



# Rapid and Nondestructive Freshness Determination of Tilapia Fillets by a Portable Near-Infrared Spectrometer Combined with Chemometrics Methods

Hai-Dong Yu<sup>1</sup> · Si-Min Zuo<sup>1</sup> · Guanghua Xia<sup>1,2</sup> · Xing Liu<sup>1</sup> · Yong-Huan Yun<sup>1</sup> · Chenghui Zhang<sup>1</sup>

Received: 24 March 2020 / Accepted: 28 June 2020 / Published online: 6 July 2020  
© Springer Science+Business Media, LLC, part of Springer Nature 2020

## Abstract

To assess the freshness of tilapia fillets rapidly and nondestructively, near-infrared (NIR) spectroscopy technology in the spectral range from 900 to 1700 nm was conducted by means of a portable spectrometer to determine the total volatile basic nitrogen (TVB-N) value of tilapia fillets stored at 4 °C. To establish a robust and high predictive NIR model between the collected NIR spectra and measured TVB-N values, a series of chemometric methods, including outlier elimination, spectral preprocessing, and optimal wavelength selection, were conducted. First, outlier elimination was utilized continuously by two times for the removal of the outlier samples. The results illustrated that outlier elimination was effective and useful to reduce outliers and develop the robust models. Second, spectral preprocessing was executed based on five preprocessing methods as well as their hybrid methods to pretreat the spectral data, and smoothing-first derivative (SM-D1) was proved to be the most effective methodology as result of its highest  $R^2_p$  and lowest RMSEP. Finally, optimal wavelength selection was carried out by several variable selection methods to establish simpler models with better predictive performance. It was manifested that the combination of iPLS-mVCPA-IRIV was proved to be the most reliable and outstanding method based on the  $R^2_p$  of 0.9201 and RMSEP of 2.0907. In general, it has been proved that NIR spectroscopy technique has the ability to measure TVB-N content of fillets for the evaluation of fish freshness in a rapid and nondestructive manner.

**Keywords** Tilapia fillets · Freshness determination · Near-infrared spectroscopy · Chemometrics · Wavelength selection · Three-step hybrid strategy

## Introduction

Tilapia has become one of the most popular fisheries not only in China but also all over the world since it has the advantages such as strong hypoxia tolerance, disease-resistance ability, fast growth, and breed (Duan et al. 2011). Moreover, tilapia flesh, which owns the characteristics of tender texture, delicious taste, and high nutritional value, has been popular in the inland and international markets (Guan et al. 2013). It has been reported that

China has become one of main countries producing tilapia whose yield has reached 1.58 million tons (FAO 2017).

However, the features of vulnerability and perishability that the flesh of fish possesses play a negative impact on the fish freshness quality (Li et al. 2017). Freshness has been seen as one of the most crucial quality characters in estimating the quality of fish, consuming directly, and processing raw materials for aquatic industries (Özogul et al. 2005). To evaluate the freshness quality, many methods and techniques have been developed, including sensory evaluation (Green et al. 2011), quality index method (QIM) (Li et al. 2017), microbial inspection with the number of *Escherichia coli* (*E. coli*) loads (Cheng and Sun 2015) and total viable counts (Song et al. 2012), biochemical methodologies of estimating adenosine triphosphate (ATP) degradation commonly employing high-performance liquid chromatography (Ocaño-Higuera et al. 2011), and other indicators of assessing freshness quality consisting of total volatile basic nitrogen (TVB-N), trimethylamine (TMA), and formation of biogenic amine (Özogul et al. 2005). However, traditional methods achieved

✉ Yong-Huan Yun  
yunyonghuan@foxmail.com

✉ Chenghui Zhang  
zchlm@163.com

<sup>1</sup> College of Food Science and Engineering, Hainan University, 58 Renmin Road, Haikou 570228, China

<sup>2</sup> Hainan Engineering Research Center of Aquatic Resources Efficient Utilization in South China Sea, Hainan University, Haikou 570228, China

to determine these indicators are destructive, time-consuming, costly, laborious, and tedious.

Near-infrared (NIR) spectroscopy as a kind of fast as well as noninvasive analytical tool has been proved possibly replacing destructive and time-consuming traditional methods. It has been employed broadly in Chinese herb (Yun et al. 2015), agricultural products (Li et al. 2020), textile (Chen et al. 2019), and pharmaceutical (Sarragaça et al. 2010). Besides the above applications, NIR spectroscopy coupled with chemometrics also has the potential of quantitatively analyzing different attributes of fish flesh. Over the past two decades, plenty of studies have been reported applying NIR spectroscopy technique to predict properties of fish flesh, especially chemical compositions which contain protein (Uddin et al. 2006), water (Uddin et al. 2006), and fat contents (Liu et al. 2015), and indicators of fish freshness quality, particularly fish freshness indicators including K value (Zhou et al. 2019), TVB-N content (Zhou et al. 2019), texture (Wu et al. 2014), thiobarbital acid content (Zhou et al. 2019), and microbial numbers (Tito et al. 2012). However, no paper was published in the evaluation of fish freshness conducting a handheld NIR spectrometer in combination with chemometrics algorithms to predict TVB-N content in tilapia. Moreover, the portable NIR spectrometer has the advantages of easiness in installation, maintenance, and movement, which is helpful to be used widely in the quantitative and qualitative analysis anytime and anywhere.

Meanwhile, collected original NIR spectra usually contain a large number of redundant and interfering variables (wavelengths) decreasing the performance of the model built. In order to improve the accuracy and efficiency of the model built on the basis of collected spectra and corresponding reference values, variable (wavelength) selection methods are used for the elimination of redundant and noisy wavelengths before modeling. In recent years, many variable selection techniques have been proposed and implemented on NIR spectrum. They have proved that variable selection is an indispensable and crucial procedure to establish the NIR model. They were divided by Yun et al. (2019b) into two types: wavelength interval selection (WIS) and wavelength point selection (WPS) methods. WIS methods often have a higher calculation efficiency, strong interpretability, but lower prediction performance of models established by remained interval combination. On the contrary, WPS methods often obtain a higher predicted performance of the model built using the selected variables but need a larger computation cost. It is obvious that they all have their own merits and drawbacks. In order to make full use of their advantages and overcome their disadvantages, hybrid methods were developed. Most of them are conducted based on two steps described as follows: the first step is to reduce the number of variables and remain the informative variables; the second step is to make further optimization for the variables remained by first step and obtain an optimal variable subset. Currently, there have been many hybrid methods reported, such as interval partial least squares-successive projections algorithm (iPLS-SPA) (Kong

et al. 2015), competitive adaptive reweighted sampling-SPA (CARS-SPA) (Tang et al. 2014), CARS-genetic algorithms-PLS (CARS-GA-PLS) (Xu et al. 2014), modified variable combination population analysis-GA (mVCPA-GA) (Yun et al. 2019a), mVCPA- iteratively retaining informative variables (mVCPA-IRIV) (Yun et al. 2019a), and Monte Carlo-uninformative variable elimination-SPA (MC-UVE-SPA) (Li et al. 2014). Additionally, four of hybrid methods, our newly proposed strategy, including iPLS-variable importance inprojection-GA (iPLS-VIP-GA) (Yu et al. 2020), iPLS-VIP-IRIV (Yu et al. 2020), iPLS-mVCPA-GA (Yu et al. 2020), and iPLS-mVCPA-IRIV (Yu et al. 2020), are implemented by three steps which are named rough selection, fine selection, and optimal selection. They have been proved that are more effective to distinctly increase the prediction capacity of the built model based on the selected variables than their single methods as well as the hybrid methods of each two methods.

