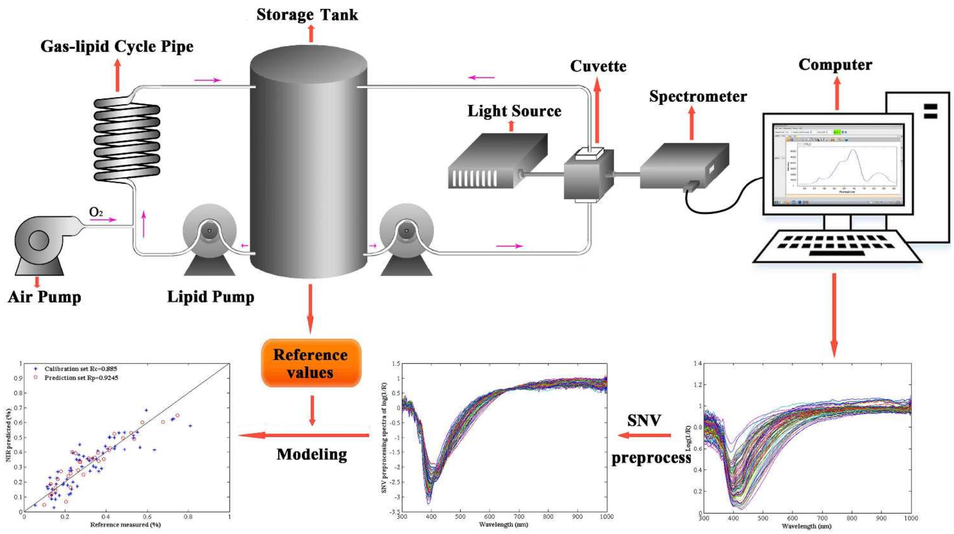
RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances



Schematic diagram for in situ monitoring tea oxidation.
320x172mm (300 x 300 DPI)

In situ monitoring total polyphenols content during tea extract oxidation using a portable spectroscopy system with variables selection algorithms

Wenxiu Pan, Jiewen Zhao, Quansheng Chen*, Lianjun Yuan

School of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, China

---

∗Corresponding author. Tel.: +86-511-88790318; fax: +86-511-88780201
E-mail address: q.s.chen@hotmail.com (Q.S. Chen)

Abstract:

This work researched on rapid monitoring total polyphenols content during tea extract oxidation using a portable spectroscopy system. Firstly, an in situ monitoring installation for tea extract oxidation was developed, including tea extract oxidation system and spectroscopy system for spectra acquisition. Then, partial least square (PLS) with several variables selection algorithms was used for modeling. Synergy interval partial least square (Si-PLS), genetic algorithm (GA), competitive adaptive reweighted sampling (CARS) and ant colony optimization (ACO) algorithm, were used comparatively for selecting the most effective variables. The performance of the final model was evaluated according to the correlation coefficient (R_p) in prediction set. Experimental results showed that the variables selection methods could significantly decrease the number of variables and improve the model performance, especially for ACO algorithm with least variables. Finally, 28 independent samples were used to test the performance of the spectroscopy system, and the coefficient of variation (CV) of the final results was used to state the stability and reliability of this system. Results also showed that the CVs for most of the samples were less than 10%. This study demonstrated that the tea extract oxidation system combined with spectroscopy system as a promising tool could be used for in situ monitoring tea fermentation.

Keywords: in situ monitoring; total polyphenols; tea extract oxidation; spectroscopy system; variables selection

1. Introduction

Tea plantation is extremely season-specific and climate dependent with different categories all over the world.[1] After spring tea leaves harvest, a mass of summer-autumn tea leaves are remained, which are usually picked for low-end products owing to their bitter flavor, leading to enormous economic losses of tea enterprises. In order to increase the economic benefits of tea industries, it is in urgent need to improve the flavor of summer-autumn tea. Currently, summer-autumn tea leaves are prepared for instant black tea products; firstly, digestion is conducted for original summer-autumn tea leaves, and the extracted solution is then concentrated, finally the concentrated solution is oxidized under flowing oxygen. In this process, oxidation is a critical operation, which is a series of biochemical reactions concentrating on the enzymatic oxidation of tea polyphenols.[2] Thus total polyphenols content is considered to be an important quality property during tea oxidation, which is mainly responsible for characteristic astringent and bitter taste of black tea infusion.[3, 4]

However, it is difficult to determine the oxidation degree. Conventionally, on the one hand, the degree of tea oxidation usually depends on experienced experts, which is entirely subjective, leading to inaccurate and inconsistent results owing to adaptation, fatigue and state of mind; on the other hand, wet chemical analysis is often used for determination of main quality parameters in tea, for instance, colorimetric measurements and titration method with potassium permanganate have been employed to measure total polyphenols content in tea,[5] but the reagents used in this process are harmful to human bodies. Besides, these conventional methods are all time-consuming and labor-intensive,[6] thus cannot feedback tea oxidation information in time, especially in flowing oxygen condition.

