

REVIEW

Factors influencing near infrared spectroscopy analysis of agro-products: a review

Xiao XU, Lijuan XIE, Yibin YING (✉)

College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China

Abstract The near infrared (NIR) spectroscopy technique has wide applications in agriculture with the advantages of being nondestructive, sensitive, safe and rapid. However, there are still more than 40 error sources influencing the robustness and accuracy of its calibration and operation. Environmental, sample and instrument factors that influence the analysis are discussed in this review, including temperature, humidity and other factors that introduce uncertainty. Error sources from livestock products, fruit and vegetables, which are the most common objects in the field of NIR analysis, are also emphasized in the second part. In addition, studies utilizing different instruments, spectral pretreatments, variable selection methods, wavelength ranges, detection modes and calibration methods are tabulated to illustrate the complications they introduce and how they influence NIR analysis. It is suggested that large scale of data with abundant varieties can be used to build a more robust calibration model, in order to improve the robustness and accuracy of the NIR analytical model, and overcome problems caused by confining analysis to too many uniform samples.

Keywords agro-product, error source, influence factor, near infrared spectroscopy

1 Introduction

The near infrared (NIR) spectroscopy technique is a powerful tool in the field of quality assessment for agro-products, because it has the advantages of being nondestructive, sensitive, safe and rapid. The first sector of commerce to give it credibility was its application in the grain industry^[1]. Now, livestock products^[2–6], fruit and vegetables^[7–15], marine products^[16] are gradually being evaluated using this technique. The analytical process is illustrated by an example in Fig. 1, including steps of NIR

spectra acquisition, reference measurement as well as multivariate calibration and prediction^[7]. Beside determination of multivariate attributes of objects, discrimination of objects from different seasons, cultivars, origins and varieties have also been successfully studied^[17–21]. With the growing attention paid to agro-product security, the NIR spectroscopy technique also has a broad application in identification of adulteration in the food industry^[22]. Since research in the laboratory have been carried out for a long time and is already yielded many satisfying results, a growing range of portable and in-field NIR equipment has come into commercial use in recent years^[23–29]. The ability and feasibility of NIR quality assessment have been improved by optimizing the performance of prediction model.

Although the NIR spectroscopy technique involves reasonably simple steps to analyze or predict the quality of products, there are still more than 40 possible error sources in the whole process of analysis^[30], which will affect the results in various ways. These error sources have been classified here into three types: errors from the external environment, the sample and the instrument. Environmental factors affect both the instrument and the sample, which consequently influences the entire analytical process and results. The sample is considered as the most significant influencing factor in the use of NIR spectroscopy, which can influence the results through the process of choosing representative samples, operations on samples^[30] and biological variations between samples^[18,19]. For different kinds of agro-products, there are various error sources related to the sample. Indeed, with different types of instruments, the accuracy and robustness of the model will differ even though measurement parameters are kept same. It is also important to ensure that the influence of the instrument can be effectively controlled or avoided.

There have been few reviews systematically examining how these factors affect the NIR analysis and how to compensate for them in practical applications. In this paper, recent studies on error source and their influence on the NIR spectroscopy analysis are reviewed, the

Received November 30, 2017; accepted January 12, 2019

Correspondence: ibeying@zju.edu.cn

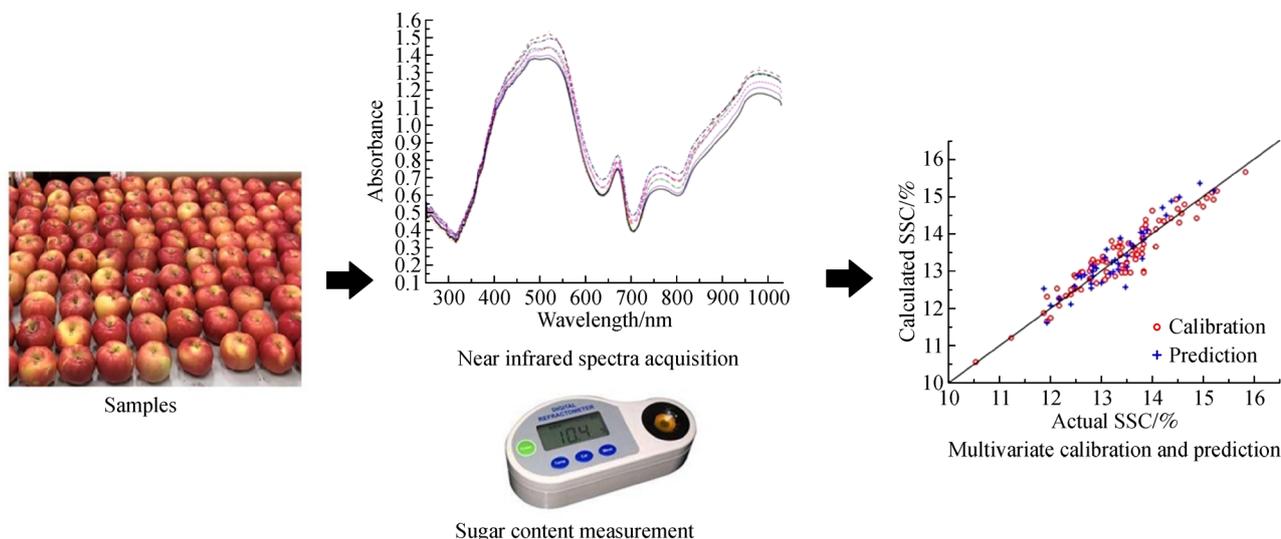


Fig. 1 Schematic diagram of experimental procedures using NIR spectroscopy for sugar content analysis

corresponding solutions and applications are discussed and solutions and recommendations made for solving these problems on the basis of this analysis.

2 Environmental factors

Higher feasibility and efficiency for measurement of agro-products with the NIR spectroscopy technique could be achieved, with only slight reduction in accuracy, if the robustness of calibration could be improved, especially for on-site applications. However, in the application of NIR spectroscopy it is difficult to keep all measurement conditions and parameters constant and the robustness of the calibration model will be influenced by various external factors, such as temperature, humidity and other uncertain variables. In this section, the effects of these environmental factors on NIR spectra and analysis are explained. In addition, some suggestions for compensating for these effects are proposed and compared.

