

Identification of Foxtail Millet Varieties Using Leaf Surface Spectral Information

Xiaoping Han,¹ Wei Yang,¹ Haiyan Song,¹ Zhiyong Zhang,¹
Yueming Zuo,¹ Zhiying Duan,¹ and Xuyuan Zhang^{2*}

¹Shanxi Agricultural University, Taigu 030801, China

²Hainan Vocational University of Science and Technology, Haikou 571126, China

(Received November 26, 2019; accepted March 10, 2020)

Keywords: foxtail millet, surface spectral information, neural network, identification of varieties

The increasing scale of plantation and production of foxtail millet (*Setaria italica*) has led to a strong demand to identify its varieties easily and quickly. It is also important for researchers to find, screen, identify, protect, and collect new mutant species and germplasm resources of foxtail millet in the early stage of growth. In this study, we present an innovative approach to identifying foxtail millet varieties using visible–near-infrared (VIS–NIR) spectral information from their growing leaves. Seven varieties of foxtail millet were successfully identified. Ten effective wavelengths (1440, 1660, 1775, 550, 410, 980, 1180, and 462 nm) were extracted. An accurate and stable prediction model for foxtail millet varieties was developed using the backpropagation (BP) neural network coupled with principal component analysis (PCA). The model can completely classify the foxtail millet varieties with a minimal number of five hidden-layer nodes. Its predictive correlation coefficient (R_v) is as high as 0.9994. Accordingly, the root-means-square error of prediction (RMSEP) and the standard error of prediction (SEP) are both 0.0026. The results show that the VIS–NIR spectral technique can be used for identifying foxtail millet varieties.

1. Introduction

Foxtail millet (*Setaria italica*) is a gramineous annual crop. The plant has a bulky root with a stem generally with a length of 1 m or larger and is wrapped by the leaf sheath. The leaf blade is linear lanceolate in shape, 10–45 cm long, and 5–33 mm wide. The panicle has a length of 20–30 cm, bearing a few hundred to thousands of grains. Foxtail millet has five growth stages: seedling, jointing, heading, flowering, and ripening.

Foxtail millet is a nutrient-rich food and feed crop. In China, it is a traditional, primary food crop, especially in dry northern areas, because foxtail millet has good quality and can tolerate drought, poor soil, and long-term storage.⁽¹⁾ China is the first in foxtail millet production in the world.⁽²⁾

*Corresponding author: e-mail: zxy80327@163.com
<https://doi.org/10.18494/SAM.2020.2718>

With increasing number of foxtail millet varieties, it is increasingly becoming difficult to distinguish them by color and form. However, we must identify them accurately to guarantee the purity of germplasm resources and meet production demand. Current methods of identifying foxtail millet are primarily based on morphological characteristics, biochemical indexes (such as proteins and isoenzymes), and DNA sequencing.⁽³⁾ These methods are all complex, tedious, and labor-intensive.

Near-infrared (NIR) spectroscopy has attracted increasing attention owing to its noninvasiveness, reliability, speed, low cost, and nonpolluting nature.^(4,5) The NIR region is the range of the electromagnetic spectrum between 750 and 2500 nm, with spectra defined by absorption bands associated with overtones and combinations of fundamental vibrations arising from functional groups of molecules (e.g., C–H, N–H, O–H, and S–H) found in many biological samples.⁽⁶⁾ In agricultural production, NIR is widely used because it can provide abundant structural and compositional information from test targets. It is used primarily for the online and field detection of the water, protein, and starch contents of cereals such as wheat and corn.^(7–10) In recent years, NIR has also been widely used to trace cereal quality⁽¹¹⁾ and in genetic breeding.^(12,13) However, it has rarely been applied to foxtail millet production.

In this study, an identification model for foxtail millet varieties was established by artificial neural network (ANN) techniques based on the spectral information from foxtail millet leaves. The aim is to explore an accurate, simple, and rapid identification method for foxtail millet.