On the contrary, near-infrared (NIR) spectroscopy is noninvasive and environmentally friendly, which provides rapid, reproducible results with minimal sample preparation, thus can be employed as an alternative to the conventional methods. During the past decade, substantial researches on the utilization of NIR spectroscopy to determine the tea quality attributes have been performed in research laboratories,[7-9] and the current trends in situ monitoring tea quality are to move the measurements to the whole processing line so as to guarantee high quality and consistency of tea products for consumers. In addition, these researches were mostly conducted with commercial NIR instrument that is expensive, enormous and very sensitive to environmental variation, and thus not suitable for online detection in industrial processing line. Accordingly, it is increasingly demand for developing a portable, cost-effective, and specific spectroscopy system for in situ monitoring tea oxidation. Currently, some advanced spectroscopy systems of small size, low-cost, and ease-of-use, have shown great potential for tea quality assessment.[10, 11] Nevertheless, these advanced spectroscopy systems have not been applied in real-time and on-line monitoring. Therefore, an in situ monitoring installation for tea extract oxidation was developed, including oxidation system and spectroscopy system to monitor total polyphenols content during tea oxidation.

In addition, it is obviously essential to develop a simple and robust calibration for this system. Partial least squares (PLS) is one of the most commonly used multivariate methods,[12, 13] but the classic PLS model based on full spectrum has some irrelevant and collinear spectral variables that may influence the stability and precision of this model. Therefore, variables selection is crucial for model performance improvement and structural simplification,[14] and its result is of great significance for modeling. Synergy interval partial least squares (Si-PLS), genetic algorithm (GA), competitive adaptive reweighted sampling (CARS) and ant colony optimization (ACO) algorithm were employed comparatively in the search for characteristic variables to calibrate models.

The main objective of this work was to study the feasibility of in situ monitoring total polyphenols content during tea oxidation by our developed spectroscopy system. The specific procedures were outlined as follows: (1) an in situ monitoring installation for tea extract oxidation consisting of oxidation system and spectroscopy system was developed; (2) four variables selection algorithms of Si-PLS, GA-PLS, CARS-PLS and ACO-PLS were comparatively used for modeling; (3) the independent samples were used for model test. This system will have a very promising application prospect in tea processing enterprises for tea oxidation monitoring.

2. Materials and methods

2.1. In situ monitoring installation for tea extract oxidation

An in situ monitoring installation for tea extract oxidation, including oxidation system and spectroscopy system, was shown in Fig. 1, which was developed by the Institution of Agro-Product Processing in Jiangsu University. The former oxidation system was mainly composed of four parts: (1) an air pump to supply oxygen; (2) a cycle pipe for gas-lipid mixing; (3) a storage tank; (4) two lipid pumps; also the latter spectroscopy system was mainly composed of four parts: (1) a light source (Halogen & LED, OTO Photonics Inc., Taiwan, China); (2) a transmittance module; (3) a spectrometer (SD1200, OTO Photonics Inc., Taiwan, China); and (4) a computer supported with a data acquisition and control software (SSDAA V1.00, Jiangsu University, China). Two optical fibers (Flight Technology Co., Ltd., Hangzhou, China) were used to connect light source and transmittance module, and also to connect the transmittance module and the spectrometer, respectively. All the connectors were standard SMA 905. The concentrated solution of summer-autumn tea was stored in the tank, and then pumped into the cycle pipe to oxidize with flowing oxygen from the air pump (1.1

L/min of flow rate); the obtained tea infusion then returned to the storage tank under stirring; meanwhile, tea infusion in the storage tank was pumped by a lipid pump (320 ml/min of flow rate) through a quartz cuvette of 1 cm optical path length, and then returned to the storage tank for further oxidation after spectral collection. Other parameters were set as follows: the exposure integration time of the CCD was set to 3 ms; in order to get a stable database, each spectrum was the average of 15 scanning spectra pretreated with Savitzky-Golay smoothing (17 smoothing points); all spectra data were saved as light intensity value in *.txt file format, and the spectrum of each sample was the average of data collection conducted three times. The range of the collected spectra was from 300 to 1000 nm, totally containing 889 variables. While working, the temperature was kept around 25^{o}C and a steady humidity level was maintained in laboratory.

[Here for Fig. 1]

To correct dark current effect of spectrometer and obtain a relative spectrum R, a spectrum B of dark and a spectrum W of water were obtained respectively, and the R was calculated as the Eq. (1):

$$R = \frac{I - B}{W - B} \tag{1}$$

Where I is the original spectrum, B is the dark spectrum recorded by closing completely the transmittance module, and W is the reference spectrum of water.

