Food Chemistry 224 (2017) 1–10

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

# Comparison of different CCD detectors and chemometrics for predicting total anthocyanin content and antioxidant activity of mulberry fruit using visible and near infrared hyperspectral imaging technique

Lingxia Huang [a,b,1], Yibin Zhou [a,1], Liuwei Meng [a], Di Wu [c,*], Yong He [d]

[a] College of Animal Sciences, Zhejiang University, Hangzhou 310058, PR China
[b] South Taihu Agricultural Technology Extension Center in Huzhou, Zhejiang University, Huzhou 313000, PR China
[c] College of Agriculture & Biotechnology/Zhejiang Provincial Key Laboratory of Horticultural Plant Integrative Biology/The State Agriculture Ministry Laboratory of Horticultural Plant Growth, Development and Quality Improvement, Zhejiang University, Zijingang Campus, Hangzhou 310058, PR China
[d] College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, PR China

ARTICLE INFO

*Article history:*
Received 4 August 2016
Received in revised form 12 December 2016
Accepted 13 December 2016
Available online 18 December 2016

*Keywords:*
Hyperspectral imaging
Mulberry
Total anthocyanins
Antioxidant activity
CCD detector
Visible and near infrared spectroscopy
Chemometrics

ABSTRACT

This study investigated the potential of using hyperspectral imaging technique in tandem with chemometrics for rapid and invasive predicting total anthocyanin content and antioxidant activity of mulberry fruit. Two calibration methods of partial least square regression and least-squares support vector machines and three wavelength selection algorithms of successive projections algorithm, uninformation variable elimination, and competitive adaptive reweighted sampling were applied. The best prediction models for the analysis of total anthocyanin content and antioxidant activity had $R^2_{val}$ of 0.959 and 0.995 respectively. The performances of two CCD detectors named silicon (Si) and indium gallium arsenide (InGaAs) were compared. The results show that hyperspectral imaging has a great potential for the assessment of total anthocyanin content and antioxidant activity of mulberry fruit.



## 1. Introduction

Mulberry fruit comes from the mulberry tree (*Morus* sp.), which grows in a wide range of climatic, topographical and soil conditions (Ercisli & Orhan, 2007). Besides as a by-product of sericulture, mulberry fruit with its sweet flavor is popularly not only consumed fresh but also made in jam, pies, tarts, wines, and liquor. Research found that mulberry fruit contains many bioactive components including flavonoids, anthocyanins, and carotenoids (Du, Zheng, & Xu, 2008; Jiang & Nie, 2015), and its extracts possess a wide scope of biochemical activities such as scavenging free radicals, anti-cancer effects, anti-atherogenic properties, anti-hyperlipidemia, and neuroprotective effects (Bae & Suh, 2007; Chen et al., 2005, 2006). Therefore, mulberry fruit is commonly considered as a high grade fruit in the market, which is more expensive than common fruits.

Nowadays, consumers are paying more attention to fruits with high quality. Besides the taste, customers also care about the health benefits of fruits, especially bioactive compounds and their bioavailability. Guaranteeing fruits with high quality is an important issue, which is directly related to people's health. There is a need for reliable techniques to evaluate the quality of mulberry fruit. The product additional value of mulberry fruit can be increased by grading the fruits according to their anthocyanin content and antioxidant activity.

Commonly, chemical analysis methods are used to predict the anthocyanin content in mulberry fruits, by extracting and separating anthocyanins from mulberry fruits tissues (Jiang & Nie, 2015; Oki et al., 2006). On the other hand, the main antioxidant activity measurement procedures include ferric reducing/antioxidant power (FRAP), 2,20-azino-bis (3-ethylbenz-thiazoline-6-sulfonic acid) (ABTS), Trolox equivalent antioxidant capacity (TEAC), 2,2-diphenyl-1-picrylhydrazyl (DPPH), and oxygen radical absorbance capacity (ORAC) (Saura-Calixto et al., 2008; Thaipong, Boonprakob, Crosby, Cisneros-Zevallos, & Byrne, 2006). However, these methods are time-consuming, laborious and destructive, resulting in a lack of capable for rapid or even on-line detection.

* Corresponding author.
*E-mail addresses:* lxhuang@zju.edu.cn (L. Huang), zhou_yi_bin@qq.com (Y. Zhou), 1193414839@qq.com (L. Meng), di_wu@zju.edu.cn, china.di.wu@gmail.com (D. Wu), yhe@zju.edu.cn (Y. He).
[1] Lingxia Huang and Yibin Zhou contributed equally to this work.

http://dx.doi.org/10.1016/j.foodchem.2016.12.037
0308-8146/

Moreover, the above methods are selective examinations, which can only measure a few samples. The mulberry industry requires rapid and non-invasive methods to predict anthocyanin content and antioxidant activity of massive mulberry fruits.

Visible and near infrared (Vis–NIR) spectroscopy is a rapid and non-invasive technique permitting an online quality measurement of fruit products (Nicolaï et al., 2007). It provides complex structural information related to the vibration behavior of combinations of bonds, such as O—H, C—H, C—O and N—H (Cen & He, 2007; Zou, Zhao, Povey, Holmes, & Mao, 2010). The information can be associated with the component content and quality of fruits. There are successful investigations using spectroscopic techniques to predict anthocyanin content and antioxidant activity of fruits, such as apple (Giovanelli, Sinelli, Beghi, Guidetti, & Casiraghi, 2014), cherries (Zude, Pflanz, Spinelli, Dosche, & Torricelli, 2011), and blueberries (Sinelli, Spinardi, Egidio, Mignani, & Casiraghi, 2008). However, to the best of our knowledge, the anthocyanin content and antioxidant activity of mulberry fruit has not been investigated.

Moreover, because the shape of mulberry fruit is not round and its size is small, when a small circle area of mulberry fruit is measured at once by setting a relatively close distance between the spectroscopic probe and the fruit surface to avoid the measured area covers background, the spectral acquisition needs to be repeated at different positions to cover the whole fruit surface. On the other hand, if only a relatively large distance is set from the probe to fruit surface, the measured area would cover the background. Therefore, the spectroscopic measurement cannot accurately get the spectral information of the whole fruit due to its inherent deficiency of containing no spatial information.

Hyperspectral imaging also called imaging spectroscopy or imaging spectrometry, integrates the main advantages of spectroscopy and imaging into one system, so that the spectral and spatial information can be acquired simultaneously (Wu & Sun, 2013a). Different from the spectroscopic methods that only acquire spectral information of one or more points on sample, hyperspectral imaging can acquire spectral information of every pixel on sample image. There are many studies using hyperspectral image to predict quality of fruits, such as apple, citrus, pear, kiwifruit, peach and grape (Wu & Sun, 2013b). Hernandez-Hierro, Nogales-Bueno, Rodriguez-Pulido, and Heredia (2013) studied the feasibility on the use of near-infrared hyperspectral imaging for the screening of anthocyanins in intact grapes during ripening. Nogales-Bueno, Baca-Bocanegra, Rodriguez-Pulido, Heredia, and Hernandez-Hierro (2015) used near infrared hyperspectral tools to screen extractable polyphenols in red grape skins. Chen, Zhang, et al. (2015) predicted the anthocyanin content of wine grapes by NIR hyperspectral imaging. Fernandes et al. (2015) studied the anthocyanin content determination in whole Port wine grape berries by hyperspectral imaging. However, besides grape, only a few studies evaluated the feasibility of predicting anthocyanin content or antioxidant activity of other fruits using hyperspectral imaging technique. Liu, Wei, Wei, Yang, and Lei (2015) used multispectral imaging to predict the content of bioactive compounds in intact tomato fruit. Yang, Sun, Pu, Wang, and Zhu (2015) detected the anthocyanin content in lychee pericarp during storage using hyperspectral imaging. To the best of our knowledge, there are few researches working on the anthocyanin content and antioxidant activity prediction of mulberry fruit using Vis–NIR spectroscopy or hyperspectral imaging techniques.

The CCD detectors of hyperspectral imaging systems have the function of quantifying the intensity of the acquired light by converting incident photons into electrons. Photodiodes made of light sensitive materials are the basic unit of solid state area detectors to convert radiation energy to electrical signals. Silicon (Si) and indium gallium arsenide (InGaAs) are two commonly used materials for hyperspectral imaging detectors, which work for acquiring the spectra at around 400–1000 nm and 900–1700 nm, respectively. Therefore, it is of our interest to investigate and compare the predictive capabilities of the two detectors for predicting anthocyanin content and antioxidant activity of mulberry fruit.