Therefore, in this study, NIR spectral technique was applied to estimate its potential for prediction of TVB-N value of tilapia fillets. The proposed three-step hybrid methods were used to conduct optimal wavelength selection. The overall objective of this study is to investigate whether freshness of the fish flesh could be evaluated by a portable NIR spectrometer.

## Materials and Methods

The key steps of experimental operation and multivariate data analysis are shown in Fig. 1. Sample preparation, NIR spectra collection, and TVB-N content measurement were included in experimental operation, and outlier elimination, spectral pre-processing, and optimal wavelength selection included in multivariate data analysis. They are described clearly in the following sections.

### Sample Preparation

Forty fresh tilapia fillets, each of which was about 110 g, were purchased from Xiang Tai Fishery Company in Chengmai City of Hainan Province, China. All fillets were put in the foam box with ice at the bottom and then covered by ice again. The fillets were then transported to the laboratory immediately. At the same time, each fillet sample was segmented into three small fillet samples with similar sizes and a total of 120 fillets were gained, which were packaged in the sealing bag for cold storage at 4 °C. When all fillets were stored for 0, 2, 4, 6, 8, and 10 days, twenty fillets were selected randomly for spectra acquisition and TVB-N measurement.

### NIR Spectra Collection

Spectra were collected by a portable NIR spectrometer (BWS025, B&W Tek, USA) equipped with an integrating

sphere of 2 cm in diameter. The spectra sampled by a portable spectrometer have the range of 900–1700 nm, and its resolution is 1.6 nm. Thus, each spectrum has 495 wavelengths. Instrument control and spectra acquisition were achieved using BWSpec (v.4.11) software (B&W Tek, USA). The integration time was 4000  $\mu$ s, and the average number of spectra was 64. Two spectra were collected from opposite sides of each fillet and then averaged for further analysis. While all samples were scanned and averaged, one hundred and twenty spectra were obtained.

### TVB-N Measurement

TVB-N values were measured based on reference (Cai et al. 2011) with some modifications. Five grams tilapia flesh was weighted after the fillets were minced. And it was blended with 45 mL perchloric acid ( $\text{HClO}_4$ ) (0.6 mol/L) in a centrifuge tube and then centrifuged applying a 5840 R centrifuge (Eppendorf, German) at 3000 rpm for 10 min at 25 °C. Afterwards, the solution was filtered by a filter paper and glass funnel. Twenty-five milliliters of 30% sodium hydroxide (NaOH) was added in the filtrate to make it alkaline, and they were then distilled with 3 g magnesium oxide and 1.5 mL defoamer for 5 min employing a K9840 Kjeldahl Distillation Unit (Hanon Instruments, China), and the mixed sample solution was replaced by 45 mL of  $\text{HClO}_4$  and 25 mL of NaOH solutions as the blank control group. Fifty milliliters of 40 g/L boric acid ( $\text{H}_3\text{BO}_3$ ) was taken into a conical flask to absorb the distillate. The gained  $\text{H}_3\text{BO}_3$  solution was titrated with 0.01 mol/L hydrochloric acid (HCl). The TVB-N value was formulized as Eq. (1) illustrates and is expressed as milligrams per 100 g.

$$\text{TVB-N} = \frac{(V_t - V_b) \times c \times 14}{m} \times 100 \quad (1)$$

where  $V_t$  and  $V_b$  are the titration volume for the fillet sample (mL) and blank control group (mL), respectively.  $c$  represents the concentration of HCl (mol/L), and  $m$  represents the weighted mass of each tested tilapia flesh.

### Multivariate Data Analysis

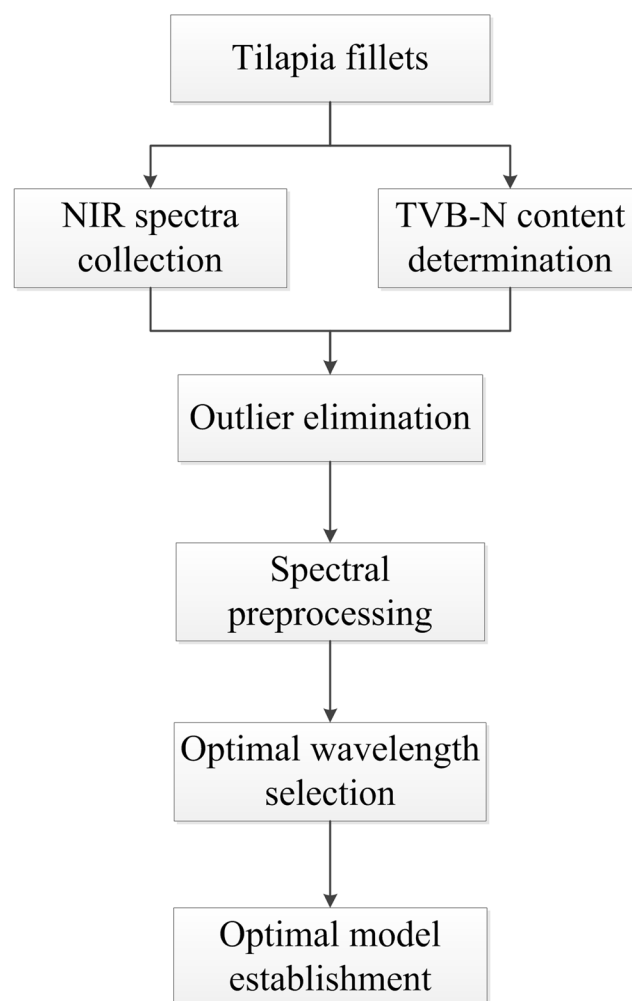
#### Outlier Elimination

To eliminate the effect of outliers' spectra on the prediction capability of those built models, the outlier elimination algorithm (Yun et al. 2015) based on Monte Carlo sampling method (MCS) was performed on all spectra to remove these outliers. The MCS method was implemented on the basis of the stages as follow: Firstly, the number of latent variables (nLVs) was calculated by cross-validation in PLS. Secondly, the whole data set was classified as the two parts (calibration set

and independent test set) by the MCS sampling method. Thirdly, the calibration set was employed for the building of the model based on the optimal nLVs. The independent test set was applied for prediction, and the prediction error of each test sample would be gained. Finally, the distribution of the prediction error for each sample would also be acquired, whose statistical parameters could be used for the observation of the outliers. Here, we employed two continuous steps to implement the operation of the outlier elimination. The latter step was employed to eliminate outliers from the spectral samples selected by the initial step. The remaining 111 spectra would be used for spectral preprocessing and optimal wavelength selection.