2.2. Preparation of samples and spectral collection

First of all, digestion is conducted for original summer-autumn tea leaves, and the extracted solution is then concentrated, finally the concentrated solution is oxidized under flowing oxygen. In this work, approximately 1 L summer-autumn concentrated tea was obtained for further oxidation. According to previous experience, the whole process of tea oxidation needs 4 hours approximately. Therefore, one sample was collected every 180 seconds, and then immediately under the preservation of liquid nitrogen for subsequent physiochemical analysis; meanwhile, spectral data was also collected in this process. Finally, 85 tea infusion samples were collected in this work, corresponding to 85 spectra.

Fig. 2(a) presents the raw spectra profile of tea infusion samples. Then standard normal variate (SNV) preprocessing method, as a mathematical transformation method of spectra, was employed to remove slope variation and to correct scatter effects due to small particles from tea leaves dissolved in concentrated solution and differences in the particle sizes.[3, 15]

[Here for Fig. 2]

2.3. Reference analysis

Total polyphenols content in tea infusion was measured by conventional chemical method, referred to the National Standard of China (GB/T 8313-2008) with slight modification. It was estimated by a photometric Folin-Ciocalteu assay, and determined in a 1 cm light-path cuvette by a UV-1601 spectrophotometer (Rayleigh Analytical Instrument, Beijing, China) at 765 nm absorbance of the reaction solution, with its own reagent blank solution for reference. Fig. 3 presents the variation trend of total polyphenols content during tea oxidation, the total polyphenols content first descended quickly, and then descended in a gentle way, eventually keeps a stable level.

[Here for Fig. 3]

In addition, all the samples were divided into two subsets, namely calibration set and prediction set. To avoid bias in subset division, samples in subsets were selected as follows: reference values were arranged in ascending order, and then one prediction sample was chosen from every three samples. Samples in calibration set (57 samples) were used for model calibration, whereas samples in prediction set (28 samples) were used for model verification. Table 1 shows the reference results of total polyphenols content of samples in calibration and prediction sets.

2.4. Multivariate analysis

NIR spectroscopy combined with variables selection methods were used to develop PLS models for predicting total polyphenols content in samples. First, PLS model was calibrated based on full spectrum; then, four commonly used variables selection methods, such as Si-PLS, GA-PLS, CARS-PLS and ACO-PLS, were operated comparatively to eliminate irrelevant variables and extract efficient variables.

Si-PLS algorithm was developed by Norgaard et al., which is an all possible interval combination procedure tests based on all possible PLS of all subsets of intervals.[16] The principle of this algorithm is to split the data set into a series of intervals (variable-wise) and then to calculate all possible PLS model combinations of two, three and four intervals.[17]

GA algorithm described by Leardi and Gonzalez, was based on the principles of genetics and natural selection.[18] It is a mainly stochastic algorithm, and its result depends very much on the randomly generated original population, resulting in enormous difference of the final results from different runs.[19, 20] Therefore, a single GA run is not enough, and it is worthwhile to conduct a series of different runs. The information of variables can be obtained from the frequency with which each variable is selected in the top chromosome of each run.

CARS algorithm first proposed by Hongdong Li et al., employs a simple but effective principle ‘survival of the fittest’ on which Darwin’s Evolution Theory is based, and realizes to some extent the selection of an optimal subset of wavelength.[21, 22] It selects N subsets of variables by N sampling runs in an iterative manner and finally chooses the subset with the lowest RMSECV value as the optimal subset.

ACO algorithm was first proposed as a cooperative learning approach to the traveling salesman problem by Dorigo et al. in the early 1990s.[23, 24] At present, it has been employed by some researchers for spectral variables selection. This algorithm is aim to choose from a large number of available spectral variables, those relevant to the estimation of analyte concentrations or sample properties when spectroscopic analysis is associated with multivariate calibrations. It resembles the behavior of ant colonies in the search for the best path to food sources, which employs the concept of cooperative pheromone accumulation, and optimizes models using a pre-defined number of variables, occupying a Monte Carlo approach to discard irrelevant variables.[25]

The performance of the final model was evaluated according to correlation coefficient of calibration (R_c) and root mean square error (RMSEC) in calibration set, and tested on the basis of correlation coefficient of prediction (R_p) and root mean square error (RMSEP) in prediction set, and the detail descriptions about them can be referred.[26-27] In addition, in order to develop a concise model with high precision for this in situ monitoring installation, the number of variables should be as low as possible in order to reduce the complexity of the developed model and improve the detection efficiency of this system. All data processing and analysis were performed in Matlab Version 7.11.0 (Mathworks, Natick, USA) under Windows 7.

3. Results and discussion

As can be seen from Fig. 1, an in situ monitoring installation for tea extract oxidation was developed to simulate industrial production. While working, the concentrated solution of summer-autumn tea stored in tank was pumped into the cycle pipe to oxidize with flowing oxygen from the air pump. The obtained tea infusion then returned to the storage tank under stirring. Meanwhile, tea infusion in the storage tank was pumped by a lipid pump through a quartz cuvette of 1 cm optical path length, and then returned to the storage tank for further oxidation after spectral collection. Fig. 2(a) presents the original spectra obtained in this process, and then SNV preprocessing method was performed to remove slope variation and to correct scatter effects due to small particles from tea leaves dissolved in concentrated solution and differences in the particle sizes, as shown in Fig. 2(b). Each spectrum is corrected individually by first centering the spectral values, and then the centered spectrum is scaled by the standard deviation calculated from individual spectral values. Thereafter, multivariate analysis methods were applied in this work, and results were described as follows.