The objective of this study was to investigate the potential of Vis–NIR hyperspectral imaging for the rapid and non-invasive prediction of total anthocyanin content and antioxidant activity of mulberry fruit. The successful outcome of the study is very advantageous to consistently assure buyers and consumers of mulberry fruit with high quality. The specific objectives of the current work were to (1) acquire Vis-NIR hyperspectral images of mulberry fruit; (2) extract spectra of mulberry fruit based on the spatial information within hyperspectral images; (3) select the important wavelengths that were most correlated to anthocyanin content and antioxidant activity; (4) build predictive models for the anthocyanin content and antioxidant activity by two calibration methods, respectively; and (5) compare the prediction abilities of two imaging detectors of Si and InGaAs.

## 2. Materials and methods

### 2.1. Sample preparation

Experiments were carried out in 2015. Fresh "Dashi" mulberry fruits were provided by Huzhou Academy of Agricultural Sciences, China. Fruits were harvested at three ripening stages at April 23, April 30 and May 5, respectively, and transported to the laboratory on the day of harvest. Those of uniform commercial maturity with absence of disease and mechanical wounding were selected. There were sixty fruits of each harvest day. As a result, there were a total of 180 mulberry fruits analyzed in this work.

### 2.2. Hyperspectral imaging system and image acquisition

Hyperspectral images of mulberry fruits were acquired by two hyperspectral imaging systems. System I had a Si detector acquiring the spectral information in visible and short-wave near infrared region (380–1030 nm), while an InGaAs detector was applied in System II to measure the long-wave near infrared spectra (874–1734 nm). System I had an imaging spectrograph (ImSpectorV10E, Spectral Imaging Ltd., Oulu, Finland), a high-resolution $1344 \times 1024$ 6.45-μm single-piece digital camera (C8484-05; Hamamatsu, Hamamatsu City, Japan), a camera lens 122 (100195; Schneider Kreuznach, Bad Kreuznach, Germany), and an operation software (V10E, Isuzu 123 Optics Corp., Taiwan, China). System II was composed of an imaging spectrograph (ImSpector N17E, Spectral Imaging Ltd., Oulu, Finland), a high performance $320 \times 256$ camera (Xeva 992; Xenics Infrared Solutions, Leuven, Belgium), a camera lens (OLES15; Specim, Spectral Imaging Ltd., Oulu, Finland), and an operation software (N17E, Isuzu Optics Corp., Taiwan, China). The two systems shared a specially assembled light unit consisting of two 150-W quartz tungsten halogen lamps as the light source (2900ER; Illumination Technologies, Inc., New York, USA) and a conveyer table operated by a stepper motor (ST-1212-300; Tanlian Company, Taiwan, China). Each fruit sample was placed on the conveyer table, and first measured by the system I (Si detector) and then measured by the system II (InGaAs detector). The measured hyperspectral images were three-dimensional, which had a two-dimensional image ($x$, $y$) and one-dimensional spectra ($\lambda$). Hyperspectral image examples measured by both System I (Si detector) and System II (InGaAs detector) for mulberry fruits at different stages with different ranges of the total anthocyanin content and antioxidant activity are shown in Fig. 1.

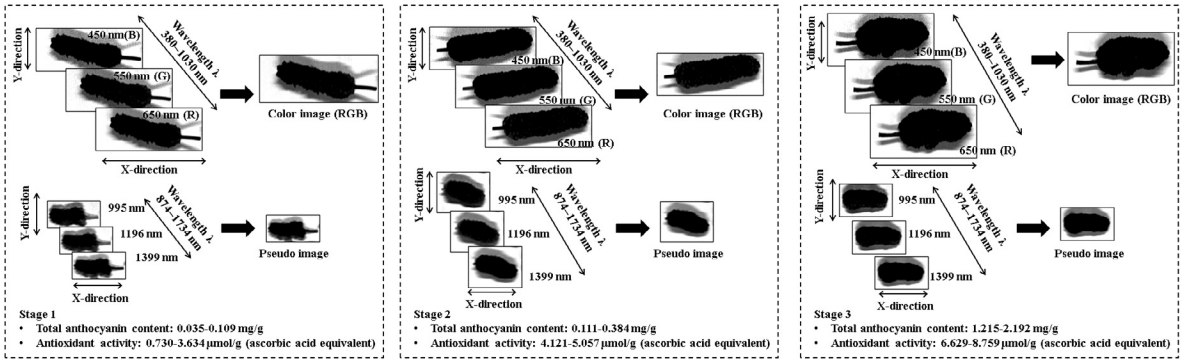


**Fig. 1.** Hyperspectral image examples measured by both System I (Si detector) and System II (InGaAs detector) for mulberry fruits at different stages with different ranges of the total anthocyanin content and antioxidant activity.

### *2.3. Reference methods for measurement of total anthocyanin content and antioxidant activity*

Mulberry fruits were froze by liquid nitrogen and grinded into powder. Due to the small weight of single mulberry fruit, five fruits were considered as one sample and their total anthocyanin content and antioxidant activity were measured together and set as the reference values of this sample. The DPPH assay was carried out by the procedure described by (Thaipong et al., 2006) with some modification. For each sample, 0.5 g of powders were mixed with 5 mL methanol. The homogenates were kept at 4 °C for 24 h and then centrifuged at 4000 rpm for 20 min to get supernatant. 0.0059g DPPH were dissolved with 250 ml methanol to get DPPH methanol solution (60 μmol/ml) and 3.9 mL of DPPH (60 μmol/L) was added to 0.1 mL of appropriately diluted extract. After reacting 1 h in the dark at room temperature (25 °C), the absorbance was recorded at 517 nm. The standard curve was linear between 0 and 0.5 μmol/ml ascorbic acid. Results are expressed in μmol ascorbic acid/g fresh mulberry fruits mass.

For the measurement of total anthocyanin content, 0.5 g mulberry powder of each sample was added to 5 mL 85% methanol (containing 0.5% acetic acid), extracting at 4 °C keep in dark for 24 h. Then, the extracts were centrifuged (4000 rpm, 20 min) to get supernatant. The pH differential method with some modification was used for measuring total anthocyanin content (Lee, Durst, & Wrolstad, 2005). 1 ml extracting solution was mixed with 9 ml 0.025 mol/L potassium chloride buffer (pH 1.0), and the other 1 ml extracting solution was mixed with 0.4 mol/L sodium acetate buffer (pH 4.5). After 1.5 h of stabilization at room temperature (25 °C), absorbance at wavelengths of 510 and 700 nm were measured for both solutions respectively. The total anthocyanin content was expressed as cyanidin-3-glucoside equivalents by the following Eq. (1):

Total anthocyanins content

$$(\mathrm{mg}\cdot\mathrm{g}^{-1}) = \frac{A}{\varepsilon \cdot L} \times MW \times DF \times \frac{V}{W} \times 10^3 \quad (1)$$

$A = (A_{510\mathrm{nm\ pH\ 1.0}} - A700_{\mathrm{nm\ pH\ 1.0}}) - (A510_{\mathrm{nm\ pH\ 4.5}} - A700_{\mathrm{nm\ pH\ 4.5}})$
$\varepsilon$ = 26,900 $L\cdot mol^{-1}\cdot cm^{-1}$ (extinction coefficient for cyanidin-3-glucoside at 510 nm)
L = 1 cm (path length)
MW = 449.2 $g\cdot mol^{-1}$ (molecular weight for cyanidin-3-glucoside)
DF = 10 (dilution factor of solution)
V = 0.008 L (volume of extracting solution)
W = weight of powder in gram
$10^3$: factor for conversion from $g\cdot g^{-1}$ to $mg\cdot g^{-1}$

### *2.4. Pixel identification, and spectral data extraction*

Both Si and InGaAs CCD detectors in two hyperspectral imaging systems measured the raw data with signal intensity rather than spectral reflectance. It is necessary to correct the raw hyperspectral images into reflectance mode before spectral extraction. The correction procedure is descripted in the literature (Wu et al., 2012). On the other hand, the accurate extraction of spectral data from mulberry fruit rather than background is important to establish calibration models related to corresponding quality attributes. The spectra of mulberry fruits were extracted from their hyperspectral images using ENVI v4.6 software (Research Systems Inc., Boulder, CO, USA). In details, the pixel selection was carried out manually, and the reflectance spectra of all pixels belonging to the fruit within the corresponding image were extracted using the region of interests (ROI) function of ENVI v4.6 software.