#### Spectral Preprocessing

NIR spectra were often affected by particle size, light scattering, and the difference, so some preprocessing methods such as standard normal variation (SNV), multiplicative scatter correction (MSC), first-order derivative (D1), smoothing (SM),



**Fig. 1** The key steps of experimental operation and multivariate data analysis

Savitzky-Golay smoothing (SG), SG-D1, SM-D1, SNV-D1, and MSC-D1 were used to weaken the effect of the noise and light scattering. SNV was to use a mathematical transformation method of the  $\log(1/R)$  for the elimination of slope variation and correction of scatter effects (Barnes et al. 1989). MSC was employed to correct scattered light caused by different particle sizes and also the additive as well as multiplicative effects in the spectra (Isaksson and Næs 1988). D1 aimed to reduce peak overlap and the drift not correlated with wavelengths (Savitzky and Golay 1964). SM and SG, as two different smoothing methods, were able to enhance the ratio of signal noise (Gorry 1990; Savitzky and Golay 1964). After comparing all methods, the optimal one with the best prediction ability would be applied in the stage of optimal wavelength selection.

### Optimal Wavelength Selection

The NIR spectra of fillet samples collected by conducting the portable NIR spectrometer with the range of 900–1700 nm could be regarded as high-dimensional data, which usually has very strong collinearity and many redundant and noisy variables making the models an unfavorable prediction performance as well as a large amount of computation time. Hence, it is necessary to decrease a large quantity of useless or unimportant wavelengths and select informative wavelengths that can reflect the changes of TVB-N content. Optimal wavelength selection is a good way to realize it. Therefore, some wavelength selection methods developed were employed to remove wavelengths being irrelevant with the TVB-N values, and some informative wavelengths then were remained and modeled. In this study, single methods such as iPLS, CARS, and VCPA and hybrid methods such as iPLS-VIP, iPLS-VCPA, VIP-GA (Yu et al. 2020), VIP-IRIV (Yu et al. 2020), mVCPA-GA, mVCPA-IRIV, iPLS-VIP-GA, iPLS-VIP-IRIV, iPLS-mVCPA-GA, and iPLS-mVCPA-IRIV were conducted to achieve wavelength selection. They are briefly described as follows.

iPLS is a widely used wavelength interval selection method with the merit of strong interpretability, which was conducted to segment the whole spectra into a number of intervals with equal size and establish a model for each interval. An overall picture of RMSECV of each wavelength interval is then obtained. Finally, we can choose the interval with the lowest RMSECV value or several intervals with lower RMSECV for further wavelength selection. Therefore, iPLS is a powerful tool for rapidly removing a large number of spectral redundancies and noises.

CARS is a widely used method, which is based on the principle of the “survival of the fittest” from Darwin’s Evolution Theory. When implementing CARS, MCS method is firstly conducted to generate a large number of variable subsets, then continuously shrink variable space by the

exponentially decreasing function (EDF) and following adaptive reweighted sampling (ARS), and in the end, the optimal subset with the lowest RMSECV will be selected. Therefore, CARS is an effective and simple method to rapidly realize the selection of an optimal subset.

VIP, which is expressed as the reference (Yu et al. 2020) shows, is to use a VIP value computed by PLS parameters for the evaluation of the importance of each variable. Meanwhile, VIP algorithm is a simple, rapid, and efficient method for determining the importance of each variable and then sorting them based on their VIP values. Calibration models can be built by using the variables selected on the basis of the predefined threshold of VIP value. More variables will be retained by VIP method to gain better model performance when it was combined with IRIV and GA.

VCPA is an approach to use EDF to continuously shrink original spectral space and get a small and optimizing spectral space, from which the optimal variable subset is searched. At each run, a large crowd of different wavelength combinations are produced by binary matrix sampling (BMS). These combinations are modeled by PLS, and then, the built models are assessed and sorted on the basis of RMSECV. The frequency of every wavelength in the best 10% of submodels will be calculated, and the variables with lower frequency will be eliminated. The whole step will be repeated, until only 14 variables remained. In order to solve this problem of fewer retained variables to remain more variable and utilize its advantage of continuously shrinking variable space for further optimization, a modified VCPA was proposed as the initial step of mVCPA-based hybrid methods and combined with GA and IRIV to make wavelength selection (Yun et al. 2019a).

GA algorithm was first proposed as a kind of global optimization method, to mimic the evolutionary course of a living species based on Darwin’s classical rules. There are four basic steps included in GA (Niazi and Leardi 2012): coding of variables, initiation of population, evaluation of the response, and reproduction. As more than 200 NIR spectral wavelengths are put into GA, it is effortless to result in over-fitting. Thus, GA algorithm was implemented optimizing variable space after the number of wavelengths had been decreased to lower than 200.

IRIV is used to assess the importance of every variable, classify them, and then remove unimportant variables. The importance of every variable was evaluated by calculating the difference of RMSECV values after every variable was included and excluded in all variable combinations generated by BMS, while the statuses (inclusion or exclusion) of other variables kept changeless. By this step, all variables will be divided by statistics analysis into four parts, such as strongly informative, weakly informative, uninformative, and interfering variables. The first two kinds of variables will be kept, and other kinds of variables removed. The elimination operation

of uninformative and interfering variables is conducted continuously until no uninformative and interfering variables exist. Thus, IRIV is very time-consuming, especially when meeting a high-dimensional data. Thus, it is suitable to further optimize variable subset in the final step after a large number of unimportant variables have been removed.

Three-step hybrid methods (iPLS-VIP-GA, iPLS-VIP-IRIV, iPLS-mVCPA-GA, and iPLS-mVCPA-IRIV) are to continuously reduce the number of variables and optimize variable space by three steps, including rough selection, fine selection, and optimal selection. Rough selection is to use wavelength interval selection methods (iPLS) with great interpretation capacity for roughly selecting variables and shrinking the spectral space. Fine selection is then to employ wavelength point selection methods (VIP and mVCPA) for continuously shrinking the variable space from large to small and retaining the very important variables. Optimal selection is to apply some optimization methods (IRIV and GA) for further optimizing variable subsets obtained by fine selection. Using these three steps can take full advantage of various concerned algorithms as well as remedy their drawbacks, while dealing with a large number of spectral wavelengths. Actually, three-step hybrid methods have been applied successfully on NIR spectral data in the reference (Yu et al. 2020).

