently applied. Spectral derivatives may also be employed to improve resolution and to highlight the selectivity towards a particular analyte when strong multicollinearity is present. Fig.1a presents the raw spectra profile of green tea sample and the raw spectra require to be preprocessed.

In this study, several spectral preprocessing methods were applied comparatively. These methods are standard normal variate transformation (SNV), multiplicative scatter correction (MSC), first derivative and second derivative. SNV is a mathematical transformation method used to remove slope variation and correct scatter effects in spectra [22]. Each spectrum is corrected individually by first centering the spectral values, and then the centered spectrum is scaled by the standard deviation calculated from the individual spectral values. MSC is another important procedure for the correction of scatter light [23]. It is used to modify the additive and multiplicative effects in the spectra on the basis of different particle sizes. First and second derivatives focus on eliminating baseline drifts and enhancing small spectral differences. Comparing with these preprocessing methods, MSC method was as good as SNV, better than first and second derivatives. The reason was that dry green tea particle solid was easy to bring in scatter light. MSC and SNV spectral preprocessing methods have better ability in correcting light scatter, and can also remove slope variation. MSC pretreated spectra was shown in Fig. 1b.

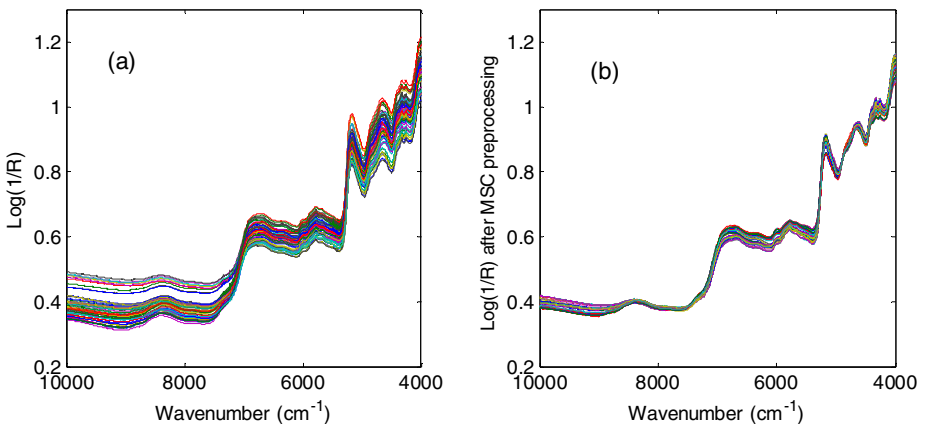


Fig. 1. Spectra of green tea obtained from (a) raw data and (b) MSC preprocessing data

2.5 ANN Description

ANN is multivariate calibration tools, which are especially suited for nonlinear systems. ANN is a powerful analytical tool because of the flexible learning algorithm, diverse network topology, fast learning algorithm and high error tolerance [24]. These advantages will lead to more sophisticated applications of ANN in food science, particularly in sensory analysis and analytical chemistry. Back propagation (BP) is the most commonly used algorithm to train feedforward neural networks for spectral data interpretation [25]. In this study, a multi-layer perception (MLP) based feedforward ANN which uses back propagation learning algorithm, was applied for modeling TFAA concentration.

The network consists of an input layer, hidden layers and an output layer. Fig. 2 illustrates the architecture of the network that is most commonly used with the back propagation algorithm-the multilayer feed-forward network. The determination of number of neurons in hidden layers is very important as it affects the training time and generalization property of neural networks. A higher value of neurons in hidden layer may force the network to memorize the patterns which it has seen during training whereas a lower value of neurons in hidden layer would waste a great deal of training time in finding its optimal representation [26]. There is no general rule for selecting the number of neurons in a hidden layer. The most popular approach to finding the optimal number of neurons in hidden layer is by trial and error [27]. In this study, trial and error approach was used to determine the optimum neurons in hidden layer of the network.