2. Data and Methods

2.1 Data acquisition

Seven varieties of foxtail millet Jingu-33, Jingu-29, Jingu-21, Changsheng-6, Zhangza-3, Zhangza-10, and Zhangza-9 were planted in a standard test field with a row distance of 30 cm and a line distance of 10 cm. Samples were collected three times during the millet's heading stage; each time, twelve leaves were randomly collected from each variety, yielding a total of 84 samples per collection. The collected leaves were grouped by variety and placed into zip-lock bags. Within half an hour, these leaves should be tested in a spectrograph lab; otherwise, the moisture loss in the leaves may affect the accuracy of the test results.

In this study, FieldSpec3, a portable visible–near-infrared (VIS–NIR) spectrometer developed by Analytical Spectral Device Company was used. Its wave band ranges from 350 to 2500 nm, with resolutions of 3, 10, and 10 nm at 700, 1400, and 2100 nm, and sampling intervals of 1.4 and 2 nm at 350–1000 and 1000–2500 nm, respectively. The data interval, wavelength accuracy, and wavelength repeatability were 1, ± 1 , and ± 0.02 nm, respectively. The field view angle was 25°.

To avoid environmental light interference during measurement, a plant probe and a leaf-clip assembly were used, referred to collectively as a leaf detector.⁽¹⁴⁾ The effective spot diameter of the detector is 10 mm and its maximum mirror reflectance loss is 5%. At room temperature, for every sample, its spectral information was collected from three spots in the middle of the foxtail millet leaf blade with a width of greater than 10 mm by the leaf detector. We got 252 spectral data and all the data were analyzed and processed together in MATLAB 7.6.

2.2 Data preprocessing

The obtained spectral data is large and redundant. We must preprocess the data to gain representative spectral information.

First, the process of obtaining spectral data is easily disturbed by the surrounding environment, such as background, light, and high-frequency signals. Although the leaf detector can avoid some of the environmental noise, systematic errors still exist.

Second, the overlapping of spectral lines can lead to the low-content component spectrum lines being masked by the high-content component in testing materials.

Third, it will be impossible to establish a good identification model on the basis of high-dimensional data. Therefore, it is necessary to preprocess spectral signals by denoising and extracting the feature spectra.⁽¹⁵⁾ In this process, Savitzky–Golay smoothing (SGS) was used to reduce random noise, multiple scattering correction (MSC) was used to correct the baseline shift, and wavelet transformation (WT) was used to remove high-frequency noise. WT is a useful tool for time–frequency signal analysis and processing. It can automatically adapt to the requirements of time–frequency signal analysis by a changing the “time–frequency” window with frequency. Moreover, it can focus on any detail of the signal by performing time subdivision at a high frequency and frequency subdivision at a low frequency. It compensates for the defect of Fourier transform (FT) and is considered as a major breakthrough in scientific methods since Fourier transformation.

2.3 Feature spectral extraction

Figure 1 shows the average reflection spectrogram of the seven varieties of foxtail millet leaf after preprocessing. There are clear peak and valley points indicated by dashed lines (Fig. 1). The peak and valley points corresponding to the functional groups are shown in Table 1.

In the visible spectrum (bands 1 to 5), the response wave band of folic acid is at 365 nm.⁽¹⁶⁾ The absorption peaks of chlorophyll *a* are at 410 and 675 nm, whereas that of chlorophyll *b* is at 462 nm;⁽¹⁷⁾ the strongest reflection peak of chlorophylls *a* and *b* is at 550 nm.⁽¹⁸⁾

In the near-infrared spectrum (bands 6 to 12), as shown in Table 1, the response wave band corresponds to the functional groups of green leaves of foxtail millet according to Siesler *et al.*⁽¹⁹⁾

The reflectivity above all of the response wave bands were subjected to principal component analysis (PCA). It is a multivariate statistical data compression technique to reduce the number of dimensions of data while retaining information by choosing a reduced number of new variables to replace the original variables. In this manner, it can be used to solve the problem of overlapping in NIR spectral bands while eliminating the effect of random factors.⁽²⁰⁾

3. Backpropagation (BP) Neural Network Identification Models

The BP neural network is a multilayer feedforward neural network that is trained using an error-reverse propagation algorithm. It is composed of input, hidden, and output layers. BP models were built to identify foxtail millet varieties with the following parameters:

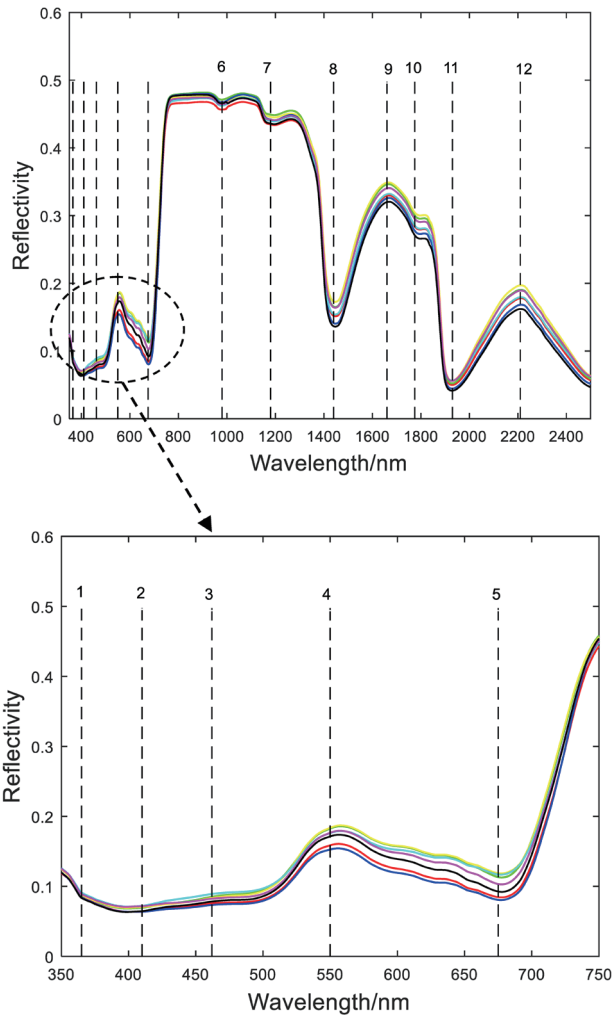


Fig. 1. (Color online) Average spectral reflectivity after denoising.

Table 1

Response wave bands and functional groups.

	Wavelength (nm)	Substance and functional groups
1	365	Folic acid
2	410	Chlorophyll <i>a</i>
3	462	Chlorophyll <i>b</i>
4	550	Chlorophylls <i>a</i> and <i>b</i>
5	675	Chlorophyll <i>a</i>
6	980	NH ₂ , OH
7	1180	CH ₂
8	1440	CH
9	1660	C=C
10	1775	CH, CH ₂ , CH ₃
11	1930	CONH, COOH, H ₂ O
12	2210	CH, NH ₂ , CHO

- *logsig* is the input layer neuron transfer function.
- *purelin* is the output layer neuron transfer function.
- *trainlm* is the Levenberg–Marquardt training algorithm implemented.

The size of a network is determined by the number of hidden-layer nodes (m), which is generally calculated using the following empirical equation:⁽²¹⁾

$$m = \sqrt{n+l} + t, \quad (1)$$

where m is the number of hidden-layer nodes; n is the number of input-layer nodes; l is the number of output-layer nodes, and t is a constant between 1 and 10.

Here, n is the number of principal components, l is the number of varieties of foxtail millet: Jingu-33, Jingu-29, Jingu-21, Changsheng-6, Zhangza-3, Zhangza-10, and Zhangza-9. Thus, l was set as 7. By calculating and rounding, the value of m was obtained from 5 to 14.

We selected randomly 168 for the training set and 84 for the test set among the 252 spectral data. Table 2 lists the evaluation parameters of the prediction models for different numbers of hidden-layer nodes.

3.1 Models

To verify the advantages of WT, we established the two models PCA-BP and WT-PCA-BP. The evaluation parameters of the two models are shown in Table 2.

From the perspective of prediction parameters, both models have high prediction correlation coefficient above on 0.97 and small prediction error below to 0.11. On the whole, WT-PCA-BP is better than PCA-BP. The average R_v , root-means-square error of prediction ($RMSEP$), and standard error of prediction (SEP) were 0.9939, 0.0299, and 0.0299, respectively, in the PCA-BP model. In the WT-PCA-BP model, the average R_v , $RMSEP$, and SEP were 0.9962, 0.0284, and 0.0188, respectively. Thus, WT is effective for improving the model's prediction accuracy. It is clear that the WT-PCA-BP model showed higher correlation and smaller prediction error and was thus chosen as the prediction model for this study.