#### Data Division

The spectral data was separated by duplex method (Snee 1977) into a calibration and independent test sets by the ratio of 2:1. Calibration set was applied for the calibration of the model while independent test set was applied for the evaluation of the performance of the model built by partial least square (PLS) regression (Geladi and Kowalski 1986) with calibration set.

#### Model Evaluation

Usually, the performance of the built model was assessed according to determination coefficients of calibration ( $R^2_c$ ), five-fold cross-validation ( $R^2_{cv}$ ) and prediction ( $R^2_p$ ), root mean square errors of calibration (RMSEC), five-fold cross-validation (RMSECV), and prediction (RMSEP). The larger the  $R^2_c$ ,  $R^2_{cv}$ , and  $R^2_p$  of the model, the higher the performance of the built model. On the contrary, the lower the RMSEC, RMSECV, and RMSEP of the model, the higher accuracy the built model. By the way, the effectiveness of outlier sample elimination, spectral preprocessing, and optimal wavelength selection was also evaluated by these assessment criteria, when all remained and preprocessed spectra data were modeled and predicted by PLS, respectively.

#### Software

All algorithms of outlier elimination, spectral pretreatment, and optimal wavelength selection were carried out in MATLAB 2013a (the MathWorks, Inc.) based on a common laptop with intel(R) Core (TM) i5-4258U CPU at 2.4 GHz and 4 GB RAM, with the operating system of Window 7. The MATLAB codes of CARS and VIP are from the website of <http://www.libpls.net/>. The MATLAB source codes of iPLS and GA could be gained freely from the websites <http://www.models.life.ku.dk/iToolbox> and <http://www.models.life.ku.dk/GAPLS>, respectively. In addition, the MATLAB source codes of VCPA, IRIV, VCPA-GA, and VCPA-IRIV can be freely downloaded from the website: <https://cn.mathworks.com/matlabcentral/profile/authors/5526470-yonghuan-yun>.

## Results and Discussion

### Original Data Analysis

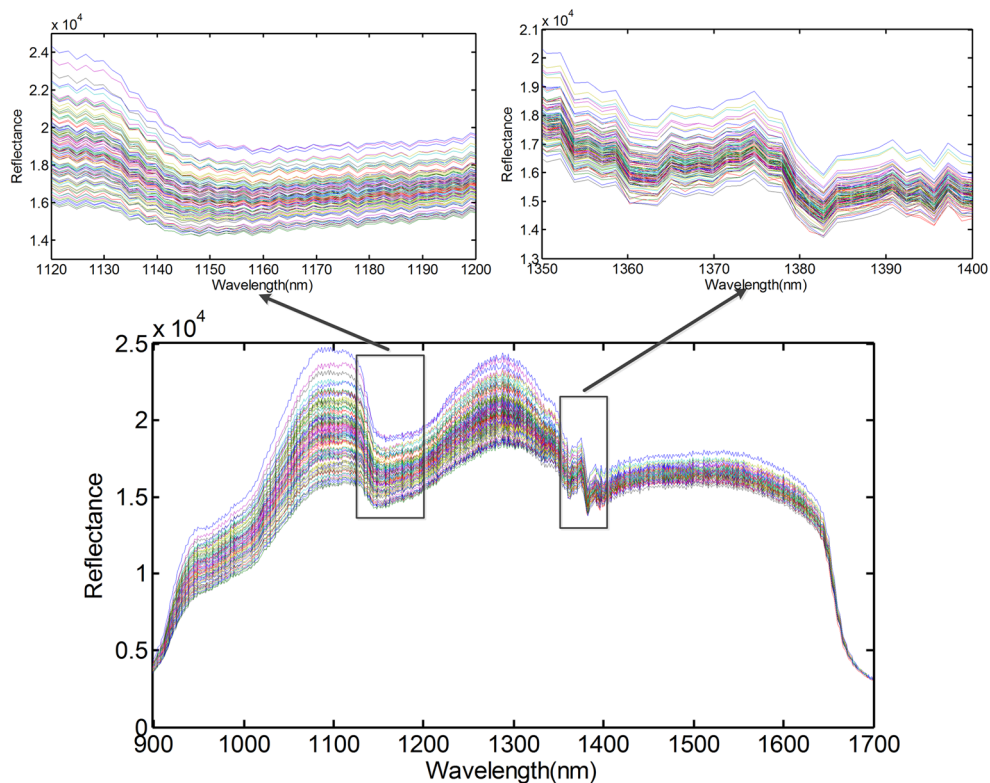
Figure 2 shows all original spectra extracted from the tilapia fillets within the wavelength range of 900–1700 nm. As shown from Fig. 2, all spectra have a similar tendency, but some differences still existed in the variation of spectral reflectance strength. It was probably because of the changes of major chemical components causing the freshness loss of tilapia fillets during storage at 4 °C. The absorbance peak at the range of 1120–1200 nm belongs to the second overtone of C-H and N-H bonds (Tito et al. 2012); the broad region from 1350 to 1400 nm is assigned to the combination of C-H (Sinelli et al. 2010).

The variation trend of TVB-N content measured from day 0 to day 10 is shown in Fig. 3. The changes from day 0 to day 10 were similar to (Özogul et al. 2005), which are stable from day 0 to day 8 and ascending sharply from day 8 to day 10. According to Chinese standard GB 2733-2015, the upper acceptability limit of TVB-N content is 30 mg/100 g for marine fish and 20 mg/100 g for freshwater fish. Therefore, it can be seen from Fig. 3 that the longest period when the tilapia fillet keeps fresh stored at 4 °C is around 9 days, which is because the mean of the TVB-N value was more than 30 mg/100 g at day 10 and less than 20 mg/100 g at day 8. This means that the fillets undergo the process from fresh to spoilage, which is appropriate and capable for the study of fish freshness change.