3.1. The result of Si-PLS model

In this work, the full spectrum (300-1000 nm) was divided into 11, 12, … , 25 intervals, and each was performed with PLS combined with two, three and four subintervals. The optimum result was obtained when full spectrum was split into 17 intervals and the combined intervals were the 5rd, 9th, 14th and 15th interval, thus totally containing 212 spectral variables corresponding to 440.976 – 478.431, 597.815 – 638.529, and 814.239 – 905.448 nm in the full-spectrum region, which is shown in Fig. 4(a).

3.2. The result of GA-PLS model

In this research, GA runs ten times, and the best running result with the lowest RMSECV value was selected to build the final PLS model; other parameters for running GA were set as follows based on experience: the initial population was set to 80; the crossover probability was set to 0.5; the mutation probability was set to 0.01; and the number of genetic iterations was set to 50. Fig. 4(b) shows the histogram of frequency selected for the variables in the final model with a dashed line indicating that frequencies of variables that were not lower than 3 could be selected, and the best result of GA-PLS model was obtained when 145 variables were involved.

3.3. The result of CARS-PLS model

Fig. 4(c) shows the variation trend of the number of sampled variables (plot 1), 8-fold RMSECV values (plot 2) and the regression coefficient path of each variable (plot 3) with the increasing of sampling runs from one CARS running. Seen from this figure, the number of sampled variables decreases fast at the first stage of exponentially decreasing function, and then very slowly at the second stage of exponentially decreasing function, demonstrating that the proposed two phase selection, fast selection and refined selection, are indeed implemented in CARS. The RMSECV values first descend quickly from sampling runs 1 to 10, which may be ascribed to the elimination of redundant variables; then descend in a gentle way from sampling runs 20 to 30, corresponding to the phase that the sampled variables reduce slowly; and finally increase fast owing to the loss of information induced by eliminating some key variables from the optimal subset (denoted by asterisk). Therefore, the best result of CARS-PLS model was obtained when 65 variables were involved.

3.4. The result of ACO-PLS model

In this work, the required parameters for running ACO algorithm were set as follows according to experience based on substantial trials: the initial population was set to 80; the maximum number of iterations was set to 50; the maximum number of cycles was set to 20; the probability threshold of variable selection was set to 0.3; and the pheromone attenuation coefficient was set to 0.65. Fig. 4(d) presents the spectral variables selected by ACO algorithm, totally containing 20 variables. As seen from this figure, some of the selected variables assemble together, and the clustering tendency is similar to the intervals previously selected by Si-PLS and GA but not the same.

[Here for Fig. 4]

3.5. Discussion of the results

In this work, we systemically studied the multivariate calibrations in order to seek an optimum model for the in situ monitoring installation of tea extract oxidation. In order to highlight the superiority of variables selection methods, four different methods of spectral variables selection such as Si-PLS, GA-PLS, CARS-PLS and ACO-PLS were attempted in contrast to the classical PLS model, and to seek the most efficient variables, these variables selection methods were also compared with each other. The overall results can be seen from Table 2, and each model only presents its optimum result. Fig. 5 are the scatter plots showing high correlation between reference measured and NIR predicted values both in calibration and prediction sets, wherein (a) for Si-PLS, (b) for GA-PLS, (c) for CARS-PLS, and (d) for ACO-PLS. Investigated from Table 2 and Fig. 5, the results show that it is feasible to measure total polyphenols content during tea oxidation using our developed in situ monitoring installation; and then, all the variables selection methods have improved model performance in contrast to the classical PLS model, which indicated that variables selection methods can not only reduce the number of variables but also improve the performance of model, especially for the approach of ACO-PLS with least variables. Main reasons could be summarized as follows.