3.2 Hidden-layer nodes

In a BP neural network, the number of hidden-layer nodes is an important parameter. If there are very few hidden layers, the network prediction accuracy will be low; if there are too many hidden layers, the network study time will increase and cause the training to fall into a local minimum-point trap.

Table 2
Prediction models' evaluation parameters.

HLN	PCA-BP			WT-PCA-BP		
	R_v	$RMSEP$	SEP	R_v	$RMSEP$	SEP
5	0.9765	0.1176	0.1102	0.9994	0.0026	0.0026
6	0.9952	0.0215	0.0216	0.9972	0.0142	0.0134
7	0.9995	0.0027	0.025	0.9958	0.0195	0.0193
8	0.9989	0.0054	0.0052	0.9996	0.0018	0.0017
9	0.9992	0.0049	0.0039	0.9957	0.0201	0.0198
10	0.9995	0.0023	0.0023	0.9976	0.0109	0.0109
11	0.9976	0.0107	0.0108	0.994	0.0285	0.0278
12	0.9956	0.0214	0.0207	0.996	0.0201	0.019
13	0.9987	0.0061	0.006	0.9899	0.0536	0.0495
14	0.9784	0.0912	0.0933	0.9966	0.0165	0.016

Note: HLN is the hidden-layer nodes. R_v is the predictive correlation coefficient. $RMSEP$ is the root-mean-square error of prediction. SEP is the standard error of prediction.

In general, increasing the number of hidden layers can reduce network error and improve accuracy; however, the increase in network complexity will in turn increase the network training time and cause the likelihood of “overfitting”. Therefore, the prediction results of WT-PCA-BP need further comparison. Figure 2 shows the prediction results for the model with the smallest number of hidden-layer nodes ($m = 5$) and the model with the greatest R_v and the smallest errors ($m = 8$).

The two models identified the foxtail millet varieties with 100% success. The absolute error is close to zero for both models. There were no clear differences between WT-PCA-BP₅ and WT-PCA-BP₈, except that there are some small fluctuations on the line of absolute error of the model with five hidden layer nodes.

The two models produced nearly identical values of evaluation parameters. Comparison of their evaluation parameter values (Table 2) showed that the number of hidden-layer nodes increased from 5 to 8, R_v is increased by 0.0002, whereas $RMSEP$ and SEP are decreased by 0.0008 and 0.0009, respectively.

Nevertheless, these changes are not significant, and WT-PCA-BP₅ has the minimum number of hidden nodes and the shortest running time. Thus, WT-PCA-BP₅ was chosen as the model for identifying foxtail millet. As shown in Table 2, the R_v is up to 0.9994, and the $RMSEP$ and SEP evaluation parameters are both 0.0026.

3.3 Characteristic wave points

The WT-PCA-BP₅ model obtains seven principal components. Its coefficient matrix is shown in Table 3. The coefficient matrix shows the correlations between the principal components and the variables, with higher coefficients indicating better correlation between the principal component and the spectral reflectance. The first principal component contains the largest amount of information. Its coefficients are all positive. We discover that 0.3721 and 0.3700