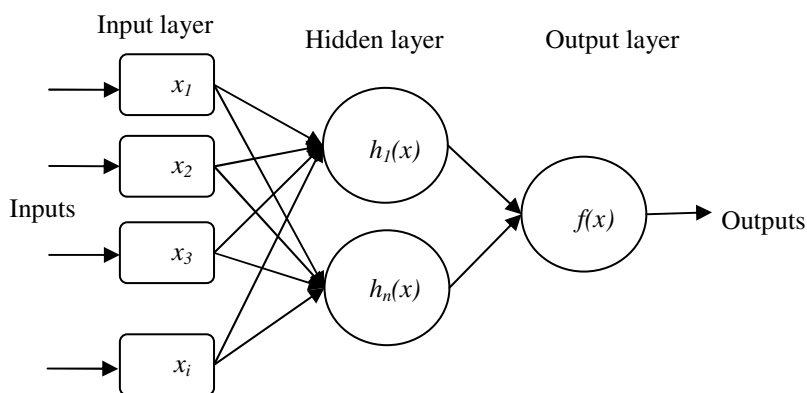


Fig. 2. Illustration of a feed-forward back propagation neural network containing input, hidden and output layers

2.6 Software

All computations were performed using a Matlab (version 7.9) Toolbox (Mathworks Inc., USA) under Windows XP. For the spectral acquisition Result Integration 3.0 (Antaris II System, Thermo Electron Co., USA) was used.

3 Results and Discussion

3.1 Reference Data Description

Green tea samples were divided randomly into calibration sets of 60 samples and prediction sets of 30 samples. Table 1 shows the descriptive statistics (mean, standard deviation and range) for calibration and prediction sets. As seen from Table 1, the range of reference measurements results of TFAA in the calibration set almost covers the range in the prediction set, and their standards deviations between the calibration and prediction sets are no significant differences. The reference values of these four properties had a broad range of variation, which was helpful for the calibrations. Therefore, the distribution of the samples is appropriate both in the calibration and the prediction sets.

Table 1. Reference measurement of free amino acid content (g/g, %) in calibration and prediction sets

Subset	No. ^a	Range	Mean	SD ^b
Calibration set	60	0.976~5.127	2.986	0.905
Prediction set	30	1.236~4.898	2.939	0.883

^aNo: Number of samples; ^bSD: Standard deviation.

3.2 Models Evaluation

The performance of the calibration model was evaluated in terms of the root mean square error of cross-validation (*RMSECV*). For *RMSECV*, a leave-one-sample-out cross-validation was performed: the spectrum of one sample of the calibration set was deleted from this set, and a model was built with the remaining spectra of the calibration set. The left-out sample was predicted with this model, and the procedure was repeated with leaving out each sample of the calibration set. Root mean square error of prediction (*RMSEP*) was used to evaluate the performance of the prediction set in the prediction process. Correlation coefficient (*R*) between the predicted and the measured value were calculated for both the calibration and the prediction sets. Generally, a good model should have lower *RMSECV*, *RMSEP* and higher *R* but small differences between *RMSEP* and *RMSECV*.

3.3 PLS Models

A PLS model performance evaluation was performed using leave-one-out cross validation for the prediction of TFAA in tea samples. It is generally known that the number of PLS factors and the spectra preprocessing are two critical factors in calibrating model for the application of PLS algorithm [18]. In the experiment, the number of PLS factors and the spectral preprocessing were optimized by cross validation, and, they were determined by the lowest value of *RMSECV*. MSC and

SNV spectral pretreatments were used to produce reasonable results for the spectra, respectively. Different PLS factors 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 for each model through different preprocessing methods are presented in Table 2. With a comparison of results for calibration and prediction by the aforementioned evaluation standards, the PLS models based on MSC spectral pretreatment turned out to be the optimal results which were achieved as follows: $R_c = 0.927$ and $RMSECV = 0.337$ in the calibration set; $R_p = 0.925$ and $RMSEP = 0.323$ in the prediction set. Fig. 3 is the scatter plot of reference measurements and NIR predicted results by the optimal PLS model.