The commonly used PLS algorithm was implemented to calibrate model based on full spectrum including 889 variables in this work. However, there are many collinear variables or irrelevant variables that are unrelated to tea polyphenols compositions, namely unwanted variables. If too many unwanted variables were included in PLS model, the performance of PLS model would be inevitably weakened. Si-PLS divides the full spectrum into a series of equidistant intervals and searches for all-possible-interval-combinations to select the optimum model, so it can remove unwanted spectral information and retain some interesting information as much as possible for modeling. The optimal spectra region selected for tea polyphenols was corresponding to 440.976 – 478.431, 597.815 – 638.529, and 814.239 – 905.448 nm. Tea polyphenols is a complex of over 30 polyphenolic compounds, generally including flavanols, anthocyanins, flavonoids, flavonols and phenolic acids,[28] involving lots of Ar-OH (phenolic hydroxyl group) groups and other groups like Ar-CH, CH_3, and CH_2. Thus the variables related to tea polyphenols may be discontinuous and disperse over the whole spectral region, indicating that there were still collinear variables in the subintervals, or some other relevant variables may be omitted in Si-PLS model. Thus other variables selection methods such as GA, CARS and ACO were attempted and all of them achieved good performances. However, GA and CARS searches variables among the whole spectral region in order to avoid local minima, leading to a potential problem: a tendency to include irrelevant variables in the final solutions together with those which are relevant to tea polyphenols.[21, 25] This may be the reason that the variables selected by GA and CARS were much more than that by ACO. Meanwhile, the larger number of ants can decrease the probability of convergence of ACO to the local optimum. As a result, ACO was adopted as the optimal variables selection method, which greatly improved the model precision with least variables. Once the model was developed, the computation time for each prediction model is less than 50 ms, which totally meets the demand for in situ monitoring tea oxidation in industrial usage.

[Here for Fig. 5]

In addition, in order to verify the stability and reliability of this developed VIS-NIR spectroscopy system, the three times spectra of one sample were respectively taken into the optimal model. Generally, coefficient of variation (CV) is defined as the "relative standard deviation" and a concept that can directly compare uncertainties of different variables given in different dimensions. It is defined as CV = SD/MV, where SD is the standard deviation and MV is the mean value.[29] Thus, CV of the three times NIR predicted values was employed to test the stability and reliability of this system. Fig. 6 shows CV of the three times NIR predicted values, we could see that the very low CV values were obtained for all the three times NIR predicted values, and most of them were less than 10%. CV results indicated that the measurements by this portable spectroscopy system have good stability and reliability in our work.

[Here for Fig. 6]

4. Conclusions

In this work, an in situ monitoring installation for tea extract oxidation including oxidation system and spectroscopy system was developed. Combined with variables selection methods, it was successfully used for monitoring total polyphenols content during tea oxidation. Results showed that the ACO-PLS algorithm was extremely suitable for this system. This research provides an important tool for in situ monitoring tea oxidation.

Acknowledgements

This work has been financially supported by the National Natural Science Foundation of China (31271875), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). We are also grateful to many of our colleagues for stimulating discussion in this field.

References

1 N. Bhattacharyya, S. Seth, B. Tudu, P. Tamuly, A. Jana, D. Ghosh, R. Bandyopadhyay and M. Bhuyan, *J. Food Eng.*, 2007, **80**, 1146-1156.
2 V. Cheynier, *Am. J. Clin. Nutr.*, 2005, **81**, 223S-229S.
3 Q. S. Chen, J. W. Zhao, M. H. Liu, J. R. Cai and J. H. Liu, *J. Pharmaceut. Biomed.*, 2008, **46**, 568-573.
4 I. Lesschaeve and A. C. Noble, *Am. J. Clin. Nutr.*, 2005, **81**, 330S-335S.
5 Q. S. Chen, J. W. Zhao, X. Y. Huang, H. D. Zhang and M. H. Liu, *Microchem. J.*, 2006, **83**, 42-47.
6 H. Jiang, Q. Chen and G. Liu, *Process Biochemistry*, 2014, **49**, 583-588.
7 Q. Chen, J. Zhao, S. Chaitep and Z. Guo, *Food Chemistry*, 2009, **113**, 1272-1277.
8 M.-S. Lee, Y.-S. Hwang, J. Lee and M.-G. Choung, *Food Chemistry*, 2014, **158**, 351-357.
9 M. N. Hall, A. Robertson and C. N. G. Scotter, *Food Chemistry*, 1988, **27**, 61-75.
10 C. Feng, X. Liu, C. S. Yan, S. L. He and L. Mei, *SPIE OPTO*, 2013, 864122-864122-6.
11 X. Liu, C. Feng and C. S. Yan, *WOCC*, 2013, 607-609.
12 W. Luo, J. Wu, X. K. Wang, X. Lin and H. Li, *Anal. Methods*, 2013, **5**, 1337-1345.
13 S. Mahramyari, E. Pourbasheer, A. Banaei, M. R. Ganjali and P. Norouzi, *RSC Advances*, 2014, **4**, 41039-41044.
14 X. B. Zou, J. W. Zhao, M. J. W. Povey, M. Holmes and H. P. Mao, *Analytica Chimica Acta*, 2010, **667**, 14-32.
15 D. Wu, Y. He and S. J. Feng, *Anal. Chim. Acta*, 2008, **610**, 232-242.
16 D. Wu, Y. He, P. Nie, F. Cao and Y. Bao, *Analytica Chimica Acta*, 2010, **659**, 229-237.
17 L. Huang, J. W. Zhao, Q. S. Chen and Y. H. Zhang, *Food Res. Int.*, 2013, **54**, 821-828.
18 E. Franco-Lara, H. Link and D. Weuster-Botz, *Process Biochemistry*, 2006, **41**, 2200-2206.
19 R. Leardi and A. Lupiáñez González, *Chemometrics and Intelligent Laboratory Systems*, 1998, **41**, 195-207.
20 J. Zhao, Q. Ouyang, Q. Chen and H. Lin, *Food Science and Technology International*, 2013, **19**, 305-314.
21 H. Li, Y. Liang, Q. Xu and D. Cao, *Analytica chimica acta*, 2009, **648**, 77-84.
22 Q. Luo, Y. Yun, W. Fan, J. Huang, L. Zhang, B. Deng and H. Lu, *RSC Advances*, 2015, **5**, 5046-5052.
23 M. Dorigo and L. M. Gambardella, *Evolutionary Computation, IEEE Transactions on*, 1997, **1**, 53-66.
24 K. R. Rao, T. Srinivasan and C. Venkateswarlu, *Process Biochemistry*, 2010, **45**, 961-972.
25 F. Allegrini and A. C. Olivieri, *Analytica chimica acta*, 2011, **699**, 18-25.
26 Q. S. Chen, P. Jiang and J. W. Zhao, *Spectrochim. Acta A*, 2010, **76**, 50-55.
27 Y. N. Shao, Y. D. Bao and Y. He, *Food bioprocess technol.*, 2011, **4**, 1376-1383.
28 J. H. Huang, K. L. Huang, S. Q. Liu, Q. Luo and M. C. Xu, *J. Colloid Interf. Sci.*, 2007, **315**, 407-414.
29 L. Håkanson, J. M. Malmaeus, U. Bodemer and V. Gerhardt, *Ecological Modelling*, 2003, **169**, 179-196.