2.1 Temperature

Unavoidable temperature fluctuation may occur because of varying weather conditions or improper conditioning of agro-products after harvest and these can have a crucial influence on NIR analysis. Ambient temperature is important in the whole process of NIR measurement and analysis by affecting the temperature of the instrument, the sample or a sub sample^[30]. From a spectral perspective, the wavelength that is most affected by temperature is the moisture-dominant part. The absorbance value in NIR spectra changes because vibration and rotation of O–H in water molecular are changed due to temperature fluctuations^[30,31]. Thus, objects with higher moisture content will suffer more from temperature fluctuations.

Factors that have an important influence on the instrument, such as temperature fluctuation, can affect both lamp output (intensity and quality) and detector response^[32] (wavelength sensitivity), as well as electronic noise^[33,34]. Any drift of the designed wavelength range in the instrument will be followed by a change to the regression coefficients of the calibration model. Martinsen et al.^[35] studied the effect of drift in wavelength calibration. Although the drift was small (< 0.1 nm) over a 12-month period, it was noted that a partial least squares regression (PLSR) model based on short-wavelength NIR could be very sensitive to such fluctuations, e.g., a kiwifruit dry matter (DM) model showed a bias shift of 7% DM for a 1 nm perturbation. However, introducing the small wavelength perturbation into the training set proved to be an effective way to reduce this error to less than 1.7% per nm. When operating an NIR instrument in different places in different ambient temperatures, changes in temperature will lead to spectral variation. Hayes et al.^[34] monitored the spectral variation of quartz halogen lamps due to temperature fluctuations. The lamp output was spectrally stable for the time of the first measurement (10 s), although total illumination was not stable until approximately 40 min from start-up. However, when predictive models were developed using second derivative absorbance data, there was no statistically significant impact in root mean square error of prediction (RMSEP) on time of lamp warm-up (after 10 s). In practical applications of the NIR spectroscopy technique, up to 50°C temperature variation can occur, where the previous reference measurement is no longer suitable for the current conditions. Thus, it is necessary to operate in the same conditions where a reference measurement has been taken in order to get high-accuracy of analytical results. Hence, explanations of the procedure and frequency to conduct reference measurements should be given to ensure and

maintain the optimal performance of an NIR measurement system. However, this also limits the versatility of a specific NIR measurement instrument and can be inconvenient for the measurement process.

There have been a large number of studies on the influence of temperature on the NIR spectra of samples and the research on wheat conducted by Williams et al.^[36] provides a representative example. It was shown that when varying sample temperature from -10 to 45°C , the read protein was significantly linearly reduced due to the sensitivity to temperature of the effective wavelength range in the calibration model. The analysis also verified that this reduction was not caused by changes in the moisture content that occurred during grinding but was strictly a result of temperature change, and was totally independent of the algorithm. Yao et al.^[37] collected and analyzed the spectra of watermelon juice sample at nine temperatures, varying from 0 to 40°C at intervals of 5°C to predict soluble solid content (SSC). The result indicated that there was an influence of temperature on NIR spectra in a nonlinear way. The average absorbance shifted with the change of temperature, is shown in Fig. 2.

There have been many studies on compensating for the effect of temperature in calibration models^[38]. For primary applications, Williams et al.^[36] proposed several methods to compensate, including choosing a wavelength insensitive to temperature for calibration, adjusting sample temperature to the calibration temperature and construction of a correction chart. Of these, the third can be considered as the most effective and direct way to solve this problem, although it suffers from a lack of automation and convenience. It is more satisfactory and convenient to utilize mathematical algorithms for compensating this bias automatically. Yao et al.^[37] used discriminant analysis and PLSR to establish calibration models for nine different temperatures. It turned out that the PLSR model established at 20°C performed better than at any other

temperature. The result also indicated that the local model was sensitive to fluctuations in temperature. Additionally, the global model showed useful prediction ability, which made temperature fluctuation a negligible interference. Besides, there are many other studies^[37–43] that have proved that the global model can significantly reduce the effect of temperature on NIR measurement. However, the drawback of this approach is that the required data collection is quite a large undertaking because the local temperature model uses a calibration set for a single selected temperature and a validation set for other temperatures, while the global temperature model contains all of the data sets. Based on large scale of data sets, Acharya et al.^[41] designed several methods of population structuring for detecting the dry matter and color of tomatoes with the aim of producing robust models that take account of sample temperature. It was concluded that temperature compensation created by adding spectra of the same set of samples at different temperatures was overwhelmed by continuing addition of 500 spectra at a uniform temperature, resulting in a model that was not robust to temperature. The use of a repeatability file was regarded as the best way for temperature compensation to predict dry matter of fruit, while selection of a wavelength region to avoid water absorption features was recommended for color prediction. Chauchard et al.^[42] categorized these temperature-compensating approaches into two major types depending on whether the temperature is measurable or not. For measurable temperatures, they tested three methods. The first used a spectrum correction while the second and third were based on regression coefficients which vary with temperature fluctuations. Studies where temperatures are non-measurable, have led to robust calibration models and to a self-correcting model, where fruit temperature is estimated using spectral data. The efficient ways to compensate for temperature fluctuation have been explored and these can also be potentially

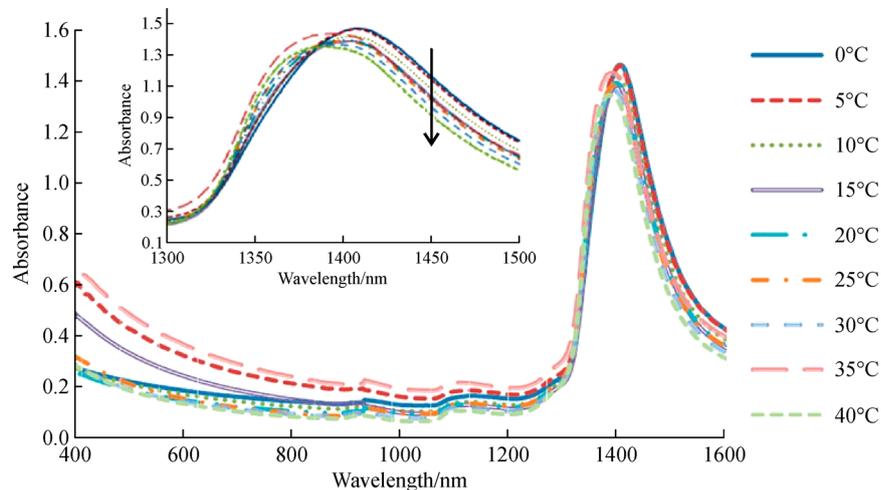


Fig. 2 The average NIR spectra of watermelon juice at different temperatures. The arrow indicates direction of temperature increase. Reprinted from Yao et al.^[37], with permission from Elsevier.

useful to minimize other external parameters in NIR calibration. Also, spectral pretreatment can be adopted to reduce the influence of temperature change in samples. Roger et al.^[43] presented a pretreatment method called external parameter orthogonalization (EPO) which aimed at removing from the calculated space the part mostly influenced by the external parameter variations. This method estimated this corresponding subspace by computing a principal component analysis on a small set of spectra measured on the same objects, while the external parameter was varied. An application for detecting the influence of the fruit temperature on the sugar measurement of intact apples is presented to better illustrate this method. Without any spectral pretreatment, the bias in the prediction of SSC was about 8 °Bx for a temperature variation of 20°C. After EPO pretreatment, the bias was less than 0.3 °Bx for the same temperature change. This approach can also be considered for use in other similar circumstances.