### *2.5. Chemometric modeling*

The use of chemometrics in modeling spectral data is widely employed, being considered as a standard procedure for establishment of predictive models in the analysis of hyperspectral images. In this study, partial least square regression (PLSR) and least-squares support vector machines (LS-SVM) were respectively used to establish calibration models between the spectral data of samples and their reference total anthocyanin content and antioxidant activity values. PLSR is a classic linear regression method that has been commonly used for spectral analysis (Huang, Ruan, Chen, Lin, & Liu, 2014; Wu, Sun, & He, 2014). It optimizes the covariance between quality data and linear combinations of spectral data by performing the decomposition on both the spectral and quality data simultaneously. To achieve that, PLSR projects the spectral data onto a set of orthogonal factors called latent variables (LVs), and explores the optimal function by minimizing the error of sum squares (finding the optimal LVs), which is typically done by cross-validation. LS-SVM is commonly used to establish non-linear calibration models (Chen et al., 2013; Huang, Liu, Zhang, & Wu, 2015). Instead of solving a convex quadratic programming problem like in classical SVM, LS-SVM finds the solution by solving a set of linear equations. Radial basis function (RBF) was used in this study because it is a non-linear function and can obtain good performance under general smoothness assumptions and reduce the computational complexity of the training procedure compared with other kernels. The optimal values of two main parameters in the LS-SVM model, namely the regularization parameter $\gamma$ and the RBF kernel function parameter $\sigma^2$, were determined using the grid-search technique with leave one out cross-validation. It was

noticed that, in this work, hyperspectral images of three spectral sets were considered for the model calibration, respectively, and their results were compared. Specific, set I was measured by the Si detector in System I, set II was measured by the InGaAs detector in System II, set III was the combination of sets I and II.

### 2.6. Important wavelength selection

Spectral wavelengths in hyperspectral images are characterized by their large degree of dimensionality with collinearity and redundancy. Only the most important wavelengths (variables) having the great influence in prediction should be kept in the model. In this work, three variable selection techniques, namely successive projections algorithm (SPA), uninformative variable elimination (UVE), and competitive adaptive reweighted sampling (CARS) were considered for wavelength selection in four strategies, namely SPA, UVE, UVE-SPA and CARS. The SPA, UVE, and CARS are all typical variable selection techniques for spectral analysis (Wu & Sun, 2013c; Zhu, Nie, Wu, He, & Chen, 2016; Zou et al., 2010). SPA selects variables with minimally redundancy to solve the collinearity problems (Araujo et al., 2001). UVE selects informative variables according to their stability calculated from PLSR regression coefficient (Centner et al., 1996). In the calculation of CARS, wavelengths with larger absolute regression coefficients of PLSR model are considered as good candidates and selected based on the principle of 'survival of the fittest' from Darwin's Evolution Theory (Li, Liang, Xu, & Cao, 2009). In addition, besides the calculation of SPA based on the full wavelength range, SPA is also commonly carried out after UVE calculation to select variables that informative but no collinearity. Therefore, in this work, three variable selection techniques of SPA, UVE, and CARS were considered for the wavelength selection in four strategies, namely SPA, UVE, UVE-SPA and CARS.

### 2.7. Model evaluation

To evaluate the predictive abilities of the calibrated models, both leave-one-out cross validation and independent test were used for the validation purpose. Leave-one-out cross validation was carried out based on the samples used for model establishment. During the calculation, only one sample is preserved at a time and all other samples are used for calibration. For the independent testing, all the samples of each stage were randomly divided into three groups, and then the samples of two groups were used for calibration and the samples of the left group for prediction, so that the samples for prediction were independent to the samples for calibration. All three groups were used for prediction respectively, resulting in three different calibration/prediction sets of A, B and C.

The performances of models were evaluated in terms of the correlation coefficients of calibration ($r_{cal}$), coefficients of determination of calibration ($R^2_{cal}$), and root mean square errors of calibration (RMSEC) for the calibration process. For the leave-one-out cross validation, the correlation coefficients of validation ($r_{val}$), coefficients of determination of validation ($R^2_{val}$), and root mean square errors of cross-validation (RMSECV) were calculated. For the independent test, the predictive abilities of the prediction set were evaluated in terms of the correlation coefficients of prediction ($r_{pre}$), coefficients of determination of prediction ($R^2_{pre}$), root mean square errors of prediction (RMSEP). Residual predictive deviation (RPD) were calculated for both leave-one-out cross validation and independent test. The above parameters are mainly used for evaluating the accuracy of spectral prediction. In addition, the models' robustness was evaluated by the absolute difference between RMSEC and RMSECV/RMSEP (AB_RMSE). Generally, a model with a good prediction ability should have large $r_{cal}$, $r_{val}$, $r_{pre}$, $R^2_{cal}$, $R^2_{val}$, $R^2_{pre}$ and RPD, and small RMSEC, RMSECV, RMSEP and AB_RMSE.

## 3. Results and discussion

### 3.1. Total anthocyanin content and antioxidant activity of mulberry fruit

The reference total anthocyanin content and antioxidant activity of mulberry fruit samples were measured using the reference methods mentioned above. The range of total anthocyanin content was 0.035–2.192 mg/g, 0.650 mg/g on average with standard deviation of 0.723 mg/g. The antioxidant activity measured using the DPPH assay were in the range 0.730–8.759 μmol/g (ascorbic acid equivalent) with average of 4.652 μmol/g (ascorbic acid equivalent) and standard deviation of 2.269 μmol/g (ascorbic acid equivalent). For both quality attributes, broad ranges of values were observed, showing that the samples were representative of possible range of total anthocyanin content and antioxidant activity, which was helpful to develop stable and robust calibration models.

### 3.2. Analysis of total anthocyanin content

In the analysis of total anthocyanins, the spectra in the full wavelength range were first used to calibrate the model to understand the overall abilities of hyperspectral imaging and also set a reference to evaluate the performances of wavelength selection. Both leave-one-out cross validation and independent test were used for analyzing the spectra in the full wavelength range. After that, the important wavelengths for total anthocyanin prediction were selected using four selection strategies, namely SPA, UVE, UVE-SPA and CARS, and their performances were compared by leave-one-out cross validation. Besides, Hyperspectral images in three spectral sets were considered respectively, and their results were compared to identify the best CCD detector.

#### 3.2.1. Prediction based on the spectra in the full wavelength range

Two calibration methods of PLSR and LS-SVM were conducted to predict total anthocyanin content of mulberry fruit using their corresponding spectra in full wavelength range measured by Si and InGaAs detectors respectively, and the results based on both the leave-one-out cross validation and three different calibration/prediction sets are shown in Table 1. When spectral set I was considered, good results were achieved in both calibration and validation/prediction conditions with an average $R^2_{val}/R^2_{pre}$ of 0.932 and an average RMSECV/RMSEP of 0.179 for all PLSR and LS-SVM models. Compared with spectral set I, the average RMSEC and an average RMSECV/RMSEP for spectral set II increased by 16.0% and 7.8%, respectively. When spectral set III was considered, its performances were similar to spectral set I. The above results show that, when the spectra in the full wavelength range were used for calibration, the Si detector and InGaAs detector had similar predictions for the total anthocyanin content, and their combination could not improve the prediction. In addition, to evaluate the robustness of hyperspectral imaging method for the prediction of the total anthocyanin content, the independent testing using three different calibration/prediction sets was considered. Their average RMSEC and RMSECV/RMSEP were similar to those of the corresponding models established based on the leave-one-out cross validation, and their AB_RMSE were small, showing that the hyperspectral imaging-based method for the prediction of the total anthocyanin content was robust.