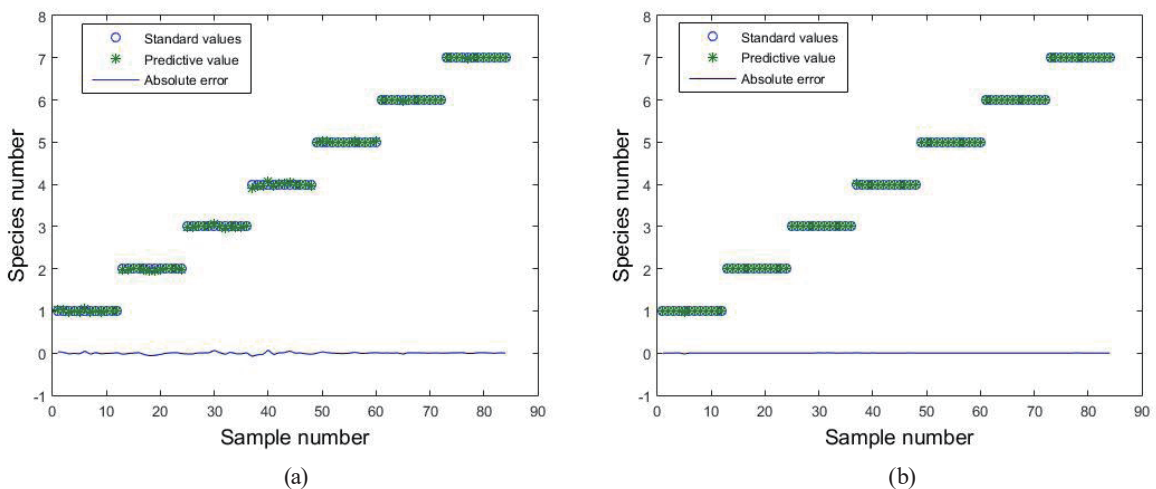


Fig. 2. (Color online) Prediction results with (a) five hidden-layer nodes and (b) eight hidden-layer nodes.

Table 3
Principal component coefficient matrix of WT-PCA-BP₅.

Wave point (nm)	Principle component						
	1	2	3	4	5	6	7
365	0.2690	0.0952	-0.4748	-0.026	-0.0224	0.2410	0.0704
410	0.2982	0.1906	0.0709	0.7541	-0.1166	-0.1178	-0.1211
462	0.3265	-0.4489	-0.2851	0.1957	0.0054	0.0260	0.6781
550	0.2655	0.4491	-0.3217	-0.5028	-0.2213	-0.0234	0.0797
675	0.2919	-0.0947	-0.3268	0.0913	-0.2329	-0.2425	-0.5644
980	0.3546	-0.3351	0.1347	-0.2688	0.6039	-0.3544	-0.1747
1180	0.3061	-0.0602	-0.1170	0.026	0.3219	0.5631	-0.3082
1440	0.3721	-0.2330	0.3028	-0.2091	-0.5144	-0.2867	0.0267
1660	0.3700	0.0619	0.5764	-0.0893	-0.1542	0.5012	0.0294
1775	0.2859	0.6070	0.1488	0.0814	0.355	-0.3031	0.2618
1930	0	0	0	0	0	0	0
2210	0	0	0	0	0	0	0

are the larger coefficients by comparison. Its main representative wave points are 1440 (CH) and 1660 nm (C=C). The higher correlations for the second principal component are 1775 (CH, CH₂, CH₃) and 550 nm (chlorophylls *a* and *b*, respectively). Similarly, the third principal component correlation maximum is still 1660 nm (C=C) and similar observations were made for the other principal components. Other wave points show the following order: 410, 980, 1180, and 462 nm. There are no clear correlations between 365 and 675 nm.

Besides, a coefficient close to zero indicates the wave point that is redundant and can be deleted. Clearly, the wave points 1930 and 2210 nm can be eliminated from the characteristic wave points for the identification of foxtail millet. It is shown that the functional groups (Table 1) at the two wave points are insensitive to the variety of foxtail millet.

Thus, the eight characteristic bands were gained for the identification of foxtail millet varieties in the following order: 1440, 1660, 1775, 550, 410, 980, 1180, and 462 nm.

3.4 Identification model PCA

The WT-PCA-BP₅ model obtains seven principal components whose contribution rates and cumulative contribution rates are listed in Table 4. Figure 3 shows the score figure of the first three principal components.

In the figure, 1 to 7 represent the foxtail millet varieties Jingu-33, Jingu-29, Jingu-21, Changsheng-6, Zhangza-3, Zhangza-10, and Zhangza-9, respectively. The foxtail millet samples were clearly separated from each other, whereas almost all the samples of the same variety were clustered densely together. The first three principal components accounted for 89.5% of the sample information. The total cumulative contribution of the seven principal components was up to 99.9%. Thus, these principal components successfully replaced the original data and were used to identify the seven varieties of foxtail millets.