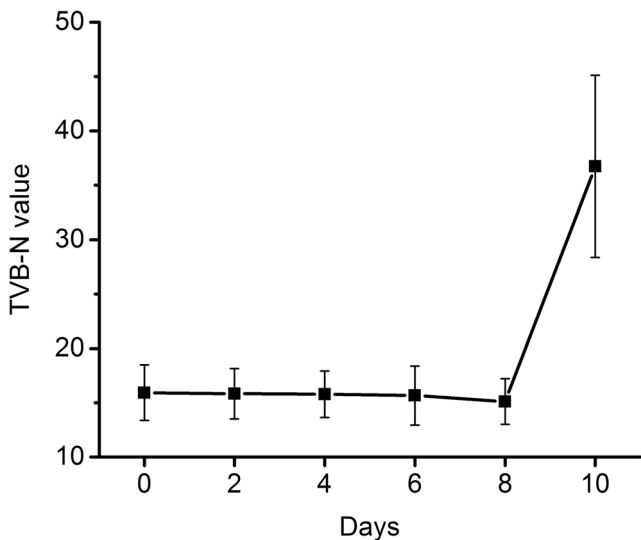
### Outlier Elimination Analysis

Outlier elimination is an important step in the multivariate calibration to remove several combinations of the spectrum and corresponding TVB-N value. The two continuous steps of outlier elimination are displayed in Fig. 4, and the samples, away from the main body of samples, are seen as outlier

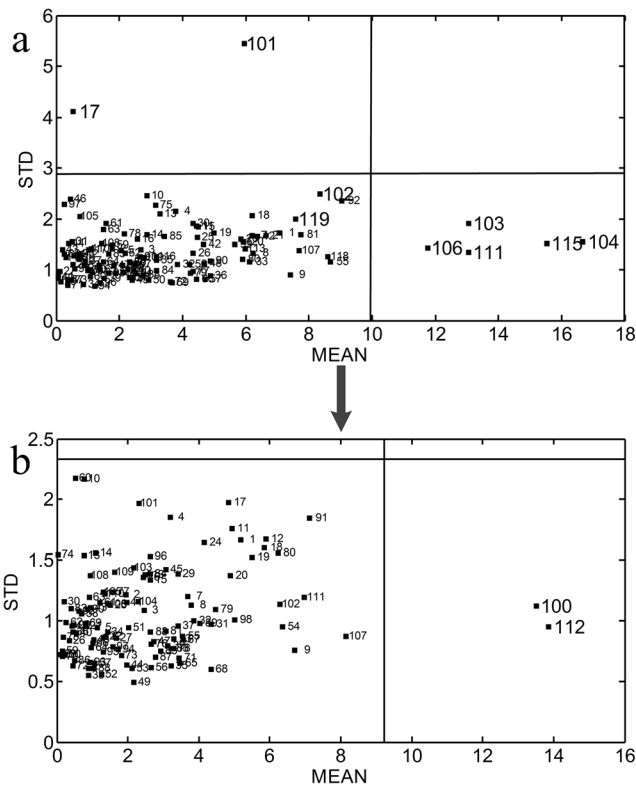
**Fig. 2** All original spectra extracted from the tilapia fillets within the wavelength



samples and will be eliminated. In the first step, it is obviously illustrated that seven samples containing 17, 101, 103, 104, 106, 111, and 115 were removed. Because among these samples, 17 and 101 were away from the principal body of samples in the *y* direction while other samples such as 103, 104, 106, 111, and 115 were away from the principal body of samples in the *x* direction. Then two samples consisting of 100 and 112 (102 and 119 in Fig. 4a, respectively), which were far from the main part of samples in the *x* direction, were removed in the second step from the samples remained by the



**Fig. 3** Changes of TVB-N values measured from day 0 to day 10



**Fig. 4** The results of the variance of residuals versus the mean of residuals for the TVB-N content of tilapia fillets. **a** The first step of MCS. **b** The second step of MCS

first step. Finally, the remaining 111 samples will be applied for further analysis.

To evaluate the effectivity of outlier elimination, PLS was employed to establish models at each step. The sample number of calibration and independent test sets and their statistic results of measured TVB-N values are illustrated in Table 1. And the evaluation results of PLS models after outlier elimination from original data to step 2 are given in Table 2. It was clearly shown in Table 2 that the operation of outlier elimination from original data (no operation of outlier elimination) to step 2 has obvious advantages to improve the accuracy and robustness of models after outlier samples were removed. This is because all determination coefficients containing  $R^2c$ ,  $R^2cv$ , and  $R^2p$  have risen by 0.0955, 0.1514, and 0.1414, respectively, from original data to step 2. Moreover, RMSEC (from 3.9564 to 2.2948), RMSECV (from 5.2212 to 3.1466), and RMSEP (from 5.9258 to 3.5737) also continuously reduced during the process, which means that the established PLS models become more and more accurate and robust after outliers were removed. These results demonstrated that outlier elimination is a necessary and effective step to remove outlier samples and develop the prediction capacity of the models built using remained samples.

### NIR Spectral Preprocessing Analysis

In order to obtain a more accurate and reliable calibration model, it is very critical to conduct preprocessing methods to preprocess NIR spectral data for the removal of the interference of background and noises. Five preprocessing methods and their hybrid methods are executed to preprocess the spectra. And their evaluation results are shown in Table 3. From Table 3, we can see that SM-D1 is the best preprocessing method in this study because of the best models established by it with the highest  $R^2p$  and lowest RMSEP. Thus, the spectra data processed by SM-D1 will be used for further analysis. It has been reported by Zhou et al. (2019) that SNV preprocessing method as the best NIR spectral pretreatment was used for the optimal wavelength selection analysis by CARS in the freshness prediction of bighead carp. However, the reference (Masoum et al. 2012) illustrated that D1 predicted more accurately than other preprocessing

techniques when used for the determination of crude protein in fishmeal, while the best prediction performance was obtained to predict moisture content with no one used.

### Optimal Wavelength Selection Analysis

NIR spectra produced by NIR spectrometers, especially the portable NIR spectrometer, usually create unwanted spectral information containing the redundancy and noise. These drawbacks not only increased the complexity and time of calibration but also decreased the accuracy and robustness of the model. In order to eliminate the spectral redundancy and noise, decrease the number of wavelengths, and enhance the efficiency of modeling, it is particularly important to achieve optimal wavelength selection. Here, the newly proposed strategy, three-step hybrid methods (iPLS-VIP-GA, iPLS-VIP-IRIV, iPLS-mVCPA-GA and iPLS-mVCPA-IRIV), two-step hybrid methods (iPLS-VIP, iPLS-VCPA, VIP-GA, VIP-IRIV, mVCPA-GA and mVCPA-IRIV), and three frequently used single methods (iPLS, CARS and VCPA) were conducted to select the important as well as informative wavelengths correlated with the measured freshness indicator from the full spectrum. Among these methods, all three-step hybrid methods have been testified that they are more effective than other listed methods to be employed for the variable selection of NIR spectral data in the reference (Yu et al. 2020).