Research on the influence of temperature will lead to more optimized temporal conditions where the measurement should be conducted to attain a more accurate and robust calibration model for the applications in the field of food quality and security, which is similar for other external parameters. For example, De Benedictis & Huck^[44] optimized NIR spectroscopy parameters by varying investigated factors like layer thickness, number of scans and temperature during measurement, in order to obtain optimal response variables in the full factorial design consisting of absorption intensity, signal-to-noise ratio and reproducibility of the spectra. Optimized combinations for liquid milk measurement have been found to be 0.5 mm layer thickness, 64 scans and 25°C ambient temperature.

2.2 Humidity

Humidity is one of the environmental factors that influence the water status, which directly affects the absorbance of near infrared spectra. However, there have been few studies on the influence of humidity fluctuations on NIR spectra and analytical model performance. Two studies^[45,46] suggested that humidity made little difference to NIR spectra, so there are only minor improvements to be made by obtaining a better quality assessment model in these cases. Compared to the high moisture of analyzed objects, such as livestock products, fruit and vegetables, dairy products and marine products, the amount of water in the air can be neglected, therefore it has little impact. To determine whether the change in humidity can influence the NIR spectrum, Zhou et al.^[45] established an airtight, humidity-controllable test bed to change the humidity of the environment. Final sample spectra were obtained by removing the background spectra from the samples at 40%, 50%, 60%, 70% and 80% relative humidity (RH). It was not known whether the influences of the sample

spectrum and the background were equal, so the trial was divided into two groups: detecting background and sample at each humidity level (Group 1) and detecting background at 40% RH only (Group 2). The result showed that in both Groups 1 and 2, humidity had little influence on NIR spectra. However, humidity fluctuation dominates absorption fluctuation, but temperature and humidity affect refraction fluctuation. If humidity fluctuations are sufficiently strong they can dominate the refraction fluctuation for some infrared frequencies, but not for visible frequencies. Thus it will have little influence on NIR spectra^[46]. There are no specific studies showing the optimal humidity condition for NIR analysis. Thus, more in-depth and appropriate research should be conducted in this area.

2.3 Other uncertain factors

Sometimes different working environments will generate problems. For instruments in a laboratory, the influence of external light can be effectively eliminated by ensuring an effective seal in order to ensure stable working conditions. When it comes to the in-field applications, especially with portable NIR equipment, the spectra acquired may be influenced by external light and other uncertain factors. With the development of NIR spectroscopy technique, although slightly better results were obtained under laboratory conditions, the results obtained in the field were also accurate enough to determine some quality attributes of fruit^[23–28], which improved the efficiency of the evaluation process. For example, by comparing spectra data recorded in two different experimental condition: on trees in the field and under laboratory conditions with the same portable NIR machine, Gracia & León^[23] and Leónmoreno^[24] proved the feasibility of use of portable NIR instrument for determination of oil and moisture contents in intact olive fruit. Bessho et al.^[25] also demonstrated that a portable nondestructive measurement system might be a powerful tool to easily estimate fruit quality in the apple canopy. In addition, the external variables affecting online agro-product NIR assessment lines are very complicated and vary under dynamic circumstances. Sun et al.^[29] examined the effect of movement speed of fruit on SSC measurement. The results showed that a speed of 0.3–0.7 m·s⁻¹ had little effect on spectra and model performance, however these can be considered as relatively slow speeds. At 0.5 m·s⁻¹, the best model for SSC was a PLSR model established with original spectra, with its coefficient of determination and RMSEP being 0.916 and 0.53, respectively.

3 Sample factors

Agro-products are object with high variation in their physical attributes and chemical composition. They vary in

shape, color, density, storage condition and so on. Most error sources come from the sample itself and the process of dealing with it. In the beginning, sample preparations (e.g., smashing and mixing) were required for the analysis of NIR measurement for agro-product, especially for seeds and grains. Gradually, with the development of NIR noninvasive detection, the need for this was reduced and now there is no need for sample preparation. And the error sources in this process are consequently eliminated. This sector will consider different factors influencing livestock products, fruit and vegetables, which lead to variations in NIR spectra and influence the performance of prediction model.

3.1 Livestock products

For livestock products, it is suggested that mincing preparations is the best way to prepare material to analyze meat by NIR spectroscopy technique^[5,47]. In addition, the performance of prediction models depends on the animal species. Comparing the analysis of chicken breast, beef and lamb, chicken breast had the best statistical analytical model for crude protein and moisture, and those for beef and lamb are worse, according to Prevolnik et al.^[5]. However, for fat, beef was better predicted than chicken breast while lamb remained the worst. This is because major constituents are easier to predict than minor compounds and beef obviously has the highest fat content. Even muscle type can have an influence on the result and this may be interpreted as the influence of texture and different models have been prepared that vary several influencing factors including muscle type. According to calibration statistics, the prediction using a common equation for several different muscle types seemed more reliable than the equations obtained with studies of single muscle, but the latter showed less bias. Mcdevitt et al.^[48] showed that NIR spectroscopy was successful able to discriminate ground carcasses from three different genotypes: fast-growing broiler, slow-growing broiler and layer-type chicken. The varied genotypes of chicken determined different contents of fat, crude protein and ash, which explains the variation of NIR spectra. To provide people with the opportunity to buy worthwhile, healthy and safe merchandise, these differences are combined with appropriate algorithms to realize the classification of livestock product quality and identification of adulteration in the industry. Several advanced algorithms have been employed to identify adulteration and sort livestock products, including artificial neural network, discriminant analysis and other techniques^[49–51]. One of the most widespread applications is to monitor and predict the spoilage of livestock products during storage and transportation. Argyri et al.^[50] used a machine learning strategy in the form of a multilayer perceptron neural network to correlate Fourier transform near infrared (FT-NIR) spectral data with beef spoilage during aerobic

storage at chilling and abuse temperatures. The result of this work indicated that the biochemical fingerprints during beef spoilage obtained by FT-NIR spectroscopy in combination with the appropriate machine learning strategy have significant potential for rapid assessment of meat spoilage. Chen et al.^[51] demonstrated that the FT-NIR spectroscopy technique combined with a classification algorithm had the potential to determine pork storage time associated with its freshness.