**Table 1**
Evaluation of visible and near infrared hyperspectral imaging with full wavelength range for predicting the total anthocyanins content in mulberry fruit based on the leave-one-out cross validation and three different calibration/prediction sets of A, B, and C.

| Spectral set | Model evaluation | Calibration | LVs | Calibration | | | Cross-validation/prediction | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $r_{cal}$ | $R^2_{cal}$ | RMSEC | $r_{val}/r_{pre}$ | $R^2_{val}/R^2_{pre}$ | RMSECV/RMSEP | RPD |
| I | CV | PLSR | 2 | 0.977 | 0.955 | 0.153 | 0.973 | 0.949 | 0.168 | 4.303 |
| I | CV | LS-SVM | / | 0.986 | 0.973 | 0.119 | 0.971 | 0.942 | 0.174 | 4.163 |
| I | Set A | PLSR | 2 | 0.983 | 0.967 | 0.133 | 0.975 | 0.914 | 0.208 | 3.693 |
| I | Set A | LS-SVM | / | 1.000 | 1.000 | 0.003 | 0.980 | 0.808 | 0.310 | 2.479 |
| I | Set B | PLSR | 3 | 0.988 | 0.976 | 0.109 | 0.981 | 0.941 | 0.184 | 4.373 |
| I | Set B | LS-SVM | / | 0.989 | 0.978 | 0.104 | 0.983 | 0.949 | 0.171 | 4.695 |
| I | Set C | PLSR | 2 | 0.973 | 0.947 | 0.168 | 0.988 | 0.968 | 0.125 | 5.580 |
| I | Set C | LS-SVM | / | 0.981 | 0.962 | 0.143 | 0.994 | 0.982 | 0.093 | 7.462 |
| II | CV | PLSR | 2 | 0.973 | 0.946 | 0.168 | 0.965 | 0.934 | 0.190 | 3.794 |
| II | CV | LS-SVM | / | 0.985 | 0.970 | 0.126 | 0.960 | 0.918 | 0.206 | 3.505 |
| II | Set A | PLSR | 4 | 0.990 | 0.980 | 0.102 | 0.969 | 0.879 | 0.246 | 2.907 |
| II | Set A | LS-SVM | / | 0.993 | 0.985 | 0.089 | 0.980 | 0.927 | 0.192 | 3.730 |
| II | Set B | PLSR | 2 | 0.976 | 0.953 | 0.152 | 0.962 | 0.917 | 0.220 | 3.465 |
| II | Set B | LS-SVM | / | 0.986 | 0.972 | 0.118 | 0.975 | 0.940 | 0.187 | 4.084 |
| II | Set C | PLSR | 2 | 0.972 | 0.944 | 0.174 | 0.980 | 0.932 | 0.181 | 3.917 |
| II | Set C | LS-SVM | / | 0.979 | 0.957 | 0.152 | 0.991 | 0.969 | 0.123 | 5.759 |
| III | CV | PLSR | 2 | 0.980 | 0.959 | 0.146 | 0.975 | 0.954 | 0.160 | 4.530 |
| III | CV | LS-SVM | / | 0.986 | 0.973 | 0.120 | 0.972 | 0.945 | 0.170 | 4.267 |
| III | Set A | PLSR | 2 | 0.986 | 0.972 | 0.121 | 0.976 | 0.916 | 0.205 | 4.059 |
| III | Set A | LS-SVM | / | 0.998 | 0.996 | 0.046 | 0.985 | 0.877 | 0.248 | 3.359 |
| III | Set B | PLSR | 2 | 0.981 | 0.963 | 0.135 | 0.979 | 0.948 | 0.174 | 4.505 |
| III | Set B | LS-SVM | / | 0.987 | 0.975 | 0.111 | 0.982 | 0.953 | 0.166 | 4.722 |
| III | Set C | PLSR | 2 | 0.976 | 0.952 | 0.162 | 0.993 | 0.972 | 0.116 | 6.227 |
| III | Set C | LS-SVM | / | 0.982 | 0.964 | 0.139 | 0.998 | 0.983 | 0.090 | 8.069 |

CV: Cross validation; LVs: Number of latent variables.

*3.2.2. Selection of important wavelengths*

When the spectra in the full wavelength range were delivered to the calibration models, and there was no consideration given to select important wavelengths contributed for the quality prediction of mulberry fruits. Locating the best set of wavelengths for predicting is important to obviate the high dimensionality of the hyperspectral data for improvement of model's predictive accuracy and robustness. Selecting some important spectral variables instead of the full wavelength range has showed better predictive results in some cases (Chen, Liu, Cai, Zhu, & Chen, 2015; Zhu, Lin, Nie, Wu, & Chen, 2016). In this study, four wavelength selection strategies of SPA, UVE, UVE-SPA, and CARS were considered. Based on the selected variables, the spectral data were then reduced to new reduced spectral matrix by selecting the hyperspectral images only at the important wavelength, which contained the most relevant spectral information of total anthocyanin content in mulberry fruit. The new matrix was then used to replace the spectra in the full wavelength range for building new PLSR and LS-SVM models to predict the total anthocyanin content. The established models were evaluated by the leave-one-out cross validation, and the results are shown in Table 2.

In the SPA calculation, after comparing the root mean square errors of different candidate subsets of variables that were obtained by a sequence of projection operations, two, five and 21 variables which had the lowest root mean square errors were elected from the spectral ranges of spectral sets I, II, and III, respectively. The selected variables then formed new reduced spectral matrixes, which were set as the inputs of PLSR and LS-SVM models instead of the spectra in the full wavelength range. For spectral set I, the SPA models with only two input variables (0.39% of variables in the full wavelength range) had similar results compared with the models with the full wavelength range (F-model), showing that SPA could select variables with minimal redundancy. For spectral sets II and III, different improvements of SPA calculation were obtained, where their corresponding RMSECV decreased by 16.77% and increased by 18.82%, respectively. However, the differences between F-models and SPA models were not much in the aspects of both accuracy and robustness, indicating that SPA could significantly reduce the multicollinearity of full wavelength range and maintain models' accuracy. In addition, the SPA-PLSR models had better predictions than the SPA-LS-SVM models when spectral sets I and III were considered. When spectral set II was applied, the SPA-LS-SVM model was better than the SPA-PLSR model. Nevertheless the above differences between SPA-PLSR and SPA-LS-SVM models were not large. Similar to the full wavelength range, the AB_RMSE values of SPA-LS-SVM models were generally larger than SPA-PLSR models but still acceptable.

In the process of UVE calculation, the UVE based models had little more accurate and robust prediction for spectral sets I and II, where the average RMSECV of corresponding PLSR and LS-SVM models decreased by 6.24% and 10.55% respectively. However, UVE didn't improved the prediction of spectral set III, but models' input variables were much reduced from 768 to 122. The results indicated that UVE selected wavelengths with no more information for the prediction of total anthocyanin content than noise. On the other hand, compared with the corresponding SPA models, UVE models had lower RMSECV for spectral sets I (decreased by 12.97%) and III (decreased by 15.83%) but larger RMSECV for spectral set II increased by 7.47). Moreover, the UVE models were more robust than the SPA models for spectral sets I and II, where AB_RMSE decreased by 29.88% and 19.45%, respectively. In addition, the PLSR models had similar prediction of LS-SVM models, but were more robust (average AB_RMSE: 0.016 vs. 0.056).

Besides the calculation of SPA based on the full wavelength range, SPA is also commonly carried out after UVE calculation to select variables that informative but no collinearity (Wu & He, 2014; Wu et al., 2014). In this study, this strategy was also applied. Compared with SPA models, corresponding UVE-SPA based models improved both accuracy and robustness for spectral sets I and III, where RMSECV decreased by 10.37% and 17.32% while AB_RMSE decreased by 25.86% and 47.52% for spectral sets I and III, respectively. On the contrary, similar results for spectral sets II were found between SPA models and UVE-SPA models, which showed that the calculation of SPA directly on the full wavelength range

**Table 2**
Evaluation of visible and near infrared hyperspectral imaging with selected wavelengths for predicting the total anthocyanins content in mulberry fruit based on the leave-one-out cross validation.