Table 2. Results of calibration and prediction for PLS and BP-NN models with MSC and SNV spectral pretreatments

Method	Pretreatment	PCs ^a	Calibration		Prediction	
			R_c	$RMSECV(\%)$	R_p	$RMSEP(\%)$
PLS	Raw	8	0.898	0.424	0.890	0.401
PLS	SNV	7	0.925	0.344	0.910	0.356
PLS	MSC	7	0.927	0.337	0.925	0.323
BP-NN	SNV	3	0.972	0.207	0.950	0.265
BP-NN	MSC	4	0.979	0.182	0.958	0.246

^a principal component factors

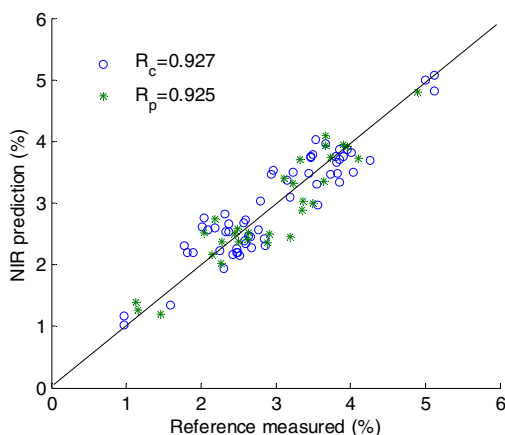


Fig. 3. PLS results: predicted vs. measured TFAA content for calibration (o) and prediction (*) set

3.4 BP-NN Models

In order to improve the training speed and reduce the training error, after spectra pretreatment, principal components (PCs) issued from principal component analysis (PCA) for dimensionality reduction were applied as input of BP-NN models because the training time increased with the square of the number of training samples and linearly with the number of variables (dimension of spectra). The PCs from the NIR spectra was used as new eigenvectors to TFAA. Four and three PCs were obtained by the criterion for an increment of explained variance lower than 0.25% [28], and the cumulative reliability was 99.39% and 99.28% for MSC and SNV spectral data, respectively.

In the model development by using BP-NN method, three layers of BP-NN were arranged. The number of PCs was also optimized according to cross validation. The output of BP-NN was TFAA content in tea. The hidden nodes were optimized by ‘trial and error’, and optimal of hidden nodes is evaluated by the minimal mean square error (MSE) value [29]. Several network architectures were tested by varying the number of neurons in the hidden layer with different initial weights. Adjusting by many times, learning rate factor and momentum factor were all set 0.1; all initial weights were set 0.3; scale function was set the ‘tanh’ function. The permitted regression error was set 0.01, and the maximal iterative epochs were set 1000 times. The optimal performance could be achieved, when 4 PCs and 8 nodes in the hidden layers were included for MSC pretreatment, and when 3 PCs and 8 nodes in the hidden layers were included for SNV pretreatment. With these optimal parameters, the BP-NN model, based on MSC spectral pretreatment, could obtain better prediction results with $R_c = 0.979$ and $RMSECV = 0.182$ in the calibration set, and $R_p = 0.958$ and $RMSEP = 0.246$ in the prediction set. Fig. 4 is the scatter plot of reference measurements and NIR predicted results by the optimal BP-NN model with MSC spectral pretreatment.

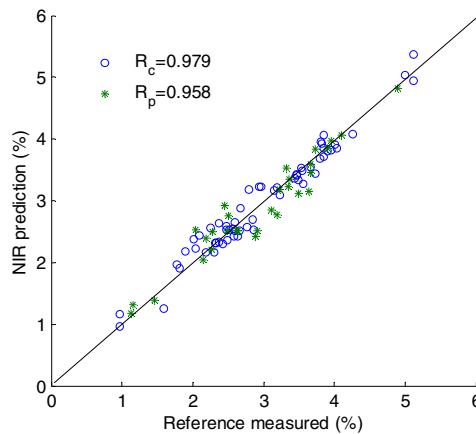


Fig. 4. Predicted values of calibration (o) and prediction (*) set against reference values using BP-NN method

3.5 Discussion of the Results

The scope of the paper is to compare the performances of linear PLS and nonlinear BP-NN analyses for the prediction accuracy of TFAA in green tea. The PLS models achieved an acceptable results for the prediction of TFAA content of green tea, and these results indicated that it was feasible to utilize spectroscopic technique for the detection and monitoring of free amino acid composition in tea leaves. The BP-NN models performed a better performance than that of PLS models, which indicated that the latent nonlinear information was helpful to improve the prediction performance. The performance of BP-NN models was a little better than that of PLS models. ANNs are suitable for solving nonlinear problems. This suggests that there are nonlinear relationships between the TFAA content and absorption of NIR spectra, more work would be done to discover the useful information or effective wavelength or wavelength bands for the nondestructive determination of TFAA content in green tea leaves.