The results of wavelength selection applying the given variable selection methods for the prediction of tilapia fillets' freshness are shown in Table 4. From Table 4, the values of  $R^2c$ ,  $R^2cv$ ,  $R^2p$ , RMSEC, RMSECV, and RMSEP were 0.9139, 0.6913, 0.8871, 1.8951, 3.5891, and 2.4851, respectively, based on the built full-spectrum PLS model. These are slightly worse than the results published in (Cheng et al. 2014), which showed that  $R^2c$ ,  $R^2cv$ , and  $R^2p$  are 0.927, 0.913, and 0.905, respectively, for the TVB-N content of grass carp fillet applying PLS regression by the VIS-NIR hyperspectral imaging. But the  $R^2p$  of 0.8871 for predicting TVB-N content was higher than the  $R_p$  of 0.6780 determined in pork and the  $R_p$  of 0.678 in bighead carp based on the references of Cai et al. (2011) and Zhou et al. (2019), respectively. As seen from Table 4, in comparison with full-spectrum PLS model, most wavelength selection methods

**Table 1** Statistic results of TVB-N content in the calibration and independent test sets

Outlier elimination	Subsets	Sample number	Range	Mean	Standard deviation
Original data	Calibration	80	10.29–51.34	19.13	8.85
	Independent test	40	12.43–51.35	19.29	8.89
Step 1	Calibration	76	10.29–44.85	18.09	7.29
	Independent test	37	12.43–44.11	18.48	7.42
Step 2	Calibration	74	10.28–44.85	17.79	7.07
	Independent test	37	11.62–36.64	18.07	6.47

**Table 2** The evaluation results of PLS models after outlier elimination from original data to step 2

Outlier elimination	R <sup>2</sup> c	RMSEC	R <sup>2</sup> cv	RMSECV	R <sup>2</sup> p	RMSEP
Original data	0.7978	3.9564	0.6479	5.2212	0.5445	5.9258
Step 1	0.8592	2.7165	0.7513	3.61	0.606	4.5917
Step 2	0.8933	2.2948	0.7993	3.1466	0.6859	3.5737

obtain better results as a result of their lower RMSEP and higher R<sup>2</sup>p values, which indicated that wavelength selection is an important step to develop the prediction performance of the models built with selected wavelengths. Although the R<sup>2</sup>p values of iPLS, iPLS-VIP, iPLS-VCPA, VCPA, VIP-GA, and mVCPA-IRIV were mildly lower than that of the full-spectrum model, they decreased the number of spectral wavelengths to a large extent, simplified the models, and shortened the computation time. It is important to note that the differences of R<sup>2</sup>p and RMSEP between CARS and PLS were 0.0267 and 0.313, respectively, which is consistent with Zhou et al. (2019) who revealed that the differences R<sub>p</sub> and RMSEP between CARS and PLS were 0.255 and 2.834, respectively.

Most importantly, the overall better results were obtained by all three-step hybrid methods because R<sup>2</sup>p values of these hybrid methods (iPLS-VIP-GA, iPLS-VIP-IRIV, iPLS-mVCPA-GA, and iPLS-mVCPA-IRIV) are 0.0088, 0.0296, 0.0019, and 0.033 higher than that of PLS, respectively. It illustrated that three-step hybrid strategy is useful and effective for the prediction of TVB-N to select optimal wavelengths, increase the accuracy of models, and optimize NIR spectral space. The four combinations of three-step hybrid strategy present better prediction results than their corresponding hybrid methods based on two steps because of their higher R<sup>2</sup>p and lower RMSEP, which was consistent with the results of the reference (Yu et al. 2020). Moreover, iPLS-mVCPA-IRIV got the best prediction results because the values of R<sup>2</sup>p and RMSEP were 0.9201 and 2.0907, respectively. The plot that showed the fitness of the iPLS-mVCPA-IRIV model is

put in Fig. 5. Therefore, these results demonstrated that developed variable selection methods were effective and reliably employed for the evaluation of fish freshness.

In order to interpret why three-step hybrid methods got the overall better predicted performance of the built models, the distribution of the selected wavelengths with the four methods (iPLS-VIP-GA, iPLS-VIP-IRIV, iPLS-mVCPA-GA, and iPLS-mVCPA-IRIV) is given in Fig. 6. From Fig. 6, we can see that all four methods have some common selected regions. Since iPLS-mVCPA-IRIV performed the best, the interpretation of the selected wavelengths would be focused on it. The wavelengths selected by iPLS-mVCPA-IRIV are principally centralized in the regions of 1014–1058 nm, 1188–1227 nm, 1336–1424 nm, and 1525–1692 nm. The absorption region of 1014–1058 nm can be ascribed to the protein functional group (Lin et al. 2006), and the absorption at 1188–1227 nm could be connected with the second overtone of C-H and N-H bonds (Tito et al. 2012). The absorption at 1336–1424 nm is assigned to the combination of C-H band (Sinelli et al. 2010), while the absorption at 1525–1692 nm is related to the first overtone of C-H stretching (Workman Jr and Weyer 2007).

## Conclusions

In this work, the portable NIR spectrometer and chemical analysis were combined with chemometrics to measure the TVB-N content of tilapia fillets for assessment of fish freshness. To obtain better prediction results of the models built

**Table 3** Comparison of the prediction performance of PLS regression models for the prediction of TVB-N content from the spectral data with different preprocessing methods for the full range

Pretreatment	R <sup>2</sup> c	RMSEC	R <sup>2</sup> cv	RMSECV	R <sup>2</sup> p	RMSEP
No pretreatment	0.8933	2.2948	0.7993	3.1466	0.6859	3.5737
D1	0.9337	1.6855	0.8106	2.8487	0.7647	3.5427
SNV	0.8706	2.6441	0.6678	4.2364	0.7058	3.0138
MSC	0.8636	2.7529	0.7512	3.7182	0.5304	3.5671
SG(5) <sup>a</sup>	0.9395	1.6278	0.7893	3.0377	0.8097	3.1362
SM(3) <sup>b</sup>	0.8487	2.7320	0.7762	3.3233	0.5667	4.1978
SNV-D1	0.9225	1.6493	0.7626	2.8865	0.8243	3.4464
MSC-D1	0.9364	1.5065	0.7224	3.1468	0.8724	2.9315
SG(5) <sup>a</sup> -D1	0.9508	1.3053	0.7481	2.9526	0.8148	3.5884
SM(3) <sup>b</sup> -D1	0.9139	1.8951	0.6913	3.5891	0.8871	2.4851

<sup>a</sup> 5 is the spectral window size in the preprocessing method of Savitzky-Golay smoothing

<sup>b</sup> 3 is the spectral window size in the preprocessing method of smoothing