| Spectral set | Wavelength selection | Variable number | Calibration | LVs | Calibration | | | Cross-validation | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | $r_{cal}$ | $R^2_{cal}$ | RMSEC | $r_{val}$ | $R^2_{val}$ | RMSECV | RPD |
| I | SPA | 2 | PLSR | 1 | 0.974 | 0.949 | 0.163 | 0.970 | 0.944 | 0.176 | 4.110 |
| I | SPA | 2 | LS-SVM | / | 0.984 | 0.968 | 0.130 | 0.964 | 0.929 | 0.193 | 3.750 |
| I | UVE | 103 | PLSR | 2 | 0.980 | 0.960 | 0.145 | 0.976 | 0.946 | 0.159 | 4.562 |
| I | UVE | 103 | LS-SVM | / | 0.987 | 0.973 | 0.119 | 0.975 | 0.950 | 0.162 | 4.467 |
| I | UVE-SPA | 4 | PLSR | 2 | 0.980 | 0.960 | 0.144 | 0.976 | 0.955 | 0.158 | 4.579 |
| I | UVE-SPA | 4 | LS-SVM | / | 0.984 | 0.968 | 0.130 | 0.971 | 0.943 | 0.172 | 4.192 |
| I | CARS | 5 | PLSR | 2 | 0.979 | 0.958 | 0.148 | 0.974 | 0.952 | 0.162 | 4.448 |
| I | CARS | 5 | LS-SVM | / | 0.988 | 0.976 | 0.113 | 0.968 | 0.938 | 0.180 | 4.005 |
| II | SPA | 5 | PLSR | 3 | 0.977 | 0.955 | 0.153 | 0.970 | 0.944 | 0.176 | 4.096 |
| II | SPA | 5 | LS-SVM | / | 0.989 | 0.979 | 0.106 | 0.977 | 0.954 | 0.154 | 4.688 |
| II | UVE | 174 | PLSR | 2 | 0.973 | 0.947 | 0.166 | 0.966 | 0.937 | 0.187 | 3.873 |
| II | UVE | 174 | LS-SVM | / | 0.986 | 0.972 | 0.121 | 0.973 | 0.946 | 0.168 | 4.294 |
| II | UVE-SPA | 6 | PLSR | 3 | 0.978 | 0.956 | 0.151 | 0.970 | 0.945 | 0.175 | 4.135 |
| II | UVE-SPA | 6 | LS-SVM | / | 0.988 | 0.976 | 0.111 | 0.974 | 0.948 | 0.165 | 4.382 |
| II | CARS | 9 | PLSR | 2 | 0.977 | 0.954 | 0.154 | 0.971 | 0.946 | 0.172 | 4.208 |
| II | CARS | 9 | LS-SVM | / | 0.988 | 0.976 | 0.112 | 0.980 | 0.959 | 0.146 | 4.964 |
| III | SPA | 21 | PLSR | 2 | 0.972 | 0.944 | 0.171 | 0.966 | 0.936 | 0.188 | 3.847 |
| III | SPA | 21 | LS-SVM | / | 0.989 | 0.978 | 0.108 | 0.961 | 0.921 | 0.203 | 3.558 |
| III | UVE | 122 | PLSR | 2 | 0.981 | 0.962 | 0.141 | 0.977 | 0.957 | 0.154 | 4.682 |
| III | UV | 122 | LS-SVM | / | 0.991 | 0.982 | 0.098 | 0.971 | 0.942 | 0.175 | 4.137 |
| III | UVE-SPA | 2 | PLSR | 2 | 0.981 | 0.963 | 0.140 | 0.977 | 0.958 | 0.153 | 4.727 |
| III | UVE-SPA | 2 | LS-SVM | / | 0.985 | 0.970 | 0.126 | 0.972 | 0.944 | 0.170 | 4.244 |
| III | CARS | 156 | PLSR | 2 | 0.981 | 0.962 | 0.142 | 0.977 | 0.957 | 0.155 | 4.662 |
| III | CARS | 156 | LS-SVM | / | 0.986 | 0.971 | 0.122 | 0.971 | 0.942 | 0.174 | 4.143 |

LVs: Number of latent variables.

could select important wavelengths efficiently, and the consideration of UVE was not required. On the other hand, UVE-SPA based models couldn't make the prediction better than corresponding UVE-based models, but the numbers of variables identified were much reduced (4 vs. 103 for set I, 6 vs. 174 for set II, and 2 vs. 122 for set III). In addition, all three spectral sets obtained similar predictions with average RMSECV of 0.166 and AB_RMSE of 0.032.

At last, CARS was carried out to select the important wavelengths based on the identification of wavelengths with higher absolute coefficients of PLSR models by the idea of 'survival of the fittest' from Darwin's Evolution Theory. Compared with the full wavelength range, similar results were found for spectral sets I and III, but a good improvement was achieved for spectral set II, in which RMSECV decreased by 20.04% and AB_RMSE decreased by 50.53%. On the other hand, compared with other wavelength section methods, most results were similar for all spectral sets, except the SPA models of spectral sets III and UVE models of spectral set II which had higher RMSECV than corresponding CARS models.

According to the evaluation standard, the best models for three spectral sets were the UVE-SPA-PLSR model, CARS-LS-SVM model, and the UVE-SPA-PLSR model, respectively. Compared with the corresponding best F-models, the RMSECV of the best models decreased by 5.95%, 18.95%, and 4.38%, the wavelength number decreased by 99.22%, 96.48%, and 99.74%, respectively and their AB_RMSE were similar. The above results showed that the prediction accuracy and robustness of the calibrated spectral models for predicting total anthocyanin content in mulberry fruit could be improved by considering the wavelength selection.

### 3.3. Analysis of antioxidant activity

Similar to the analysis of total anthocyanins, the investigation of antioxidant activity started at the calibration based on the spectra in the full wavelength range with the analysis of three spectral sets from two detectors, and the results based on both the leave-one-out cross validation and three different calibration/prediction sets are shown in Table 3. When full wavelength range was considered, the three sets had similar predictions for antioxidant activity. The average RMSECV/RMSEP for all PLSR and LS-SVM models were 0.614, 0.656, and 0.652 for spectral sets I, II and III, respectively. Moreover, the evaluation by the independent testing with three different calibration/prediction sets obtained the similar average RMSECV/RMSEP and AB_RMSE, compared with the corresponding models established based on the leave-one-out cross validation, showing that the hyperspectral imaging-based method for the prediction of the antioxidant activity was robust.

The wavelength selection was then implemented to evaluate whether the performances of three spectral sets could be further improved. The established models were evaluated by the leave-one-out cross validation, and the results are shown in Table 4. When SPA was carried out for selecting important wavelength based on the full wavelength range. There were 20, 10, and 25 wavelengths selected for spectral sets I, II and III, respectively. New reduced spectral matrixes were then formed based on these variables to establish PLSR and LS-SVM models. Compared with the F-models of three spectral sets, the application of SPA improved the corresponding PLSR and LS-SVM models with the decrements of average RMSECV by 15.36%, 26.69% and 32.76%, respectively. Average AB_RMSE were also decreased by 24.38% and 41.05% for spectral sets I and II, while similar to the F-model of spectral III. It was noticed that the SPA-LS-SVM models had better prediction than corresponding SPA-PLSR models, in which the RMSECV of former ones were only 71.75%, 62.24%, and 66.82% of latter ones, respectively.

When UVE was used to eliminate uninformative variables, the models of spectral sets I and III had improved results with the decrements of average RMSECV by 11.84% and 20.57%. However, UVE calculation could not improve the prediction for spectral set II. Moreover, most UVE models had poorer predictive accuracy than corresponding SPA models. The other strategy of SPA process for wavelength selection was to execute it based on the informative wavelengths selected by UVE instead of the spectra in the full wavelength range. On one side, compared with UVE models, the UVE-SPA models had similar results for spectral sets I and II, and

**Table 3**
Evaluation of visible and near infrared hyperspectral imaging with full wavelength range for predicting the DPPH antioxidant activity in mulberry fruit based on the leave-one-out cross validation and three different calibration/prediction sets of A, B, and C.

| Spectral set | Model evaluation | Calibration | LVs | Calibration | | | Cross-validation/Prediction | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | $r_{cal}$ | $R^2_{cal}$ | RMSEC | $r_{val}/r_{pre}$ | $R^2_{val}/R^2_{pre}$ | RMSECV/RMSEP | RPD |
| I | CV | PLSR | 2 | 0.963 | 0.928 | 0.609 | 0.956 | 0.918 | 0.667 | 3.402 |
| I | CV | LS-SVM | / | 0.999 | 0.998 | 0.098 | 0.956 | 0.912 | 0.672 | 3.419 |
| I | Set A | PLSR | 2 | 0.956 | 0.914 | 0.645 | 0.973 | 0.945 | 0.559 | 4.459 |
| I | Set A | LS-SVM | / | 0.994 | 0.988 | 0.239 | 0.967 | 0.930 | 0.631 | 3.947 |
| I | Set B | PLSR | 2 | 0.967 | 0.936 | 0.581 | 0.953 | 0.905 | 0.684 | 3.237 |
| I | Set B | LS-SVM | / | 0.972 | 0.943 | 0.550 | 0.977 | 0.952 | 0.487 | 4.549 |
| I | Set C | PLSR | 2 | 0.968 | 0.937 | 0.579 | 0.953 | 0.907 | 0.669 | 3.414 |
| I | Set C | LS-SVM | / | 0.990 | 0.978 | 0.339 | 0.973 | 0.939 | 0.542 | 4.211 |
| II | CV | PLSR | 1 | 0.961 | 0.923 | 0.628 | 0.955 | 0.917 | 0.673 | 3.371 |
| II | CV | LS-SVM | / | 0.999 | 0.998 | 0.104 | 0.960 | 0.922 | 0.635 | 3.573 |
| II | Set A | PLSR | 1 | 0.950 | 0.902 | 0.691 | 0.981 | 0.959 | 0.485 | 4.932 |
| II | Set A | LS-SVM | / | 0.981 | 0.962 | 0.431 | 0.932 | 0.868 | 0.867 | 2.759 |
| II | Set B | PLSR | 1 | 0.968 | 0.937 | 0.574 | 0.949 | 0.889 | 0.737 | 3.006 |
| II | Set B | LS-SVM | / | 0.981 | 0.963 | 0.441 | 0.956 | 0.911 | 0.660 | 3.355 |
| II | Set C | PLSR | 1 | 0.965 | 0.931 | 0.604 | 0.954 | 0.904 | 0.681 | 3.367 |
| II | Set C | LS-SVM | / | 0.989 | 0.978 | 0.340 | 0.975 | 0.946 | 0.508 | 4.510 |
| III | CV | PLSR | 2 | 0.966 | 0.932 | 0.591 | 0.959 | 0.923 | 0.647 | 3.506 |
| III | CV | LS-SVM | / | 0.988 | 0.976 | 0.351 | 0.949 | 0.899 | 0.720 | 3.163 |
| III | Set A | PLSR | 4 | 0.990 | 0.980 | 0.313 | 0.973 | 0.898 | 0.762 | 3.843 |
| III | Set A | LS-SVM | / | 0.995 | 0.989 | 0.232 | 0.966 | 0.931 | 0.627 | 3.229 |
| III | Set B | PLSR | 2 | 0.971 | 0.942 | 0.553 | 0.954 | 0.904 | 0.686 | 3.813 |
| III | Set B | LS-SVM | / | 0.970 | 0.939 | 0.568 | 0.965 | 0.931 | 0.581 | 3.511 |
| III | Set C | PLSR | 2 | 0.969 | 0.939 | 0.568 | 0.959 | 0.915 | 0.640 | 4.053 |
| III | Set C | LS-SVM | / | 0.991 | 0.981 | 0.317 | 0.973 | 0.936 | 0.554 | 3.163 |