4 Conclusion

Near Infrared spectroscopy combined with a suitable preprocessing method and with BP-NN regression is shown to be a promising method for predicting free amino acid composition in whole tea samples. The BP-NN models were developed for the prediction of properties of interest, and BP-NN models outperformed PLS models. The correlation coefficient R_p and $RMSEP$ in prediction set by BP-NN models were 0.958 and 0.246 for TFAA, respectively. The NIR technique can estimate free amino acid compositions in whole green tea, nondestructively with good accuracy and has enormous potential in improving the efficiency for quality control, evaluation and classification in tea processing.

Based on results achieved in this study, it is recommended to adopt BP-NN analysis as the proper calibration method to predict TFAA in green tea. However, further studies are needed to provide in depth interpretation of the successful measurement of tea properties that do not possess direct spectral response in NIR region.

Acknowledgements. This research was financially supported by National Natural Science Foundation (No.31071324) and National High Technology Research and Development Program (No. 2011AA100703).

References

1. Moreda-Pineiro, A., Fisher, A., Hill, S.J.: The classification of tea according to region of origin using pattern recognition techniques and trace metal data. *Journal of Food Composition and Analysis* 16(2), 195–211 (2003)
2. Chen, Q., Guo, Z., Zhao, J.: Identification of green tea's (*Camellia sinensis* (L.)) quality level according to measurement of main catechins and caffeine contents by HPLC and support vector classification pattern recognition. *Journal of Pharmaceutical and Biomedical Analysis* 48(5), 1321–1325 (2008)

3. Naithani, V., Nair, S., Kakkar, P.: Decline in antioxidant capacity of Indian herbal teas during storage and its relation to phenolic content. *Food Research International* 39(2), 176–181 (2006)
4. Hung, Y.-T., Chen, P.-C., Chen, R.L.C., Cheng, T.-J.: Sequential determination of tannin and total amino acid contents in tea for taste assessment by a fluorescent flow-injection analytical system. *Food Chemistry* 118(3), 876–881 (2010)
5. Juneja, L.R., Chu, D.-C., Okubo, T., Nagato, Y., Yokogoshi, H.: L-theanine: a unique amino acid of green tea and its relaxation effect in humans. *Trends in Food Science & Technology* 10(6), 199–204 (1999)
6. Kuroda, Y., Hara, Y.: Antimutagenic and anticarcinogenic activity of tea polyphenols. *Mutation Research/Reviews in Mutation Research* 436(1), 69–97 (1999)
7. Sakanaka, S., Juneja, L.R., Taniguchi, M.: Antimicrobial effects of green tea polyphenols on thermophilic spore-forming bacteria. *Journal of Bioscience and Bioengineering* 90(1), 81–85 (2000)
8. Ding, Y., Yu, H., Mou, S.: Direct determination of free amino acids and sugars in green tea by anion-exchange chromatography with integrated pulsed amperometric detection. *Journal of Chromatography A* 982(2), 237–244 (2002)
9. Alcazar, A., Ballesteros, O., Jurado, J.M., Pablos, F., Martin, M.J., Vilches, J.L.: Differentiation of green, white, black, oolong, and Puerh teas according to their free amino acids content. *Journal of Agricultural and Food Chemistry* 55, 5960–5965 (2007)
10. Kim, T., Lee, Y.K., Park, S.G., Choi, I.S., Ban, J.O., Park, H.K.: L-theanine, an amino acid in green tea, attenuates β -amyloid-induced cognitive dysfunction and neurotoxicity: Reduction in oxidative damage and inactivation of ERK/p38 kinase and NF- κ B pathways. *Free Radical Biology and Medicine* 47(11), 1601–1610 (2009)
11. Reich, E., Schibli, A., Widmer, V., Jorns, R., Wolfram, E., DeBatt, A.: HPLC methods for identification of green tea and green tea extract. *Journal of Liquid Chromatography and Related Technologies* 29, 2141–2151 (2006)
12. Wang, L., Xu, R., Hu, B., Li, W., Sun, Y., Tu, Y.: Analysis of free amino acids in Chinese teas and flower of tea plant by high performance liquid chromatography combined with solid-phase extraction. *Food Chemistry* 123(4), 1259–1266 (2010)
13. Aucamp, J.P., Hara, Y., Apostolides, Z.: Simultaneous analysis of tea catechins, caffeine, gallic acid, theanine and ascorbic acid by micellar electrokinetic capillary chromatography. *Journal of Chromatography A* 876, 235–242 (2000)
14. Kato, M., Gyoten, Y., Sakai-Kato, K., Toyooka, T.: Rapid analysis of amino acids in Japanese green tea by microchip electrophoresis using plastic microchip and fluorescence detection. *Journal of Chromatography A* 1013(1), 183–189 (2003)
15. Luybaert, J., Zhang, M.H., Massart, D.L.: Feasibility study for the use of near infrared spectroscopy in the qualitative and quantitative analysis of green tea, *Camellia sinensis* (L). *Analytica Chimica Acta* 478(2), 303–312 (2003)
16. Chen, Q., Zhao, J., Zhang, H., Wang, X.: Feasibility study on qualitative and quantitative analysis in tea by near infrared spectroscopy with multivariate calibration. *Analytica Chimica Acta* 572(1), 77–84 (2006)
17. Sinija, V.R., Mishra, H.N.: FT-NIR spectroscopy for caffeine estimation in instant green tea powder and granules. *LWT - Food Science and Technology* 42(5), 998–1002 (2009)
18. Chen, Q., Zhao, J., Chaitep, S., Guo, Z.: Simultaneous analysis of main catechins contents in green tea (*Camellia sinensis* (L.)) by Fourier transform near infrared reflectance (FT-NIR) spectroscopy. *Food Chemistry* 113(4), 1272–1277 (2009)