3.2 Fruit and vegetables

For fruit and vegetables, there are more influencing factors from the growth stage and postharvest storage period. Samples of various cultivars, orchards, districts and seasons can all have an influence on NIR spectra. Peirs et al.^[18] researched the effect of cultivar, orchard and season. Cultivar and season were responsible for a major part of the spectral variability, whereas the influence of the orchard was low, and only occurred for certain cultivars during specific seasons. Bobelyn et al.^[19] used functional analysis of variance to interpret the variance in spectra with respect to biological variability. The effect of cultivar, storage period, origin and season on the accuracy of NIR calibration models for the SSC and firmness of apple was studied based on a large spectral data set for about 6000 apples of different cultivars and storage periods, and from different origins and seasons, which all proved to be significant. The largest difference in spectra was found around the water absorption peaks (970, 1170 and 1450 nm). External validations using an independent data set showed that the accuracy of the models increased considerably when more variability was included in the calibration data set. Guthrie et al.^[20] found that calibrations for total soluble sugar in intact pineapple were not transferrable between the summer and winter growing seasons. A combined calibration (data of three harvest dates) validated reasonably well against a population set drawn from all harvest dates. However, they came to the conclusion that a lack of robustness of calibration was indicated by poor validation within populations of fruit harvested at different time. León et al.^[21] studied the influence of parent and harvest year on NIR determinations of oil, moisture, oleic acid and linoleic acid contents in intact olive fruit. When samples for each year and female parent were analyzed separately by a PLS method with validation against the other groups, calibration models were accurate enough to predict all constituents in new samples from a different female parents but were not transferable across years. However, a calibration equation of sufficient accuracy was obtained from the combined data sets, which should prove to be a useful tool in olive breeding programs.

Deciding how to choose samples appropriately is the first step in obtaining a robust and representative prediction model^[52]. It was shown that having more variability

between samples could, to some extent, improve the robustness of calibration models. Peirs et al.^[18] tested the robustness of the calibration models of SSC with respect to cultivars, orchards and seasons based on external validations. It was found that the accuracy of the models increased considerably when more variability was included in the calibration set. Furthermore, overfitting of the calibration model was avoided. However, adding more data to the calibration set increased the chance of adding atypical data, which can reduce the accuracy of model. Thus, it was suggested that when the effect of a certain factors is known a priori, e.g., cultivar, a specific calibration model should be used. It is important to collect calibration data that are sufficiently representative for future samples to be analyzed with the calibration model developed and to develop simple procedure for model adaptation during application^[19].

Also, a non-homogenized spectral distribution of measured values of objects may make a difference to the analysis for intact fruit and vegetables^[53]. Thus, the selection of measurement position can influence the accuracy and robustness of NIR analytical results. Slaughter et al.^[54] determined SSC of intact tomato using NIR and suggested that when used at a random position along the equator of the fruit, rather than at the blossom-end, the technique had significantly greater accuracy. Thus, it was recommended that the fruit should be oriented to allow measurement at the equator when developing high-speed sorting equipment for SSC. Fan et al.^[55] studied the influence of variation of spectrum measurement position on the NIR analysis of SSC for apples (Fig. 3). This result indicated that the measurement

position influenced the prediction accuracy of SSC. Compared with the local position model, the global position model was well suited to controlling the prediction accuracy of the calibration model for SSC with respect to the variation of spectrum measurement position. Next, competitive adaptive reweighted sampling was used for the robust global position model to select the most effective wavelengths. This indicated that a global model established with effective wavelengths achieved more promising results.

Last but not least, the size and texture of different fruit types differ a lot, which will influence the choice of NIR spectra's transmission mode. Transmittance mode can acquire the most complete internal information of fruit, which will improve the prediction ability of NIR calibration models. Due to limited penetration depth of NIR light, the transmittance mode is more suitable for small fruits, such as jujube and cherry. However, by enhancing the intensity of light source, even watermelon can be analyzed using transmittance mode. On the other hand, for those fruits with a core, reflectance mode will be more suitable because it will carry more information from the part near the surface. Therefore, it is suggested that the transmission path should be well designed according to the features of the fruit sample to be tested.

4 Instrument and algorithm factors

To highlight the factors that may influence the accuracy and robustness of measurement, spectral pretreatment, detection mode, spectral range, modeling method and