CV: Cross validation; LVs: Number of latent variables.

**Table 4**
Predictive results of DPPH antioxidant activity in mulberry fruit using visible and near infrared hyperspectral imaging with chemometrics based on the leave-one-out cross validation.

| Spectral set | Wavelength selection | Variable number | Calibration | LVs | Calibration | | | Cross-validation | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | $r_{cal}$ | $R^2_{cal}$ | RMSEC | $r_{val}$ | $R^2_{val}$ | RMSECV | RPD |
| I | SPA | 20 | PLSR | 4 | 0.972 | 0.946 | 0.530 | 0.957 | 0.920 | 0.660 | 3.439 |
| I | SPA | 20 | LS-SVM | / | 0.998 | 0.997 | 0.126 | 0.979 | 0.956 | 0.473 | 4.793 |
| I | UVE | 109 | PLSR | 3 | 0.969 | 0.940 | 0.558 | 0.960 | 0.926 | 0.633 | 3.586 |
| I | UVE | 109 | LS-SVM | / | 0.992 | 0.984 | 0.287 | 0.971 | 0.942 | 0.547 | 4.145 |
| I | UVE-SPA | 16 | PLSR | 2 | 0.961 | 0.923 | 0.630 | 0.953 | 0.913 | 0.689 | 3.292 |
| I | UVE-SPA | 16 | LS-SVM | / | 0.992 | 0.984 | 0.284 | 0.971 | 0.943 | 0.540 | 4.202 |
| I | CARS | 2 | PLSR | 1 | 0.967 | 0.934 | 0.582 | 0.962 | 0.929 | 0.621 | 3.656 |
| I | CARS | 2 | LS-SVM | / | 0.986 | 0.972 | 0.379 | 0.969 | 0.939 | 0.561 | 4.041 |
| II | SPA | 10 | PLSR | 4 | 0.983 | 0.965 | 0.422 | 0.966 | 0.936 | 0.591 | 3.848 |
| II | SPA | 10 | LS-SVM | / | 0.996 | 0.992 | 0.198 | 0.987 | 0.974 | 0.368 | 6.167 |
| II | UVE | 9 | PLSR | 1 | 0.950 | 0.902 | 0.709 | 0.943 | 0.895 | 0.757 | 2.999 |
| II | UVE | 9 | LS-SVM | / | 0.977 | 0.955 | 0.480 | 0.950 | 0.903 | 0.707 | 3.212 |
| II | UVE-SPA | 3 | PLSR | 2 | 0.958 | 0.919 | 0.648 | 0.951 | 0.909 | 0.703 | 3.229 |
| II | UVE-SPA | 3 | LS-SVM | / | 0.993 | 0.986 | 0.272 | 0.938 | 0.878 | 0.792 | 2.874 |
| II | CARS | 9 | PLSR | 3 | 0.985 | 0.970 | 0.395 | 0.972 | 0.947 | 0.537 | 4.232 |
| II | CARS | 9 | LS-SVM | / | 0.998 | 0.995 | 0.153 | 0.966 | 0.932 | 0.590 | 3.844 |
| III | SPA | 25 | PLSR | 4 | 0.986 | 0.972 | 0.383 | 0.970 | 0.944 | 0.551 | 4.123 |
| III | SPA | 25 | LS-SVM | / | 1.000 | 0.999 | 0.072 | 0.987 | 0.974 | 0.368 | 6.225 |
| III | UVE | 78 | PLSR | 4 | 0.986 | 0.973 | 0.372 | 0.974 | 0.951 | 0.518 | 4.382 |
| III | UV | 78 | LS-SVM | / | 0.993 | 0.987 | 0.264 | 0.968 | 0.937 | 0.568 | 3.996 |
| III | UVE-SPA | 8 | PLSR | 4 | 0.987 | 0.974 | 0.369 | 0.979 | 0.961 | 0.462 | 4.916 |
| III | UVE-SPA | 8 | LS-SVM | / | 0.992 | 0.984 | 0.290 | 0.980 | 0.961 | 0.449 | 5.060 |
| III | CARS | 18 | PLSR | 5 | 0.992 | 0.983 | 0.294 | 0.982 | 0.966 | 0.433 | 5.255 |
| III | CARS | 18 | LS-SVM | / | 1.000 | 1.000 | 0.049 | 0.998 | 0.995 | 0.159 | 14.255 |

LVs: Number of latent variables.

improvements were achieved for set III, where RMSECV were decreased by 10.85% and 21.02% for PLSR and LS-SVM models respectively. It should be noted that such maintain or even improvement of prediction were achieved based on dramatic decrement of input variables. Only 14.68%, 33.33% and 10.26% of input variables remained for spectral sets I to III. However, on the other side, most UVE-SPA models had poorer prediction than corresponding UVE models. Especially for spectral set II, in which the RMSECV of PLSR and LS-SVM models increased by 18.88% and 115.3% respectively. Therefore, the consideration of UVE before SPA had no improvement for the wavelength selection for the antioxidant activity analysis.

On the basis of the wavelengths selected by CARS, both PLSR and LS-SVM models were also established for three spectral sets. When set I was considered, the CARS models had an improvement of prediction compared with the corresponding F-models, in which

the average RMSECV and AB_RMSE decreased by 11.70% and 64.89%, respectively. When the performances of CARS models were compared with the corresponding models established based on the wavelengths identified by other wavelength selection methods, it was found that there were not much differences among these models. For spectral set II, the CARS models had a better prediction than corresponding F-models, UVE models, and UVE-SPA models, but worse than SPA models in both accuracy and robustness. At last, when spectral set III was considered, the CARS models were better than all the models established based on full wavelength range or the wavelengths selected by other methods. Especially for the CARS-LS-SVM model, its RMSECV and AB_RMSE were only 32.22% and 43.20% of the average values of those from other models.

In general, the results of SPA-LS-SVM models were better than other corresponding models for all three spectral sets, except the CARS-LS-SVM model which was the best model for not only spectral set III, but also the best for antioxidant activity prediction. Moreover, similar to the analysis of total anthocyanins, the wavelength selection improved the prediction of the spectral models. Compared with the corresponding best F-models, the RMSECV of the best models with wavelength selection decreased by 29.02%, 45.34%, and 75.41% and the wavelength number decreased by 96.09%, 96.09%, and 97.66% for three spectral sets, respectively.