Figures Captions

Figure 1 Schematic diagram for in situ monitoring tea oxidation.

Figure 2 Raw spectra (a) and SNV preprocessed spectra (b) of tea infusion samples.

Figure 3 Variation trend of total polyphenols content during tea oxidation.

Figure 4 Spectral variables selected by Si-PLS (a), GA (b), CARS (c), and ACO (d).

Figure 5 Scatter plots of reference values versus NIR predicted values in calibration and prediction sets for Si-PLS (a), GA-PLS (b), CARS-PLS (c), and ACO-PLS (d).

Figure 6 CV (%) value of three times NIR predicted values for total polyphenols content.

Table 1 Reference results of total polyphenols content of samples in calibration and prediction sets.

Subsets	Unit	S.N.[a]	Range	Mean	S.D.[b]
Calibration set	%	57	6.73~72.58	22.95	12.85
Prediction set	%	28	7.91~53.40	22.12	11.12

[a] S.N.: sample number.

[b] S.D.: standard deviation.

Table 2 Results of different regression models for predicting total polyphenols content in tea infusion.

Methods	Variables	Calibration set		Prediction set	
		R_c	*RMSEC*	R_p	*RMSEP*
PLS	889	8	0.8358	0.07	0.7659
Si-PLS	212	10	0.9751	0.0283	0.8766
GA-PLS	145	7	0.9651	0.0334	0.8993
CARS-PLS	65	9	0.9751	0.0283	0.8897
ACO-PLS	20	9	0.9148	0.0515	0.8853

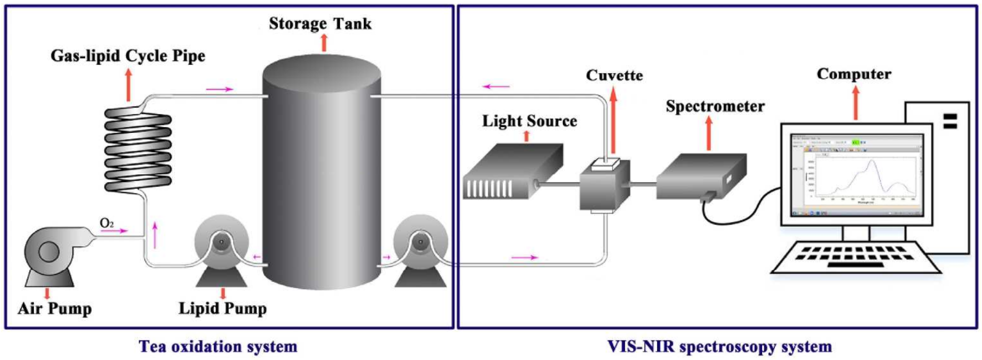


Figure 1 Schematic diagram for in situ monitoring tea oxidation.
316x120mm (300 x 300 DPI)