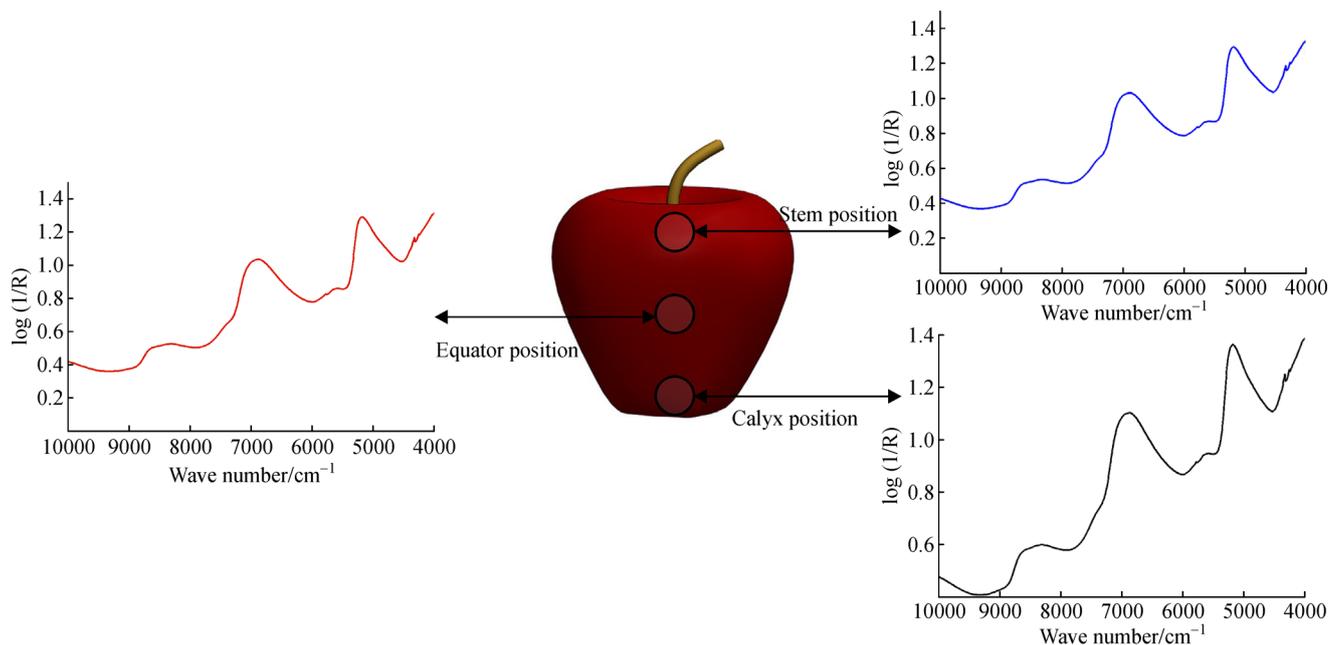


Fig. 3 Schematic of the measurement positions of NIR reflectance spectra within an individual apple fruit. Reprinted from Fan et al.^[55], with permission from Elsevier.

instrument all need to be taken into consideration. The recent studies for quality analysis of livestock products (Table 1), and fruit and vegetables (Table 2) are listed here to illustrate these factors.

Table 1 Analysis of livestock product quality by NIR spectroscopy

Object	Attribute	Detection mode	Spectral range	Pretreatment	Modeling method	Instrument	Optimal model performance
Pork ^[2]	Color (L*, a*, b*), pH value, TVB-N	Diffuse reflectance	400–1000 nm	S-G, SNV	PLS	Portable device	R_p : 0.92, 0.91, 0.92, 0.95, 0.96 (L*, a*, b*, pH value, TVB-N)
Pork ^[3]	TVB-N, WBSF	Reflectance	10000–4000 cm^{-1}	SNV	SI-PLS	Antaris II FT-NIR spectrophotometer	R_C : 0.8398, R_p : 0.8084 (TVB-N) R_C : 0.7533, R_p : 0.7041 (WBSF)
Pork ^[4]	Water content, cooking loss, tenderness	Reflectance	350–1100 nm; 1000–2500 nm	S-G, SNV	PLS	Online detection system	R_p : 0.9123, 0.9200, 0.9019 (respectively)
Pork ^[5]	IMF	Reflectance	1100–1830 nm	*	Modified PLS	NIRS 6500	R_{CV}^2 : 0.84–0.99 SECV: 0.14%–0.53%
Beef ^[5]							
Beef ^[56]	Fat, moisture, protein, myoglobin, stress 20%, stress 80%, WBSF, tenderness, juiciness, overall appraisal	Reflectance	408–2492 nm; 1108–2492 nm; 1500–2460 nm	MSC, SNV, SNVD, none	PLS, modified PLS, PCR	NIRS 6500	Optimal R_p^2 : 0.98 (tenderness)
Broiler breast ^[57]	Fatty acid	Reflectance	400–2498 nm	SNVD, WMSC	Modified PLS	NIRS 6500	R_C^2 : 0.86–0.98 R_p^2 : 0.83–0.97
Broiler breast ^[58]	Fatty acid	Reflectance	1100–1830 nm	2nd derivative	Modified PLS	LabSpec®2500	R_{CV}^2 : less than 0.60

Note: L*, the value of brightness; a*, the red-green axis; b*, the blue-yellow axis; TVB-N, total volatile basic nitrogen; WBSF, Warner-Bratzler shear force; IMF, intramuscular fat; S-G, Savitzky-Golay filter; SNV, noise removal standard normal variable transformation; *, not mentioned; MSC, multiplicative scatter correction; SNVD, noise removal standard normal variable transformation and detrend; WMSC, weighted multiplicative scatter correction; PLS, partial least square; SI-PLS, synergy interval partial least square; PCR, partial component regressions; R_p , correlative coefficient of prediction; R_C , correlative coefficient of calibration; R_{CV}^2 , determination coefficient of cross validation; SECV, standard error of cross validation; R_p^2 , determination coefficient of prediction.

Table 2 Analysis of the quality of fruit and vegetables by NIR spectroscopy

Object	Attribute	Detection mode	Spectral range	Pretreatment	Modeling method	Instrument	Optimal model performance
Apple ^[7]	SSC	Diffuse reflectance	500–1100 nm; 1000–4000 nm	*	ICA-SVM	Ocean Optics model USB2000 fiber spectrometer; antaris TM II method development sampling system	R_p : 0.9455 RMSEP: 0.3691%
Orange ^[8]	SSC	Interactance reflectance transmittance	460–1000 nm	*	PLS	VIS-SWNIR CCD spectrometer	R_{CV} : 0.778–0.866 RMSECV: 0.329–0.518
Jujube ^[9]	Inner insect-infestation	Interactance, reflectance, transmittance	310–1100 nm; 1000–2150 nm	*	DA	Handy Lambda II & Solid lambda NIR2.2t2	100% (interactance) 90% (reflectance) 97% (transmittance)
Peach ^[10]	SSC, pH	Diffuse reflectance	325–1075 nm	S-G, MSC	PLS, LS-SVM	Fieldspec Pro FR, Analytical Spectral Devices, Inc.	R_p : 0.9537, R_{CV} : 0.9485 (SSC) R_p : 0.9638, R_{CV} : 0.9657 (pH)
Apple ^[11]	ITB	Diffuse transmittance	650–950 nm	*	PLS	Prototype based on time-delayed integration spectroscopy; Prototype based on large aperture spectrometer	R_p^2 : 0.7–0.9 RMSECV: 4%–7% R_p^2 : ~0.9 RMSEP: ~4.1%

(Continued)

Object	Attribute	Detection mode	Spectral range	Pretreatment	Modeling method	Instrument	Optimal model performance
Pear ^[12]	DM, SSC	Reflectance	680–1000 nm; 1100–2350 nm	S-G, SNV	PLS	Agriquant FT-NIR spectrometer	R_{CV}^2 : 0.78–0.84
Blueberry ^[13]	Inner insect-infestation	Reflectance	650–1100 nm; 600–1700 nm	*	PLS		Two scanning spectrometers 82%, 76.9%
Tomato ^[14]	Total soluble solids, lycopene and β -carotene	Diffuse reflectance	*	MSC, 2nd derivative	PLS	Buchi NIRlab N-200 spectrometer	R_p : 0.9998, 0.9996, 0.9981 RMSEP: 0.4157, 21.5779, 0.7296
Mango ^[15]	TSS, acidity, firmness, storage period	Reflectance	1200–2400 nm	1st derivative, 2nd derivative	MLR, PCA, PLS	Quantum 120	R_p^2 : 0.9276, 0.6085, 0.8226, 0.9380 (respectively)