The above results confirm that WT-PCA-BP₅ is an ideal identification model with a prediction accuracy of 100%. In the model, WT is an advanced and ideal tool in terms of

Table 4
Contribution rates of WT-PCA-BP₅

Component	Contribution rate (%)	Cumulative contribution rate (%)
1	69.4851	69.4851
2	11.2691	80.7542
3	8.7511	89.5053
4	4.9343	94.4396
5	2.8283	97.2679
6	1.7811	99.0490
7	0.8745	99.9235

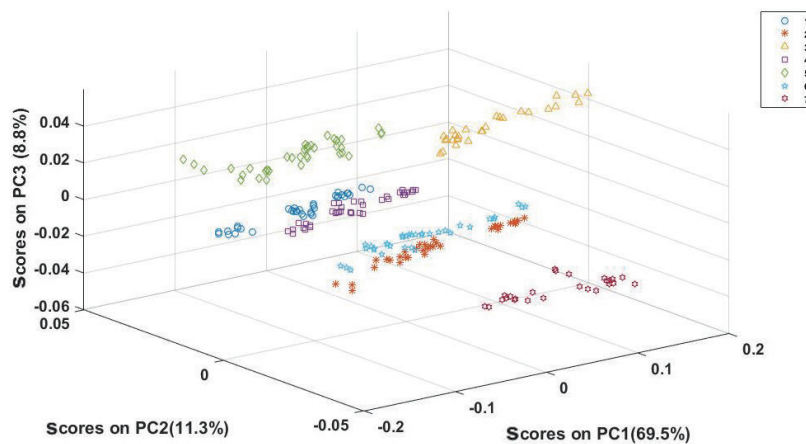


Fig. 3. (Color online) PCA scores for NIR spectra of seven varieties of foxtail millet.

time scale and multiresolution signal analysis; PCA reduced the spectral data dimension and extracted foxtail millet characteristic waves; the BP neural network can realize complex nonlinear mapping. This unique combination of artificial intelligence and other data processing methods is the key to obtaining an ideal model.

4. Conclusions and Suggestions

In this paper, an innovative method of identifying foxtail millet varieties was established by VIS–NIR spectroscopy. This method has the following advantages: it is noninvasive, highly accurate, and pollution-free, and the measurement methodology does not affect the normal growth of foxtail millet plants.

There are two key factors that guarantee the accuracy and reliability of the predication model. One is that the leaf detector apparatus can decrease the effect of environmental illumination and ensure the uniformity of experimental conditions. The other is the adoption of a set of appropriate data processing methods.

The WT-PCA-BP₅ model for the identification of varieties has a relatively simple structure and a higher running speed owing to the small number of hidden-layer nodes. WT-PCA

processing guaranteed the model accuracy, removed redundant information, and extracted feature spectra.

The extracted feature spectra, in the order from the most important to the least important, were 1440, 1660, 1775, 550, 410, 980, 1180, and 462 nm. There are five characteristic spectra in the near-infrared spectrum region. The CH_a ($a = 1, 2, 3$) and $\text{C}=\text{C}$ groups played an important role in distinguishing the varieties of foxtail millet. There are three characteristic spectra in the visible spectral region. The spectral information of chlorophylls a and b play a key role in identifying the varieties of foxtail millet.

However, the robustness and adaptability of the model will require further validation before it can be used in the actual production of the detector. In future work, we will focus on two primary issues: we will increase the sample size to verify the feasibility of the prediction model, and we will look at developing an inexpensive instrument for identifying foxtail millet varieties.

Acknowledgments

We thank all our friends for supporting us during our lab experiments. We extend our thanks to the College of Engineering, Shanxi Agricultural University, Taigu, for supporting and providing research facilities for this study. We also thank the National Key R&D Plan (2018YFD0700300) and Jinzhong City science and technology project (Y192007) for funding this research work.

Conflict of Interest

The authors declare no conflicts of interest for this study.