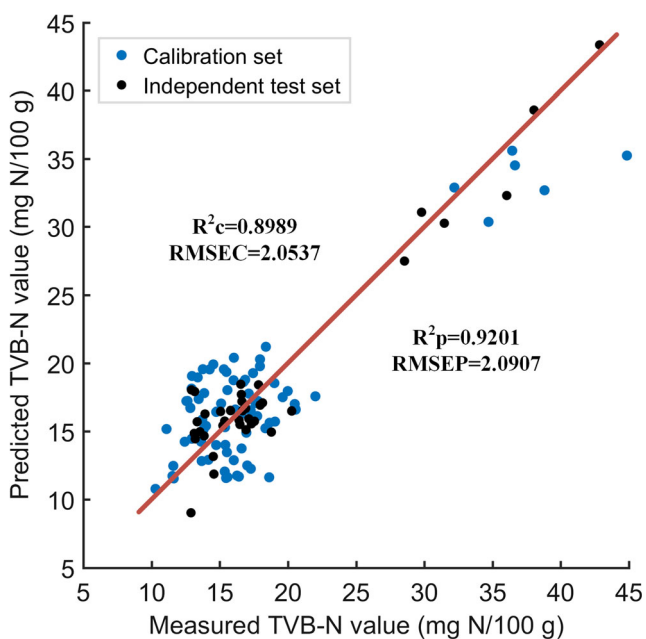
**Table 4** Calibration, cross-validation, and prediction results of the TVB-N values of tilapia fillets by NIR spectral technique

Combination	nVAR <sup>a</sup>	nLVs <sup>b</sup>	Calibration		Cross-validation		Prediction	
			R <sup>2</sup> <sub>c</sub>	RMSEC	R <sup>2</sup> <sub>cv</sub>	RMSECV	R <sup>2</sup> <sub>p</sub>	RMSEP
PLS	495	10	0.9139	1.8951	0.6913	3.5891	0.8871	2.4851
iPLS	198	9	0.8770	2.2655	0.7481	2.9166	0.8445	3.2422
CARS	24	8	0.9264	1.7529	0.8721	2.3098	0.9138	2.1721
VCPA	11	6	0.8670	2.3555	0.8287	2.6732	0.8785	2.5780
VIP-GA	42	9	0.8779	2.2570	0.7752	3.0625	0.8744	2.6215
VIP-IRIV	21	10	0.8732	2.2998	0.8049	2.8529	0.8920	2.4309
iPLS-VIP	168	9	0.8750	2.2835	0.7536	4.7755	0.8430	2.9308
iPLS-VCPA	11	9	0.8780	2.2564	0.8338	2.6332	0.8777	2.5872
mVCPA-GA	78	10	0.9118	1.9182	0.8178	2.7570	0.8887	2.4672
mVCPA-IRIV	52	10	0.9287	1.7250	0.8869	2.1726	0.8703	2.6643
iPLS-VIP-GA	73	9	0.9044	1.9974	0.8118	2.8019	0.8959	2.3864
iPLS-VIP-IRIV	35	8	0.9074	1.9657	0.8659	2.3652	0.9167	2.1351
iPLS-mVCPA-GA	53	10	0.8994	2.0486	0.8151	2.7773	0.8890	2.4641
iPLS-mVCPA-IRIV	33	8	0.8989	2.0537	0.8538	2.4701	0.9201	2.0907

<sup>a</sup> nVAR is the number of selected variables by the listed wavelength selection methods

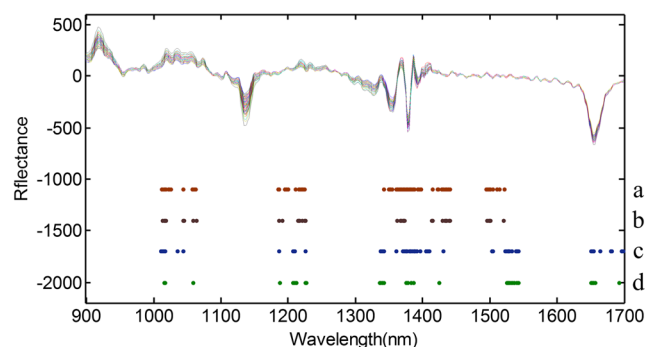
<sup>b</sup> nLVs is the number of latent variables in PLS

based on PLS regression, NIR spectral data were processed by the steps of outlier elimination, spectral preprocessing, and optimal wavelength selection. First, outlier spectra were eliminated by continuously implementing the outlier elimination method two times, and then the PLS regression model, established utilizing the remained samples at each step, had a significantly improved prediction performance based on the ascending determination coefficients ( $R^2$ ) and decreasing root



**Fig. 5** Predicted and measured TVB-N values after variable selection using iPLS-mVCPA-IRIV

mean square errors (RMSE). Second, the operation of pre-treatment for original spectra was conducted by some preprocessing algorithms to clear background information, light scattering, and noises. The results indicated, in this study, that SM-D1 was the most suitable preprocessing method as a result of the highest  $R^2_p$  and lowest RMSEP. Final, optimal wavelength selection was carried out applying several variable selection methods produced to establish more robust and simple models. It is illustrated that four methods of three-step hybrid strategy gained overall preferable results, and the combination of iPLS-mVCPA-IRIV was also seen as the most reliable and outstanding method based on the  $R^2_p$  of 0.9201 and RMSEP of 2.0907. Therefore, it can be said that the portable NIR spectrometer was qualified and powerful for fast and nondestructive evaluation of tilapia freshness.



**Fig. 6** The distribution of remained variables by three-step hybrid methods in the preprocessed spectra. **a** iPLS-VIP-GA. **b** iPLS-VIP-IRIV. **c** iPLS-mVCPA-GA. **d** iPLS-mVCPA-IRIV

**Funding Information** This work is funded by the National Natural Science Foundation of China (grants no. 21705162), the Program of Hainan Association for Science and Technology Plans to Youth R & D Innovation (QCXM201924) and the Scientific Research Funds of Hainan University, China (KYQD(ZR)-1844).

## Compliance with Ethical Standards

**Conflict of Interest** Hai-Dong Yu declares that he has no conflict of interest. Si-Min Zuo declares that she has no conflict of interest. Guanghua Xia declares that he has no conflict of interest. Xing Liu declares that he has no conflict of interest. Yong-Huan Yun declares that he has no conflict of interest. Chenghui Zhang declares that he has no conflict of interest.