19. Kasemsumran, S., Du, Y.P., Maruo, K., Ozaki, Y.: Improvement of partial least squares models for in vitro and in vivo glucose quantifications by using near-infrared spectroscopy and searching combination moving window partial least squares. *Chemometrics and Intelligent Laboratory Systems* 82(2), 97–103 (2006)
20. Ngetich, W.K., Stephens, W.: Responses of tea to environment in kenya. 1. Genotype environment interactions for total dry matter production and yield. *Experimental Agriculture* 37(3), 333–342 (2001)
21. Lin, H., Chen, Q., Zhao, J., Zhou, P.: Determination of free amino acid content in *Radix Pseudostellariae* using near infrared (NIR) spectroscopy and different multivariate calibrations. *Journal of Pharmaceutical and Biomedical Analysis* 50(5), 803–808 (2009)
22. Fearn, T., Riccioli, C., Garrido-Varo, A., Guerrero-Ginel, J.E.: On the geometry of SNV and MSC. *Chemometrics and Intelligent Laboratory Systems* 96(1), 22–26 (2009)
23. Ely, D.R., Thommes, M., Teresa Carvajal, M.: Analysis of the effects of particle size and densification on NIR spectra. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* 331(1), 63–67 (2008)
24. Huang, Y., Kangas, L.J., Rasco, B.A.: Applications of artificial neural networks (ANNs) in food science. *Critical reviews in food science and nutrition* 47(2), 113–126 (2007)
25. Liu, Y., Sun, X., Ouyang, A.: Nondestructive measurement of soluble solid content of navel orange fruit by visible–NIR spectrometric technique with PLSR and PCA-BPNN. *LWT - Food Science and Technology* 43(4), 602–607 (2010)
26. Benardos, P.G., Vosniakos, G.-C.: Optimizing feedforward artificial neural network architecture. *Engineering Applications of Artificial Intelligence* 20(3), 365–382 (2007)
27. Abutaleb, A.S.: A neural network for the estimation of forces acting on radar targets. *Neural Networks* 4(5), 667–678 (1991)
28. Urbano-Cuadrado, M., de Castro, M.D.L., Pérez-Juan, P.M., García-Olmo, J., Gómez-Nieto, M.A.: Near infrared reflectance spectroscopy and multivariate analysis in enology: Determination or screening of fifteen parameters in different types of wines. *Analytica Chimica Acta* 527(1), 81–88 (2004)
29. Svozil, D., Kvasnicka, V., Pospichal, J.: Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems* 39(1), 43–62 (1997)