Note: SSC soluble solids content; ITB, internal tissue browning; DM, dry matter; TSS, total suspended solids; *, not mentioned; S-G, Savitzky-Golay filter; MSC, multiplicative scatter correction; SNV, standard normal variable transformation; ICA-SVM, independent component analysis-support vector machine; PLS, partial least square; DA, discriminant analysis; LS-SVM, least square-support vector machine; MLR, multilinear regression; PCA, principal component analysis; R_p , correlative coefficient of prediction; RMSEP, root mean square of prediction; R_{CV} , correlative coefficient of cross validation; RMSECV, root mean square of cross validation; R_p^2 , determination coefficient of prediction; R_{CV}^2 , determination coefficient of cross validation.

From the instrument perspective, for different samples the wavelength range and detection mode vary, which is consistent with the previous discussion. When it comes to detection mode, reflectance is the most common spectral acquisition mode in NIR analysis shown in the tables. This may be due to its convenience for configuration and transmittance mode will sometimes lead to overheating of samples. However, the transmittance mode shows better prediction ability than reflectance mode. Interactance mode combines the features of reflectance and transmittance mode, which can also yield satisfying results. The NIR instrumental system normally consists of light source, monochromator, detector, sample compartment and optical accessory^[59]. In recent years, the instrument has become portable and automated. From the tables, it can also be seen that online and portable equipments have shown reasonable prediction ability while laboratory studies yielded better prediction models and results.

From the algorithm perspective, parts of pretreatment, variable selection and modeling are important. There are various spectral pretreatment algorithms, such as Savitzky-Golay (S-G) smoothing, standard normal variable (SNV) transformation and multiplicative scatter correction, all with the same aim of removing useless information in NIR spectra. An S-G filter and SNV are commonly used to preprocess spectra before establishing models. S-G filters can be used to minimize noise effects. SNV can be applied to remove slope variation and correct for scatter effects^[2,60,61]. Variable selection methods are used to make the model more efficient and effective. However, these two steps are not always necessary for prediction. Last but not least, mathematical algorithms, such as partial least squares, multilinear regression and principal component analysis are employed to develop calibration models.

Sometimes different algorithms will be combined to improve the performance of a calibration model.

5 Conclusions

The influencing factors that can affect NIR analysis have been summarized and classified into several types in this review. Some appropriate methods are discussed that can be considered to compensate for these influences in order to improve the robustness and accuracy of NIR prediction model.

Among environment factors, temperature has the most obvious influence on NIR spectra, while humidity has a minor influence. Ambient temperature affects both spectral output of instruments and also the sample. To minimize the influence of temperature, it is suggested to carry out experiments at a controlled temperature. However, for occasions with inevitable temperature fluctuation, global temperature models can be used to compensate for its influence. In current applications, NIR analysis is usually conducted in a specific temperature range according to the guidelines for the conditions where reference measurements are taken.

For livestock products, sample status (being minced or intact), muscle species (texture) and storage time will affect NIR spectra. These differences are combined with appropriate algorithms to realize the classification of livestock product quality and identification of adulteration in the industry. For fruit and vegetables, cultivar, parent, orchard, district, season, storage period, measurement position, size, texture and corresponding transmittance mode are all factors that will influence the performance of NIR spectra and analysis. Therefore, robust models can

only be developed with sufficiently representative samples varying in the corresponding variables.

In addition, instrument and algorithm also play an important role in the process of NIR spectroscopy analysis. The choice of instrument and algorithm should suit the features of the samples being tested. There is a trend that the range of portable and online equipment will grow fast, in order to meet the needs of the agro-products industry. Improving model accuracy and robustness when applied to samples that are moving is still a crucial problem to solve.

To conclude, the performance of NIR prediction model can be influenced by many environmental factors that affect the instrument and the sample itself. To improve model performance in NIR analysis, choosing an appropriate instrument, spectral range, modeling method, spectral pretreatment and representative sample is crucial. Sometimes the outcome depends greatly on the experience and knowledge of the operator and the guidelines are highly recommended in order to minimize sources of error for different batches of experiments. Thus, a database for NIR spectra of different objects should be established, based on standard and fixed reference measurement procedures. Calibration models in NIR instruments and systems should be updated regularly and in accordance with the specific operating conditions.

Acknowledgements The authors gratefully acknowledge the financial support provided by the Science and Technology Cooperation Project between Hong Kong, Macao and Taiwan (2015DFT30150).

Compliance with ethics guidelines Xiao Xu, Lijuan Xie, and Yibin Ying declare that they have no conflicts of interest or financial conflicts to disclose.

This article is a review and does not contain any studies with human or animal subjects performed by any of the authors.