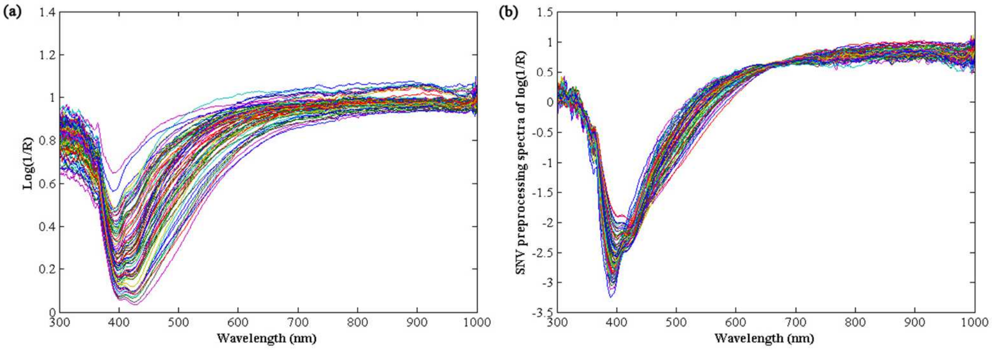


Figure 2 Raw spectra (a) and SNV preprocessed spectra (b) of tea infusion samples.
333x123mm (96 x 96 DPI)

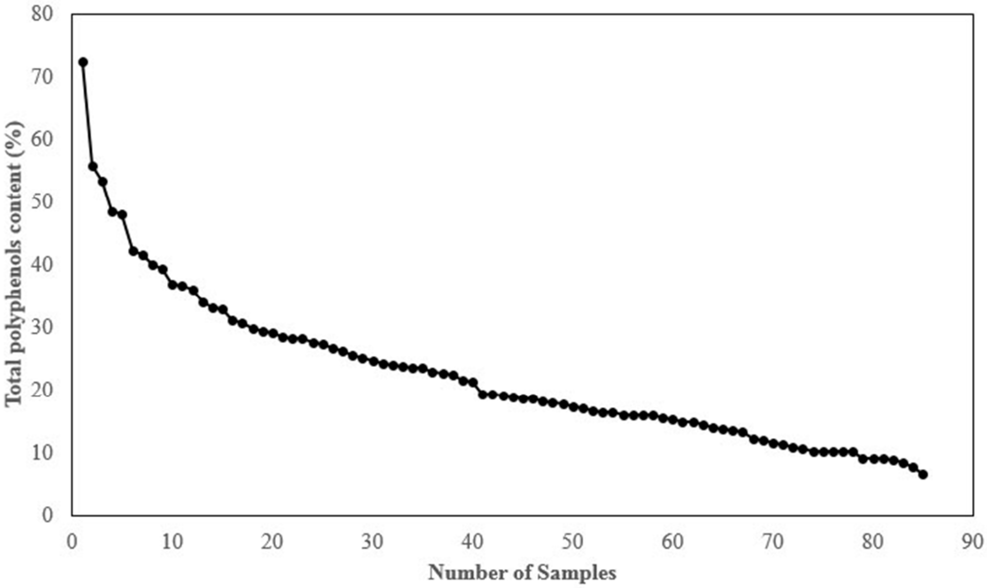


Figure 3 Variation trend of total polyphenols content during tea oxidation.
60x35mm (300 x 300 DPI)

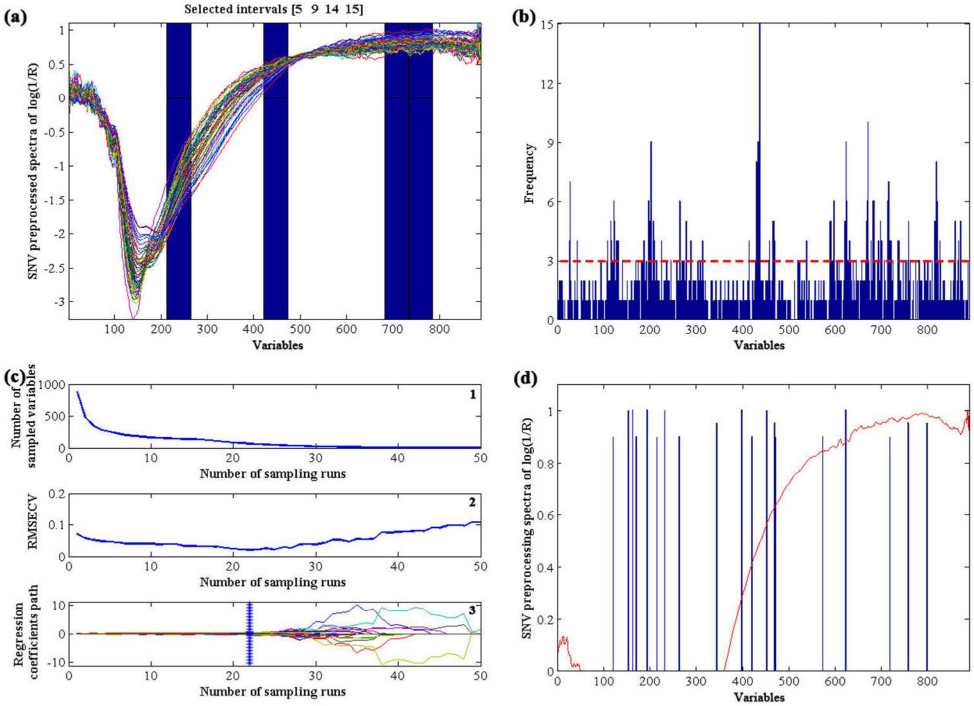


Figure 4 Spectral variables selected by Si-PLS (a), GA (b), CARS (c), and ACO (d).
108x78mm (300 x 300 DPI)