### 3.4. Discussion

The potential of three spectral sets measured by two hyperspectral imaging detectors of Si and InGaAs was investigated for the rapid and reliable assessment of total anthocyanin content and antioxidant activity of mulberry fruit, and the results are shown in Tables 1 and 2. Specifically, for the total anthocyanins analysis, the results of the best models for three spectral sets were similar, in which the RPD values were between 4.5 and 5 for these models. For antioxidant activity, the best predictive results were obtained by the CARS-LS-SVM model which was established based on the spectra from spectral set III. It had the highest RPD of 14.255 with a good AB_RMSE of 0.110. On the other hand, the best models for spectral sets I and II were also acceptable, which were both SPA-LS-SVM models. Their RPD were 4.793 and 6.167, respectively. According to the standard from the literature (Nicolaï et al., 2007), RPD over 2.5 corresponds to excellent prediction accuracy. Therefore, it was confirmed the potential of hyperspectral imaging for the rapid and reliable assessment of total anthocyanin content and antioxidant activity of mulberry fruit in a rapid and non-destructive manner. It would help the fruit industry to better understand and control the quality of mulberry fruit and further truthfully label and price the fruits according to their quality, leading to increased competitiveness in commercial markets.

To evaluate two detectors of Si (spectral set I) and InGaAs (spectral set II) from the hyperspectral imaging systems used in this study to predict the anthocyanin content and antioxidant activity of mulberry fruit, all the multivariate models were established based on the spectra measured by the two detectors and their combination. In the analysis of total anthocyanin, three spectral sets with the best models had similar results. For the antioxidant activity analysis, the combination of two detectors obtained the best prediction with RMSECV lower than 0.2. The prediction of InGaAs detector was also acceptable with RMSECV lower than 0.4, and even the Si detector had a good prediction with RMSECV lower than 0.5.

Multivariate modeling methods are often implemented to establish predictive models between hyperspectral imaging data and the desired attributes of tested samples. In this work, two typical modeling methods of PLSR and LS-SVM were considered and compared, in which the former one is commonly used to establish linear calibration models, while the latter is a classical non-linear regression method. For total anthocyanin content, among the fifteen models (three spectral sets $\times$ five wavelength sets of full wavelength range, SPA, UVE, UVE-SPA, and CARS), eleven PLSR models had better predictive accuracy than corresponding LS-SVM models. On the other hand, for antioxidant activity analysis, among the fifteen models, ten PLSR models were worse than corresponding LS-SVM models in the aspect of accuracy. Moreover, the best models for total anthocyanins and antioxidant activity were both LS-SVM models. The above statistical data showed that in order to obtain the best calibration model, it is required to use different modeling methods and compare their performances. In addition, the average AB_RMSE of fifteen PLSR models were only 29.98% of that of the corresponding LS-SVM models, showing PLSR could commonly get much robust models than LS-SVM method. Nevertheless, the AB_RMSE of some LS-SVM models were also accepted, especially those who were the best models for total anthocyanins and antioxidant activity.

The elimination of irrelevant spectral wavelengths might predigest calibration modeling and improve the results in terms of accuracy and robustness. In this study, among all six cases (three spectral sets $\times$ two quality attributes), all the best models with wavelength selection had better prediction than the corresponding best F-models. For total anthocyanin content, the RPD increased by 6.41%, 30.86%, and 4.34% for sets I, II, and III, respectively. Much improvement was obtained in the antioxidant activity analysis, where the spectral set I to III had the increment of RPD by 40.90%, 82.95%, and 306.64, respectively. Such improvement was achieved after over 95 % wavelengths from the full wavelength range were not considered for the modeling. Therefore, the wavelength selection with the help of SPA, UVE or CARS improved the predictions of both the anthocyanin content and antioxidant activity of mulberry fruit. However, the performances of different wavelength selection methods were different, especially the best models with wavelength selection were achieved based on different methods. Therefore, it is necessary to compare different wavelength selection methods and choose the best one according to models' accuracy and robustness. On the other hand, to interpret the selected wavelengths, the correlations between the total anthocyanin content and antioxidant activity and the spectra were calculated for both spectral sets I and II. It was found that the wavelengths with the correlation coefficients larger than 0.9 were those near 600 nm and between 900 and 1100 nm for the total anthocyanin content and those near 580 nm and between 950 and 1300 nm for the antioxidant activity. Although different selection methods selected different wavelengths, most selected wavelengths were within the above wavelength ranges. These wavelengths were selected and successfully used for the prediction, which might be because they are highly correlated to the total anthocyanin content/antioxidant activity. The spectra at wavelengths between 580 and 600 nm mainly reflect the colour information of red, and anthocyanin is the main reason causing the mulberry fruits having red and purple colours (Aramwit, Bang, & Srichana, 2010). Also anthocyanin is a main component making the mulberry fruits with the function of antioxidant activity (Bae & Suh, 2007; Koca, Ustun, Koca, & Karadeniz, 2008). That might be the reason why the wavelengths between 580 and 600 nm were selected for the prediction of total anthocyanin content and antioxidant activity. In the future, further analysis should be carried out to investigate why the total anthocyanin content and antioxidant activity of mulberry fruits could be predicted based on the spectra at the selected wavelengths, especially why the spectral at the wavelengths between 900–1100 nm and 950–1300 nm had high correlations with the total anthocyanin content and the antioxidant activity, respectively, and were selected for the prediction. In addition, it should be noted that an optimized multispectral

imaging system could be developed based on the selected important wavelengths for industrial purpose. The use of multispectral imaging systems can signify a reduction in the costs, compared with the hyperspectral imaging systems. Therefore, the selection of several key wavelengths is very important. In the future works, more samples from different varieties, different climate and soil conditions, different preharvest planting patterns and postharvest storage strategies, different ripening stages, and different years will be considered. On the basis of a large number of samples with different calibration and independent prediction sets, the final best wavelengths and calibration methods will be determined to further produce multispectral imaging systems for industry use.

## 4. Conclusions

This study was carried out to develop a Vis-NIR hyperspectral imaging inspection method to predict total anthocyanin content and antioxidant activity of mulberry fruit. Two detectors of Si and InGaAs in two hyperspectral imaging systems were compared to choose the best detector for inspecting the quality of mulberry fruit. Considering the predictive results and the costs of detectors, it was recommended to use the Si detector for the prediction of total anthocyanin content in mulberry fruit. On the other hand, the choice of detectors for the prediction of antioxidant activity should depend on the required predictive accuracy and affordable cost. It should be noted that hyperspectral imaging was very advantageous to extract the spectra of mulberry fruits on the basis of spatial information contained in the acquired hyperspectral images. Such spectral extraction could not be done efficiently by common spectroscopy instruments, as the shape of mulberry fruits is not round and their size is small. The best prediction models for the analysis of total anthocyanin content and antioxidant activity had $R^2_{val}$ of 0.959 and 0.995 and RPD of 4.964 and 14.255, respectively. The results of this study indicate that Vis-NIR spectra together with chemometrics of wavelength selection and multivariate modeling hold the advantages of this method to be a rapid, accurate, objective and non-invasive tool for the prediction of total anthocyanin content and antioxidant activity of mulberry fruit. The capability of hyperspectral imaging to predict the quality of mulberry fruit and further grade and price the fruits is critical to both consumers and industries in public-health and economic terms.

## Ethical statement

This article does not contain any studies with human participants or animals performed by any of the authors.

## Informed consent

Not applicable.

## Conflict of interest statement

Lingxia Huang declares that she has no conflict of interest. Yibin Zhou declares that he has no conflict of interest. Liuwei Meng declares that he has no conflict of interest. Di Wu declares that he has no conflict of interest. Yong He declares that he has no conflict of interest.

## Acknowledgements

This work was supported by Zhejiang Silkworm Industry Science and Technology Innovation Team (2011R50028), National Natural Science Foundation of China (31401541), and Huzhou Public Welfare Technology Application Research Project (2015GZ12).