References

- Williams P C, Sobering D C. Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. *Journal of Near Infrared Spectroscopy*, 1993, **1**(1): 25–32
- Sun H, Peng Y. A portable nondestructive real-time detection system for inspection of pork quality attributes using Vis/NIR spectral technique. In: SPIE Commercial + Scientific Sensing and Imaging 2016, Maryland. San Francisco: *International Society for Optics and Photonics*, 2016, 9864
- Cai J, Chen Q, Wan X, Zhao J. 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 Chemistry*, 2011, **126**(3): 1354–1360
- Wang W, Peng Y, Zheng X, Tian F, Wei W. A non-destructive detection system for determination of multi-quality parameters of meat. In: ASABE Annual International Meeting 2016, Orlando. Orlando: *American Society of Agricultural and Biological Engineers*, 2016, 1
- Prevolnik M, Čandek-Potokar M, Škorjanc D, Velikonja-Bolta Š, Škrlep M, Znidaršič T, Babnik D. Predicting intramuscular fat content in pork and beef by near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, 2005, **13**(2): 77–85
- Cozzolino D, Barlocco N, Vadell A, Ballesteros F, Gallieta G. The use of visible and near-infrared reflectance spectroscopy to predict colour on both intact and homogenised pork muscle. *LWT-Food Science and Technology*, 2003, **36**(2): 195–202
- Guo Z, Huang W, Peng Y, Chen Q, Ouyang Q, Zhao J. Color compensation and comparison of shortwave near infrared and long wave near infrared spectroscopy for determination of soluble solids content of ‘Fuji’ apple. *Postharvest Biology and Technology*, 2016, **2015**(115): 81–90
- Wang A, Hu D, Xie L. Comparison of detection modes in terms of the necessity of visible region (VIS) and influence of the peel on soluble solids content (SSC) determination of navel orange using VIS-SWNIR spectroscopy. *Journal of Food Engineering*, 2013, **2014**(126): 126–132
- Wang J, Nakano K, Ohashi S, Takizawa K, He J. Comparison of different modes of visible and near-infrared spectroscopy for detecting internal insect infestation in jujubes. *Journal of Food Engineering*, 2010, **101**(1): 78–84
- Shao Y, Bao Y, He Y. Visible/near-infrared spectra for linear and nonlinear calibrations: a case to predict soluble solids contents and pH value in peach. *Food and Bioprocess Technology*, 2011, **4**(8): 1376–1383
- McGlone V A, Martinsen P J, Clark C J, Jordan R B. On-line detection of brownheart in Braeburn apples using near infrared transmission measurements. *Postharvest Biology and Technology*, 2005, **37**(2): 142–151
- Travers S, Bertelsen M G, Petersen K K, Kucheryavskiy S V. Predicting pear (cv. Clara Frijs) dry matter and soluble solids content with near infrared spectroscopy. *LWT-Food Science and Technology*, 2014, **59**(2): 1107–1113
- Peshlov B N, Dowell F E, Drummond F A, Donahue D W. Comparison of three near infrared spectrophotometers for infestation detection in wild blueberries using multivariate calibration models. *Journal of Near Infrared Spectroscopy*, 2009, **17**(4): 203–212
- Pedro A M K, Ferreira M M C. Nondestructive determination of solids and carotenoids in tomato products by near-infrared spectroscopy and multivariate calibration. *Analytical Chemistry*, 2005, **77**(8): 2505–2511
- Schmilovitch Z, Mizrach A, Hoffman A, Egozi H, Fuchs Y. Determination of mango physiological indices by near-infrared spectrometry. *Postharvest Biology and Technology*, 2000, **19**(3): 245–252
- Uddin M, Ishizaki S, Okazaki E, Tanaka M. Near-infrared reflectance spectroscopy for determining end-point temperature of heated fish and shellfish meats. *Journal of the Science of Food and Agriculture*, 2002, **82**(3): 286–292
- Cao F, Wu D, He Y. Soluble solids content and pH prediction and varieties discrimination of grapes based on visible-near infrared spectroscopy. *Computers and Electronics in Agriculture*, 2010, **71**(S1): S15–S18
- Peirs A, Tirry J, Verlinden B, Darius P, Nicolai B M. Effect of

- biological variability on the robustness of NIR models for soluble solids content of apples. *Postharvest Biology and Technology*, 2003, **28**(2): 269–280
19. Bobelyn E, Serban A S, Nicu M, Lammertyn J, Nicolai B M, Saeys W. Postharvest quality of apple predicted by NIR-spectroscopy: study of the effect of biological variability on spectra and model performance. *Postharvest Biology and Technology*, 2010, **55**(3): 133–143
 20. Guthrie J, Wedding B, Walsh K. Robustness of NIR calibrations for soluble solids in intact melon and pineapple. *Journal of Near Infrared Spectroscopy*, 1998, **6**(1): 259–266
 21. León L, Garrido-Varo A, Downey G. Parent and harvest year effects on near-infrared reflectance spectroscopic analysis of olive (*Olea europaea* L.) fruit traits. *Journal of Agricultural and Food Chemistry*, 2004, **52**(16): 4957–4962
 22. Ding H B, Xu R J. Near-infrared spectroscopic technique for detection of beef hamburger adulteration. *Journal of Agricultural and Food Chemistry*, 2000, **48**(6): 2193–2198
 23. Gracia A, León L. Non-destructive assessment of olive fruit ripening by portable near infrared spectroscopy. *Grasas y Aceites*, 2011, **62**(3): 268–274
 24. Leónmoreno L. Usefulness of portable near infrared spectroscopy in olive breeding programs. *Spanish Journal of Agricultural Research*, 2012, **10**(1): 141
 25. Bessho H, Kudo K, Omori J, Inomata Y, Wada M, Masuda T, Nakamoto Y, Fujisawa H, Suzuki Y. A portable non-destructive quality meter for understanding fruit soluble solids in apple canopies. *Acta Horticulturae*, 2007, **2007**(732): 593–597
 26. Bellincontro A, Taticchi A, Servili M, Esposito S, Farinelli D, Mencarelli F. Feasible application of a portable NIR-AOTF tool for on-field prediction of phenolic compounds during the ripening of olives for oil production. *Journal of Agricultural and Food Chemistry*, 2012, **60**(10): 2665–2673
 27. Saranwong S, Sornsrivichai J, Kawano S. Prediction of ripe-stage eating quality of mango fruit from its harvest quality measured nondestructively by near infrared spectroscopy. *Postharvest Biology and Technology*, 2004, **31**(2): 137–145
 28. Camps C, Christen D. Non-destructive assessment of apricot fruit quality by portable visible-near infrared spectroscopy. *LWT-Food Science and Technology*, 2009, **42**(6): 1125–1131
 29. Sun T, Lin H, Xu H, Ying Y. Effect of fruit moving speed on predicting soluble solids content of ‘Cuiguan’ pears (*Pomaceae pyrifolia* Nakai cv. Cuiguan) using PLS and LS-SVM regression. *Postharvest Biology and Technology*, 2009, **51**(1): 86–90
 30. Yan Y, Zhao L, Han D. Analytical basis and application of near infrared spectroscopy. Beijing: *China Light Industry Press*, 2005 (in Chinese)
 31. Maeda H, Ozaki Y, Tanaka M, Hayashi N, Kojima T. Near infrared spectroscopy and chemometrics studies of temperature-dependent spectral variations of water: relationship between spectral changes and hydrogen bonds. *Journal of Near Infrared Spectroscopy*, 1995, **3**(4): 191–202
 32. Walsh K B, Guthrie J A, Burney J W. Application of commercially available, low-cost, miniaturised NIR spectrometers to the assessment of the sugar content of intact fruit. *Functional Plant Biology*, 2000, **27**(12): 1175–1186
 33. Greensill C V. Non-invasive assessment of fruit quality by near-infrared spectroscopy for fruit grading in an in-line setting. Dissertation for the Doctoral Degree. Australia: *Central Queensland University*, 2000
 34. Hayes C J, Greensill C V, Walsh K B. Temporal and environmental sensitivity of a photodiode array spectrophotometric system. *Journal of Near Infrared Spectroscopy*, 2014, **22**(4): 297–304
 35. Martinsen P, McGlone V A, Jordan R B, Gaastra P. Temporal sensitivity of the wavelength calibration of a photodiode array spectrometer. *Applied Spectroscopy*, 2010, **64**(12): 1325–1329
 36. Williams P C, Norris K H, Zarowski W S. Influence of temperature on estimation of protein and moisture in wheat by near-infrared reflectance. *Cereal Chemistry*, 1982, **59**(6): 473–477
 37. Yao Y, Chen H, Xie L, Rao X. Assessing the temperature influence on the soluble solids content of watermelon juice as measured by visible and near-infrared spectroscopy and chemometrics. *Journal of Food Engineering*, 2013, **119**(1): 22–27
 38. Zhang X, Chang M, Xing L, Hu J. Influence and correction of temperature on optical measurement for fat and protein contents in a complex food model system. *Infrared Physics & Technology*, 2010, **53**(3): 177–181
 39. Peirs A, Scheerlinck N, Nicolai B M. Temperature compensation for near infrared reflectance measurement of apple fruit soluble solids contents. *Postharvest Biology and Technology*, 2003, **30**(3): 233–248
 40. Acharya U K, Walsh K B, Subedi P P. Robustness of partial least-squares models to change in sample temperature: I. A comparison of methods for sucrose in aqueous solution. *Journal of Near Infrared Spectroscopy*, 2014, **22**(4): 279–286
 41. Acharya U K, Walsh K B, Subedi P P. Robustness of partial least-squares models to change in sample temperature: II. Application to fruit attributes. *Journal of Near Infrared Spectroscopy*, 2014, **22**(4): 287–295
 42. Chauchard F, Roger J M, Bellon-Maurel V. Correction of the temperature effect on near infrared calibration-application to soluble solid content prediction. *Journal of Near Infrared Spectroscopy*, 2004, **12**(1): 199–205
 43. Roger J M, Chauchard F, Bellon-Maurel V. EPO-PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. *Chemometrics and Intelligent Laboratory Systems*, 2003, **66**(2): 191–204
 44. De Benedictis L, Huck C. New approach to optimize near-infrared spectra with design of experiments and determination of milk compounds as influence factors for changing milk over time. *Food Chemistry*, 2016, **212**: 552–560
 45. Zhou Y, Fu X P, Ying Y B. Effect of humidity on detection of near-infrared spectra. *Spectroscopy and Spectral Analysis*, 2007, **27**(11): 2197–2199
 46. Hill R J, Clifford S F, Lawrence R S. Refractive-index and absorption fluctuations in the infrared caused by temperature, humidity, and pressure fluctuations. *Journal of Science and Arts*, 1980, **70**(10): 1192–1205
 47. Cozzolino D, Murray I. Effect of sample presentation and animal muscle species on the analysis of meat by near infrared reflectance spectroscopy. *Journal of Near Infrared Spectroscopy*, 2002, **10**(1):