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

**Informed Consent** Informed consent not applicable.

## References

- Barnes R, Dhanoa MS, Lister SJ (1989) Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl Spectrosc* 43:772–777
- Cai J, Chen Q, Wan X, Zhao J (2011) Determination of total volatile basic nitrogen (TVB-N) content and Warner–Bratzler shear force (WBSF) in pork using Fourier transform near infrared (FT-NIR) spectroscopy. *Food Chem* 126:1354–1360. <https://doi.org/10.1016/j.foodchem.2010.11.098>
- Chen H, Lin Z, Tan CJMJ (2019) Classification of different animal fibers by near infrared spectroscopy and chemometric models. 144:489–494
- Cheng J-H, Sun D-W (2015) Rapid quantification analysis and visualization of *Escherichia coli* loads in grass carp fish flesh by hyperspectral imaging method. *Food Bioprocess Technol* 8:951–959
- Cheng J-H, Sun D-W, Zeng X-A, Pu H-B (2014) Non-destructive and rapid determination of TVB-N content for freshness evaluation of grass carp (*Ctenopharyngodon idella*) by hyperspectral imaging. *Innov Food Sci Emerg Technol* 21:179–187
- Duan Z-h, Jiang L-n, Wang J-l, Yu X-y, Wang T (2011) Drying and quality characteristics of tilapia fish fillets dried with hot air-microwave heating. *Food Bioprod Process* 89:472–476
- FAO (2017) Global aquaculture production. <http://www.fao.org/fishery/statistics/global-production/en>. Accessed Feb 2020
- Geladi P, Kowalski BR (1986) An example of 2-block predictive partial least-squares regression with simulated data. *Anal Chim Acta* 185:19–32
- Gorry PA (1990) General least-squares smoothing and differentiation by the convolution (Savitzky-Golay) method. *Anal Chem* 62:570–573
- Green DP, Alasalvar C, Shahidi F, Miyashita K, Wanasundara U (2011) Sensory evaluation of fish freshness and eating qualities. *Handb Seaf Qual Saf Health Appl*:29–38
- Guan Z, Wang X, Li M, Jiang X (2013) Mathematical modeling on hot air drying of thin layer fresh tilapia fillets. *Polish J Food Nutr Sci* 63:25–33
- Isaksson T, Næs T (1988) The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy. *Appl Spectrosc* 42:1273–1284
- Kong Q, Su Z, Shen W, Zhang B, Wang J, Ji N, Ge H (2015) Research of straw biomass based on NIR by wavelength selection of IPLS-SPA. *Guang pu xue yu guang pu fen xi= Guang pu* 35:1233–1238
- Li J, Zhao C, Huang W, Zhang C, Peng Y (2014) A combination algorithm for variable selection to determine soluble solid content and firmness of pears. *Anal Methods* 6:2170–2180
- Li X et al (2017) Freshness assessment of turbot (*Scophthalmus maximus*) by quality index method (QIM), biochemical, and proteomic methods. 78:172–180
- Li X et al. (2020) Review of NIR spectroscopy methods for nondestructive quality analysis of oilseeds and edible oils
- Lin M, Mousavi M, Al-Holy M, Cavinato AG, Rasco BA (2006) Rapid near infrared spectroscopic method for the detection of spoilage in rainbow trout (*Oncorhynchus mykiss*) fillet. *J Food Sci* 71:S18–S23
- Liu Y, Ma D-h, Wang X-c, Liu L-p, Fan Y-x, Cao J-x (2015) Prediction of chemical composition and geographical origin traceability of Chinese export tilapia fillets products by near infrared reflectance spectroscopy. *LWT-Food Sci Technol* 60:1214–1218
- Masoum S, Alishahi AR, Farahmand H, Shekarchi M, Prieto N (2012) Determination of protein and moisture in fishmeal by near-infrared reflectance spectroscopy and multivariate regression based on partial least squares. *Iran J Chem Chem Eng* 31:51–59
- Niazi A, Leardi R (2012) Genetic algorithms in chemometrics. *J Chemom* 26:345–351
- Ocaño-Higuera V et al (2011) Freshness assessment of ray fish stored in ice by biochemical, chemical and physical methods. 125:49–54
- Özogul Y, Özyurt G, Özogul F, Kuley E, Polat A (2005) Freshness assessment of European eel (*Anguilla anguilla*) by sensory, chemical and microbiological methods. *Food Chem* 92:745–751
- Sarraguça MC, Cruz AV, Soares SO, Amaral HR, Costa PC, Lopes JA (2010) Determination of flow properties of pharmaceutical powders by near infrared spectroscopy. *J Pharm Biomed Anal* 52:484–492
- Savitzky A, Golay MJ (1964) Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem* 36:1627–1639
- Sinelli N, Casale M, Di Egidio V, Oliveri P, Bassi D, Tura D, Casiraghi E (2010) Varietal discrimination of extra virgin olive oils by near and mid infrared spectroscopy. *Food Res Int* 43:2126–2131
- Snee RD (1977) Validation of regression models: methods and examples. *Technometrics* 19:415–428
- Song Y, Luo Y, You J, Shen H, Hu S (2012) Biochemical, sensory and microbiological attributes of bream (*Megalobrama amblycephala*) during partial freezing and chilled storage. *J Sci Food Agric* 92:197–202
- Tang G et al (2014) A new spectral variable selection pattern using competitive adaptive reweighted sampling combined with successive projections algorithm. *Analyst* 139:4894–4902
- Tito NB, Rodemann T, Powell S (2012) Use of near infrared spectroscopy to predict microbial numbers on Atlantic salmon. *Food Microbiol* 32:431–436
- Uddin M, Okazaki E, Fukushima H, Turza S, Yumiko Y, Fukuda Y (2006) Nondestructive determination of water and protein in surimi by near-infrared spectroscopy. *Food Chem* 96:491–495
- Workman J Jr, Weyer L (2007) Introduction to near-infrared spectra. Practical guide to interpretive near-infrared spectroscopy. CRC Press, London
- 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 Chem* 145:417–426
- Xu F, Huang X, Dai H, Chen W, Ding R, Teye E (2014) Nondestructive determination of bamboo shoots lignification using FT-NIR with efficient variables selection algorithms. *Anal Methods* 6:1090–1095
- Yu H-D et al (2020) Three-step hybrid strategy towards efficiently selecting variables in multivariate calibration of near-infrared spectra. *Spectrochim Acta A Mol Biomol Spectrosc* 224:117376
- Yun Y-H, Wei Y-C, Zhao X-B, Wu W-J, Liang Y-Z, Lu H-M (2015) A green method for the quantification of polysaccharides in *Dendrobium officinale* RSC Adv 5:105057–105065

- Yun Y-H, Bin J, Liu D-L, Xu L, Yan T-L, Cao D-S, Xu Q-S (2019a) A hybrid variable selection strategy based on continuous shrinkage of variable space in multivariate calibration. *Anal Chim Acta* 1058:58–69
- Yun Y-H, Li H-D, Deng B-C, Cao D-S (2019b) An overview of variable selection methods in multivariate analysis of near-infrared spectra. *TrAC Trends Anal Chem*
- Zhou J, Wu X, Chen Z, You J, Xiong S (2019) Evaluation of freshness in freshwater fish based on near infrared reflectance spectroscopy and chemometrics *LWT* 106:145–150

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.