## References

Aramwit, P., Bang, N., & Srichana, T. (2010). The properties and stability of anthocyanins in mulberry fruits. *Food Research International, 43*(4), 1093–1097.
Araujo, M. C. U., Saldanha, T. C. B., Galvao, R. K. H., Yoneyama, T., Chame, H. C., & Visani, V. (2001). The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemometrics and Intelligent Laboratory Systems, 57*(2), 65–73.
Bae, S. H., & Suh, H. J. (2007). Antioxidant activities of five different mulberry cultivars in Korea. *Food Science & Technology, 40*(6), 955–962.
Cen, H. Y., & He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science & Technology, 18*(2), 72–83.
Centner, V., Massart, D. L., deNoord, O. E., deJong, S., Vandeginste, B. M., & Sterna, C. (1996). Elimination of uninformative variables for multivariate calibration. *Analytical Chemistry, 68*(21), 3851–3858.
Chen, C. C., Liu, L. K., Hsu, J. D., Huang, H. P., Yang, M. Y., & Wang, C. J. (2005). Mulberry extract inhibits the development of atherosclerosis in cholesterol-fed rabbits. *Food Chemistry, 91*(4), 601–607.
Chen, P. N., Chu, S. C., Chiou, H. L., Kuo, W. H., Chiang, C. L., & Hsieh, Y. S. (2006). Mulberry anthocyanins, cyanidin 3-rutinoside and cyanidin 3-glucoside, exhibited an inhibitory effect on the migration and invasion of a human lung cancer cell line. *Cancer Letters, 235*(2), 248–259.
Chen, S., Zhang, F., Ning, J., Liu, X., Zhang, Z., & Yang, S. (2015a). Predicting the anthocyanin content of wine grapes by NIR hyperspectral imaging. *Food Chemistry, 172*(172C), 788–793.
Chen, X., Liu, K., Cai, J., Zhu, D., & Chen, H. (2015b). Identification of heavy metal-contaminated Tegillarca granosa using infrared spectroscopy. *Analytical Methods, 7*, 2172–2181.
Chen, X., Wu, D., Guan, X., Liu, B., Liu, G., Yan, M., et al. (2013). Feasibility of infrared and Raman spectroscopies for identification of juvenile black seabream (*Sparus macrocephalus*) intoxicated by heavy metals. *Journal of Agricultural and Food Chemistry, 61*(50), 12429–12435.
Du, Q., Zheng, J., & Xu, Y. (2008). Composition of anthocyanins in mulberry and their antioxidant activity. *Journal of Food Composition and Analysis, 21*(5), 390–395.
Ercisli, S., & Orhan, E. (2007). Chemical composition of white (Morus alba), red (Morus rubra) and black (Morus nigra) mulberry fruits. *Food Chemistry, 103*(4), 1380–1384.
Fernandes, A. M., Franco, C., Mendes-Ferreira, A., Mendes-Faia, A., da Costa, P. L., & Melo-Pinto, P. (2015). Brix, pH and anthocyanin content determination in whole Port wine grape berries by hyperspectral imaging and neural networks. *Computers and Electronics in Agriculture, 115*, 88–96.
Giovanelli, G., Sinelli, N., Beghi, R., Guidetti, R., & Casiraghi, E. (2014). NIR spectroscopy for the optimization of postharvest apple management. *Postharvest Biology and Technology, 87*(1), 13–20.
Hernandez-Hierro, J. M., Nogales-Bueno, J., Rodriguez-Pulido, F. J., & Heredia, F. J. (2013). Feasibility study on the use of near-infrared hyperspectral imaging for the screening of anthocyanins in intact grapes during ripening. *Journal of Agricultural and Food Chemistry, 61*(41), 9804–9809.
Huang, G., Ruan, X., Chen, X., Lin, D., & Liu, W. (2014). A segmented PLS method based on genetic algorithm. *Analytical Methods, 6*(9), 2900–2908.
Huang, L., Liu, H., Zhang, B., & Wu, D. (2015). Application of electronic nose with multivariate analysis and sensor selection for botanical origin identification and quality determination of honey. *Food & Bioprocess Technology, 8*(2), 359–370.
Jiang, Y., & Nie, W. J. (2015). Chemical properties in fruits of mulberry species from the Xinjiang province of China. *Food Chemistry, 174*, 460–466.
Koca, I., Ustun, N. S., Koca, A. F., & Karadeniz, B. (2008). Chemical composition, antioxidant activity and anthocyanin profiles of purple mulberry (Morus rubra) fruits. *Journal of Food Agriculture and Environment, 6*(2), 39.
Lee, J., Durst, R. W., & Wrolstad, R. E. (2005). Determination of total monomeric anthocyanin pigment content of fruit juices, beverages, natural colorants, and wines by the pH differential method: Collaborative study. *Journal of AOAC International, 88*(5), 1269–1278.
Li, H., Liang, Y., Xu, Q., & Cao, D. (2009). Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Analytica Chimica Acta, 648*(1), 77–84.
Liu, C., Wei, L., Wei, C., Yang, J., & Lei, Z. (2015). Feasibility in multispectral imaging for predicting the content of bioactive compounds in intact tomato fruit. *Food Chemistry, 173*, 482–488.
Nicolaï, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., et al. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology, 46*(2), 99–118.
Nogales-Bueno, J., Baca-Bocanegra, B., Rodriguez-Pulido, F. J., Heredia, F. J., & Hernandez-Hierro, J. M. (2015). Use of near infrared hyperspectral tools for the screening of extractable polyphenols in red grape skins. *Food Chemistry, 172*, 559–564.
Oki, T., Kobayashi, M., Nakamura, T., Okuyama, M., Masuda, M., Shiratsuchi, H., et al. (2006). Changes in radical-scavenging activity and components of mulberry fruit during maturation. *Journal of Food Science, 71*(1), C18–C22.
Saura-Calixto, F., Perez-Jimenez, J., Arranz, S., Tabernero, M., Diaz-Rubio, M. E., Serrano, J., et al. (2008). Updated methodology to determine antioxidant capacity in plant foods, oils and beverages: Extraction, measurement and expression of. *Food Research International, 41*(3), 274–285.
Sinelli, N., Spinardi, A., Egidio, V. D., Mignani, I., & Casiraghi, E. (2008). Evaluation of quality and nutraceutical content of blueberries (Vaccinium corymbosum L.) by

near and mid-infrared spectroscopy. *Postharvest Biology and Technology, 50*(1), 31–36.
Thaipong, K., Boonprakob, U., Crosby, K., Cisneros-Zevallos, L., & Byrne, D. H. (2006). Comparison of ABTS, DPPH, FRAP, and ORAC assays for estimating antioxidant activity from guava fruit extracts. *Journal of Food Composition and Analysis, 19* (6–7), 669–675.
Wu, D., Chen, X., Cao, F., Sun, D.-W., He, Y., & Jiang, Y. (2014). Comparison of infrared spectroscopy and nuclear magnetic resonance techniques in tandem with multivariable selection for rapid determination of ω-3 polyunsaturated fatty acids in fish oil. *Food and Bioprocess Technology, 7*(6), 1555–1569.
Wu, D., & He, Y. (2014). Potential of spectroscopic techniques and chemometric analysis for rapid measurement of docosahexaenoic acid and eicosapentaenoic acid in algal oil. *Food Chemistry, 158*, 93–100.
Wu, D., Shi, H., Wang, S., He, Y., Bao, Y., & Liu, K. (2012). Rapid prediction of moisture content of dehydrated prawns using online hyperspectral imaging system. *Analytica Chimica Acta, 726*, 57–66.
Wu, D., & Sun, D.-W. (2013a). Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review – Part I: Fundamentals. *Innovative Food Science and Emerging Technologies, 19*, 1–14.
Wu, D., & Sun, D.-W. (2013b). Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review – Part II: Applications. *Innovative Food Science and Emerging Technologies, 19*, 15–28.
Wu, D., & Sun, D.-W. (2013c). Potential of time series-hyperspectral imaging (TS-HSI) for non-invasive determination of microbial spoilage of salmon flesh. *Talanta, 111*, 39–46.
Wu, D., Sun, D.-W., & He, Y. (2014). Novel non-invasive distribution measurement of texture profile analysis (TPA) in salmon fillet by using visible and near infrared hyperspectral imaging. *Food Chemistry, 145*, 417–426.
Yang, Y. C., Sun, D. W., Pu, H., Wang, N. N., & Zhu, Z. (2015). Rapid detection of anthocyanin content in lychee pericarp during storage using hyperspectral imaging coupled with model fusion. *Postharvest Biology and Technology, 103*, 55–65.
Zhu, N., Lin, M., Nie, Y., Wu, D., & Chen, K. (2016a). Study on the quantitative measurement of firmness distribution maps at the pixel level inside peach pulp. *Computers and Electronics in Agriculture, 130*, 48–56.
Zhu, N., Nie, Y., Wu, D., He, Y., & Chen, K. (2016b). Feasibility study on quantitative pixel-level visualization of internal quality at different cross-sections inside postharvest loquat fruit. *Food Analytical Methods*. http://dx.doi.org/10.1007/s12161-016-0581-8, in press.
Zou, X. B., Zhao, J. W., Povey, M. J. W., Holmes, M., & Mao, H. P. (2010). Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta, 667*(1–2), 14–32.
Zude, M., Pflanz, M., Spinelli, L., Dosche, C., & Torricelli, A. (2011). Non-destructive analysis of anthocyanins in cherries by means of Lambert-Beer and multivariate regression based on spectroscopy and scatter correction using time-resolved analysis. *Journal of Food Engineering, 103*(1), 68–75.