37–44

48. Mcdevitt R M, Gavin A J, Andrés S, Murray I. The ability of visible and near infrared reflectance spectroscopy to predict the chemical composition of ground chicken carcasses and to discriminate between carcasses from different genotypes. *Hispania*, 2005, **13** (3): 109–117
49. Nunes K M, Andrade M V, Santos Filho A M, Lasmar M C, Sena M M. Detection and characterisation of frauds in bovine meat in natura by non-meat ingredient additions using data fusion of chemical parameters and ATR-FTIR spectroscopy. *Food Chemistry*, 2016, **205**: 14–22
50. Argyri A A, Panagou E Z, Tarantilis P A, Polysiou M, Nychas G J E. Rapid qualitative and quantitative detection of beef fillets spoilage based on Fourier transform infrared spectroscopy data and artificial neural networks. *Sensors and Actuators B: Chemical*, 2010, **145**(1): 146–154
51. Chen Q, Cai J, Wan X, Zhao J. Application of linear/non-linear classification algorithms in discrimination of pork storage time using Fourier transform near infrared (FT-NIR) spectroscopy. *LWT-Food Science and Technology*, 2011, **44**(10): 2053–2058
52. Bureau S, Ruiz D, Reich M, Gouble B, Bertrand D, Audergon J M, Renard C M G C. Rapid and non-destructive analysis of apricot fruit quality using FT-near-infrared spectroscopy. *Food Chemistry*, 2009, **113**(4): 1323–1328
53. Martinsen P, Schaare P. Measuring soluble solids distribution in kiwifruit using near-infrared imaging spectroscopy. *Postharvest Biology and Technology*, 1998, **14**(3): 271–281
54. Slaughter D C, Barrett D, Boersig M. Nondestructive determination of soluble solids in tomatoes using near infrared spectroscopy. *Journal of Food Science*, 1996, **61**(4): 695–697
55. Fan S, Zhang B, Li J, Huang W, Wang C. Effect of spectrum measurement position variation on the robustness of NIR spectroscopy models for soluble solids content of apple. *Biosystems Engineering*, 2016, **143**(3): 9–19
56. Ripoll G, Albertí P, Panea B, Olleta J L, Sanudo C. Near-infrared reflectance spectroscopy for predicting chemical, instrumental and sensory quality of beef. *Meat Science*, 2008, **80**(3): 697–702
57. Zhou L J, Wu H, Li J T, Wang Z Y, Zhang L Y. Determination of fatty acids in broiler breast meat by near-infrared reflectance spectroscopy. *Meat Science*, 2012, **90**(3): 658–664
58. De M M, Riovanto R, Penasa M, Cassandro M. At-line prediction of fatty acid profile in chicken breast using near infrared reflectance spectroscopy. *Meat Science*, 2012, **90**(3): 653–657
59. Xie L, Wang A, Xu H, Ying Y. Applications of near-infrared systems for quality evaluation of fruits: a review. *Transactions of the ASABE*, 2016, **59**(2): 399–419
60. Chu X L, Yuan H F, Lu W Z. Progress and application of spectral data pretreatment and wavelength selection methods in NIR analytical technique. *Progress in Chemistry*, 2004, **16**(4): 528–542
61. Fernández-Cabanás V M, Garrido-Varo A, Olmo J G, Pedro E D, Dardenne P. Optimisation of the spectral pre-treatments used for Iberian pig fat NIR calibrations. *Chemometrics and Intelligent Laboratory Systems*, 2007, **87**(1): 104–112