Comparison of Near-Infrared Spectrum Pretreatment Methods for Jujube Leaf Moisture Content Detection in the Sand and Dust area of Southern Xinjiang

BAI Tie-cheng[1,2], WANG Tao[2], CHEN You-qi[3], MERCATORIS Benoît[1*]

1. TERRA Teaching and Research Centre, Biosystems Dynamics and Exchanges, Gembloux Agro-Bio Tech, Liège University, 5030, Gembloux, Belgium
2. College of Information Engineering, Tarim University, Alaer 843300, China
3. Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China

Abstract Precision irrigation for jujube crop in southern Xinjiang, China, is underlying to optimize the water use in such a drought-affected region. Water stress can be remotely assessed by evaluating the leaf moisture content using spectroscopy. These measurements are however affected by the presence of coarse sand and dust of the leaves induced by dry climates. This paper studied different methods to correct the spectral data in order to reduce the scattering noise with a baseline induced by such a jujube leaf covering. The reflectance of 120 leaf samples were measured by means of a near-infrared spectrometer (1 000～1 800 nm) and their moisture content was obtained by conventional drying method. The original reflectance spectrums were pre-processed by the normalization method, the moving smoothing method, the Savitzky-Golay (SG) convolution smoothing method, the SG first derivative method, the standard normal variables (SNV) method and the multiple scatter correction (MSC) method. The results of these different methods were compared and analyzed by means of partial leastsquares regressions (PLSR) allowing selecting sensitive spectral bands and establishing prediction models. The results showed that a significant reflectance peak related to the water content of the jujube leaves was located at 1 443 nm and that a local minimum of reflectance occurred at 1 661 nm. The predicition model based on the MSC method presented the best scattering noise reduction. The model performances were $R^2 = 0.750\ 4$, RMSEP=0.034 3 and RMSEPCV=0.021 5. The five characteristic wavelengths were 1 002, 1 383, 1 411, 1 443 and 1 661 nm. In this experiment, the MSC method had a good ability to reduce the scattering noise generated by sand and dust covering. The preprocessing improved the selection ability of characteristic wavelengths and the accuracy of the prediction model. The results can therefore provide an effective detection method for the jujube leaf water in the sandy and dusty area of Southern Xinjiang, China.

Keywords Leaf moisture content; Near-infrared spectrum; Scattering noise; Pretreatment methods; Multiple scatter correction

中图分类号: S66 **文献标识码**: A **DOI**: 10.3964/j.issn.1000-0593(2019)04-1323-06

Received: 2018-03-24; **accepted**: 2018-07-29

Foundation item: National Natural Science Foundation of China (61501314, 41561088), Principal fund of Tarim University (TDZKGG201502)

Biography: BAI Tie-cheng, (1982—), PhD at Liège University e-mail: baitiecheng1983@163.com

* Corresponding author e-mail: benoit.mercatoris@uliege.be

Introduction

Southern Xinjiang is one of the most important jujube production areas in China. This region however suffers from water shortage due to severe drought. The real-time monitoring of water stress conditions in jujube fields is a crucial step to improve the water use efficiency. In this context of precision irrigation, the leaf moisture content is considered as an important trait to assess the water status of jujube trees at different growing stages. Detection methods based on near-infrared (NIR) spectroscopy allow evaluating physiological traits of leaves in a fast and non-invasive way. Such methods have been studied and validated for the evaluation of rice leaf nitrogen content[1], corn, populus euphraticaand walnut leaf water content[2-4], cotton canopy chlorophyll content[5], sharp tooth oak leaf pigment content[6], total polyphenol in fresh tea leaf[7], persimmon leaf chloride contents[8] and leaf traits[9]. They mainly include three steps: (1) the pre-treatment of the reflectance spectrum, (2) the selection of characteristic wavelength bands and vegetation indexes, and (3) the development of regression models. Regardless of the spectral indexes selection and the modeling method, the data preprocessingremainsan underlying step to improve the detection accuracy. Errors can be indeed induced during the measurement process due to background, instrument and scattering noises. For in-field acquisition, the results can be impacted by other noise sources, such as the blade morphology, the impurities on leaf surface, the sensor-leaf distance, and the instrument jitter. Common pre-treatment methods including normalization, moving-average[10], convolution smoothing[11], Savitzkky-Golay (SG) derivative[12], Standard Normal Variate (SNV) correction[13], and Multiple Scattering Correction (MSC)[14] are now available to correct the results. Although some of these pre-treatment methods, such as derivative[5, 15], MA[17], SG[3, 17], SNV[16] and MSC[17] methods have been analyzed for the leaf water content detection in cotton, wheat, corn and lettuce, only few studies have been reported for jujube. More specifically, the dry weather of Southern Xinjiang, China, results in a large amount of sand and dust covering the leaf surface of jujube trees, which particularly affects the spectral response of the leaves. Therefore, the ability of different pretreatment methods to reject the scattering noise needs to be further studied for the measurement of the moisture content of jujube leaves in such conditions. This article was focusing on the spectral response of jujube leaves in order to evaluate their water content. The reduction of spectral noise, in particular the scattering noise caused by sand and dust on leaves, was analyzed and compared for different pretreatment methods. The comparison was performed by looking at the selection of the sensitive wavelength bands obtained by a Partial Least Square Regression (PLSR) method, the decision coefficient (R^2) and the root mean square error (RMSE) of the model. The results can provide some references for jujube leaf moisture content detection and remote sensing monitoring in sandy and dusty areas.