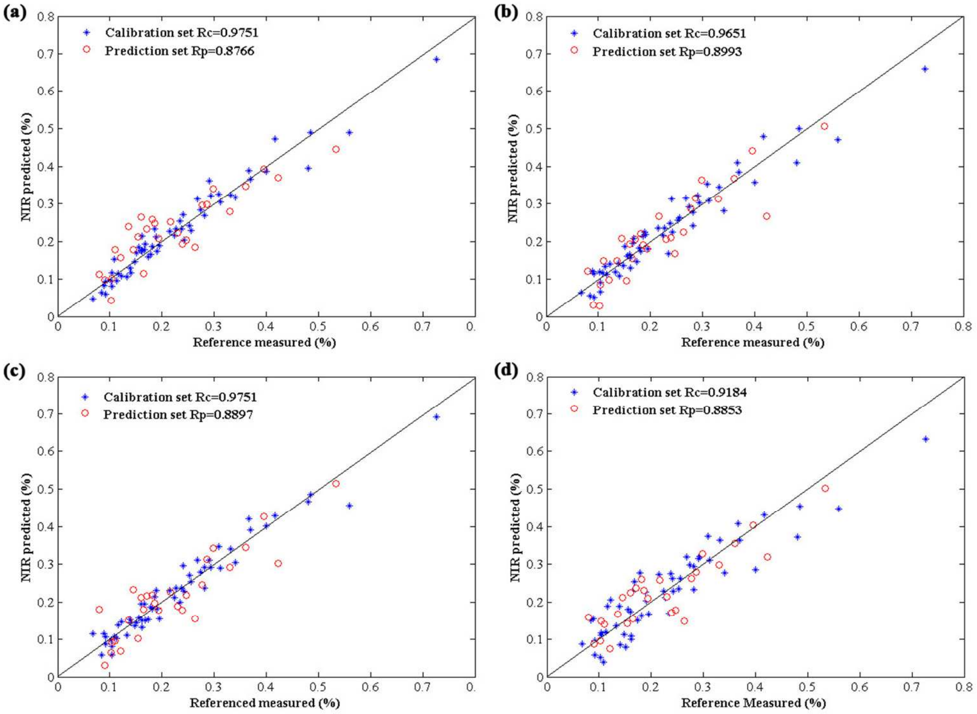


Figure 5 Scatter plots of reference values versus NIR predicted values in calibration and prediction sets for Si-PLS (a), GA-PLS (b), CARS-PLS (c), and ACO-PLS (d).
107x78mm (300 x 300 DPI)

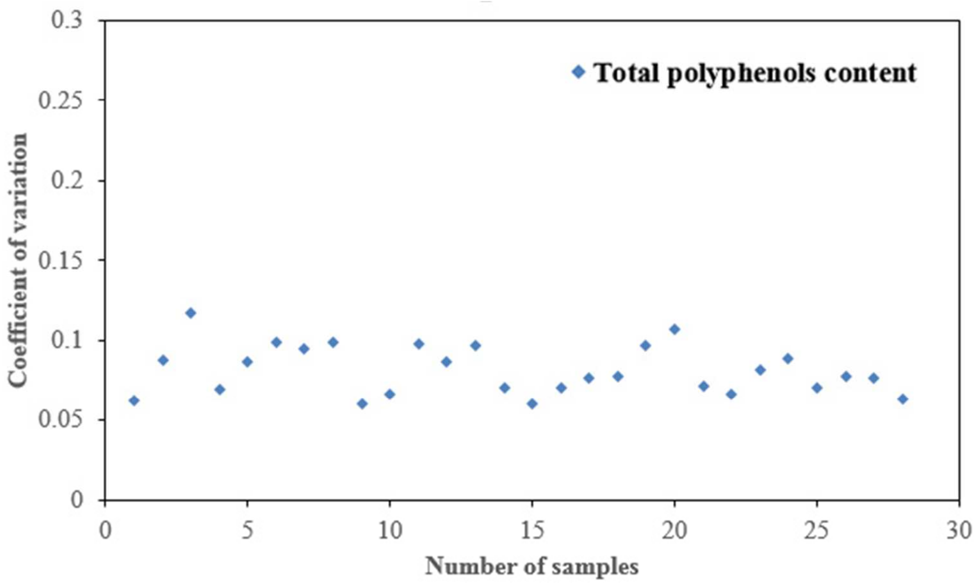


Figure 6 CV (%) value of three times NIR predicted values for total polyphenols content.
52x31mm (300 x 300 DPI)