References

- 1 Y. W. Cui: *J. Agr. Sci. Technol.* **10** (2008) 54.
- 2 Y. P. Chi, Y. Wang, X. P. Nan, and Z. N. Yang: *Agr. Food Products Sci. Technol.* **4** (2010) 52. http://qikan.cqvip.com/Qikan/Article/Detail?id=33237995&from=Qikan_Search_Index
- 3 J. Y. Xiang and Z. M. Wang. *J. Handan Agri. Coll.* **16** (1999) 1. <https://www.ixueshu.com/do>
- 4 J. Fernández-Navales, J. Tardaguila, S. Gutiérrez, M. Marañón, and M. P. Diago: *Biosyst. Eng.* **165** (2018) 47. <https://dx.org/10.1016/j.biosystemseng.2017.08.018>
- 5 T. Sun, H. R. Xu, and Y. B. Ying: *Spectrosc. Spectral Anal.* **29** (2009) 122. [https://dx.org/10.3964/j.issn.1000-0593\(2009\)01-0122-05](https://dx.org/10.3964/j.issn.1000-0593(2009)01-0122-05)
- 6 P. Williams and K. Norris: *Irish J. Food Sci. Technol.* **12** (1988) 103. <https://www.jstor.org/stable/25558183>
- 7 K. Maertens, P. Reyns, and J. D. Baerdemaeker: *Trans. ASAE* **47** (2004) 1135. <https://search.proquest.com/docview/49200641?accountid=50700>
- 8 D. S. Long, R. E. Engel, and F. M. Carpenter: *Crop Manage.* **4** (2005) 1. <http://dx.doi.org/10.1094/CM-2005-0527-01-RS>
- 9 D. S. Long and T. Rosenthal: *Precis. Agr.* **5** (2005) 385.
- 10 J. M. Montes, H. F. Utz, W. Schipprack, B. Kusterer, J. Muminovic, C. Paul, and A. E. Melchinger: *Plant Breed* **125** (2006) 591. <https://search.proquest.com/docview/925434384?accountid=50700>
- 11 Y. Zhang, Y. T. Zhang, Z. H. Wang, and W. H. Min: *Food Ind.* **36** (2015) 242. http://qikan.cqvip.com/Qikan/Article/Detail?id=664684528&from=Qikan_Search_Index
- 12 X. Liu, S. Q. Jia, C.Y. Wang, Z. Liu, J. C. Gu, W. Zhai, and D. An: *Spectrosc. Spectral Anal.* **39** (2015) 2510. <https://www.ncbi.nlm.nih.gov/pubmed/26>
- 13 L. J. Zhang, L. L. Liu, M. C. Ma, and P. Zhou: *J. Anhui Agr. Sci.* **45** (2017) 10. <https://dx.org/10.13989/j.cnki.0517-6611.2017.08.004>

- 14 X. P. Han, Y. M. Zuo, and L. Z. Li: *Spectrosc. Spectral Anal.* **30** (2010) 2479. [http://dx.org/10.3964/j.issn.1000-0593\(2010\)09-2479-05](http://dx.org/10.3964/j.issn.1000-0593(2010)09-2479-05)
- 15 F. H. Wang, H. L. Zhu, and Z. Y. Ge: *Agr. Eng.* **1** (2011) 56. http://qikan.cqvip.com/Qikan/Article/Detail?id=678897169201101020&from=Qikan_Search_Index
- 16 The National Pharmacopoeia Commission of the People's Republic of China. *Pharmacopoeia of the People's Republic of China: Part Two*, Beijing: Chemical industry Press, 2000.
- 17 S. G. Wang: *Plant Physiology and Biochemistry* (China Agricultural Press, Beijing, 2001) p. 105.
- 18 Y. J. Ding, M. Z. Li, L. H. Zheng, R. J. Zhao, X. H. Li, and D. K. An: *Spectrosc. Spectral Anal.* **31** (2011) 2936. [https://doi.org/10.3964/j.issn.1000-0593\(2011\)11-2936-04](https://doi.org/10.3964/j.issn.1000-0593(2011)11-2936-04)
- 19 H. W. Siesler, Y. Ozaki, S. Kawata, and H. M. Heise: *Near-Infrared Spectroscopy-Principles, Instruments, Applications* (Wiley-VCH, Germany, 2002) pp. 335–339. <https://www.doc88.com/p-1963430>
- 20 R. Ríos-Reina, D. L. García-González, R. M. Callejón, and J. M. Amigo: *Food Control* **89** (2018) 108. <https://doi.org/10.1016/j.foodcont.2018.01.031>
- 21 *The Feisi R&D Center of Technological Products: Neural Network Theory and Matlab7 Application* (Publishing House of the Electronics Industry, Beijing, 2005).