1 Materials and methods

1.1 Data collection

Experimental data were collected in a jujube orchard located in Alaer, Southern Xinjiang, China (40°34′47″N, 81°13′13″E). A set of 120 leaf samples of jujube trees were collected at the emergence, flowering, fruit development and maturity stages from the upper, middle and lower canopy layers. The portable near-infrared spectrometer SupNIR-1520 (Focused Photonics lnc, China) was used to measure the reflectance of jujube leaves. The spectrometer operated in the range of 1 000～1 800 nm with a resolution of 10 nm. The noise was less than 50 μA and the scattered light was less than 0.1%. A standard white board was used as a reference to calibrate the device. Each measurement was averaged on 10 repetitions. The acquisition was controlled using the RIMP software. The indoor temperature was 25 ℃ and humidity 60%. The reference measurement of moisture content was measured by drying the leaves at 105 ℃ until a constant weight was reached. Among the 120 samples, 90 were used for the regression and the remaining was used for validation.

1.2 Spectral pre-treatment

Normalization, moving-average, SG convolution smoothing, SG derivative, SNV and MSC were considered as pre-treatment methods to reduce the noise and correct the reflectance spectrum. A PLSR method was used for selecting sensitive wavelengths and extracting a model. The pretreatment methods were evaluated by considering the pre-processing results, the sensitive wavelength selection as well as the model performance (R^2 and RMSE). The pre-treatment of spectral data, wavelength selection and model establishment were implemented in MATLAB2010b.

2 Results

2.1 Leaf moisture content

The statistical information of the leaf moisture content for the 120 samples was shown in Table 1.

2.2 Leaf reflectance spectrum

The original spectra of jujube leaf reflectance, acquired for different water content, showed a significant scattering noise with a baseline drift [Fig. 1(a)]. Despite this scattering, the spectral curves presented a similar tendency. A peak occurred around 1 450 nm, which corresponded to a strong reflectance band of water[15]. A local minimum was also observed near 1 650 nm for all the curves. There was significant scattering noise with baseline drift in the range of 1 000～1 400 and 1 500～1 800 nm. The normalized method reduced significantly the scattering noise in the range of 1 500～1 800 nm but did not improve the curves from 1 000 to 1 400 nm [Fig. 1(b)]. Moving window smoothing [Fig. 1(c)] and SG convolution smoothing [Fig. 1(d)] had no effect on the original spectral curves. The SG first derivative method seemed to not lead to relevant results [Fig. 1(e)]. The SNV [Fig. 1(f)] and MSC [Fig. 1(g)] pre-treatment rather significantly reduced the scattering noise and baseline drift allowing extracting the curves characteristics.

Table 1　Leaf moisture content

Statistical parameter	Maximum	Minimum	Average	Standard deviation
Leaf moisture content	0.702	0.277	0.538	0.089

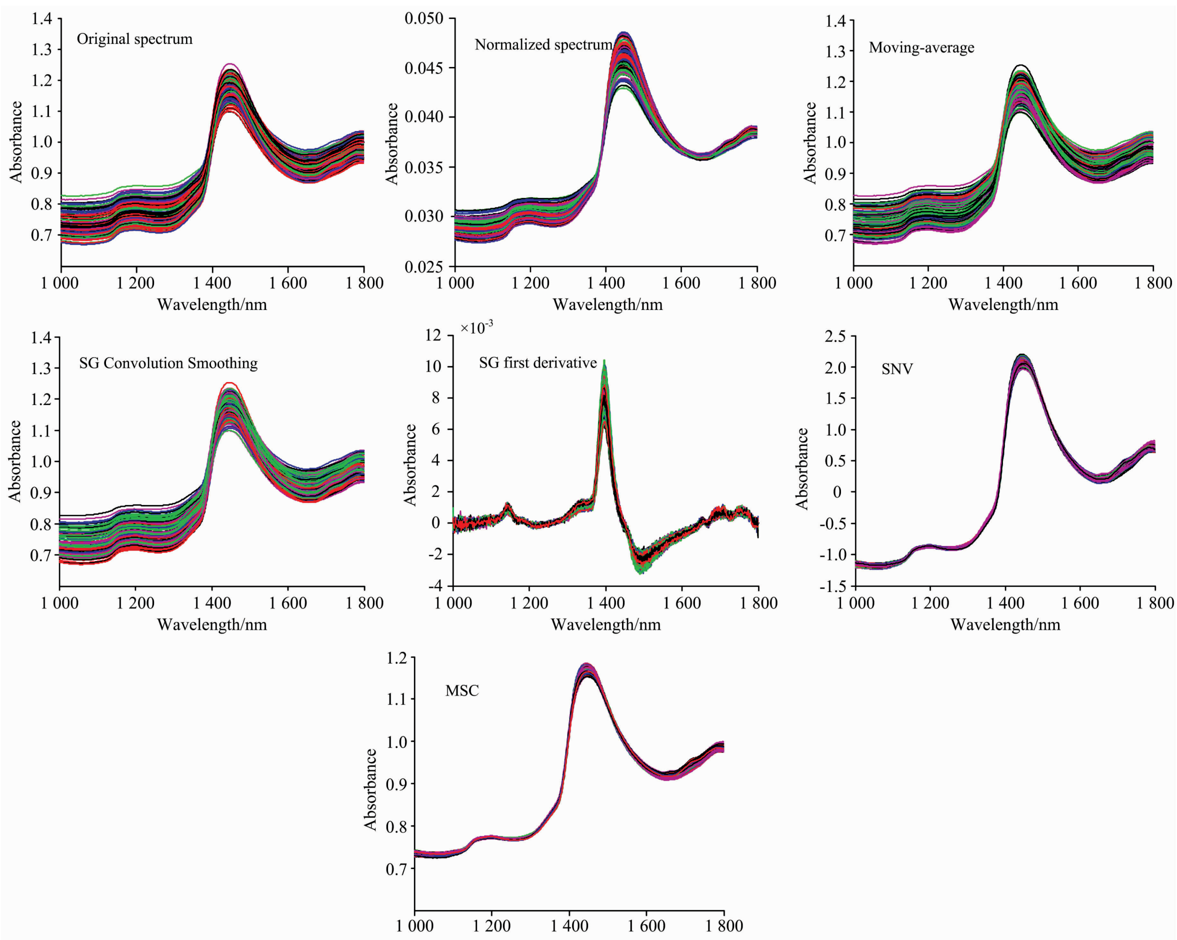

Fig. 1　Pre-treatment of the original spectrum

2.3　Characteristic wavelength selection

A principal component analysis was performed to select the characteristic wavelength bands (PLS has indicated that the optimal numbers was 5 in terms of R^2). This selection was shown in Fig. 2 for each considered pre-treatment method. The bands near the strong wave peakd at 1 443 nm and the local minimum at 1 661 nm were not selected based on the original spectrum [Fig. 2(a)]. There was a certain deviation in the selected bands around 1 443 nm according to the normalized spectrum [Fig. 2(b)], and the typical local maximum position was not selected. The selected bands around 1 661 nm were significantly higher than 1 700 nm. Moving average smoothing [Fig. 2(c)] and SG convolution smoothing [Fig. 2(d)] had good selected capability around 1 443 nm, but weak around 1 661 nm. There was the difference in the selection of the typical peak (1 443nm) and local minimum (1 661 nm) positions due to the deviation of the first derivative spectrum [Fig. 2(e)]. Both SNV [Fig. 2(f)] and MSC [Fig. 2(g)] selected sensitive spectral bands near the position of local maximum and minimum, where MSC method happened to be

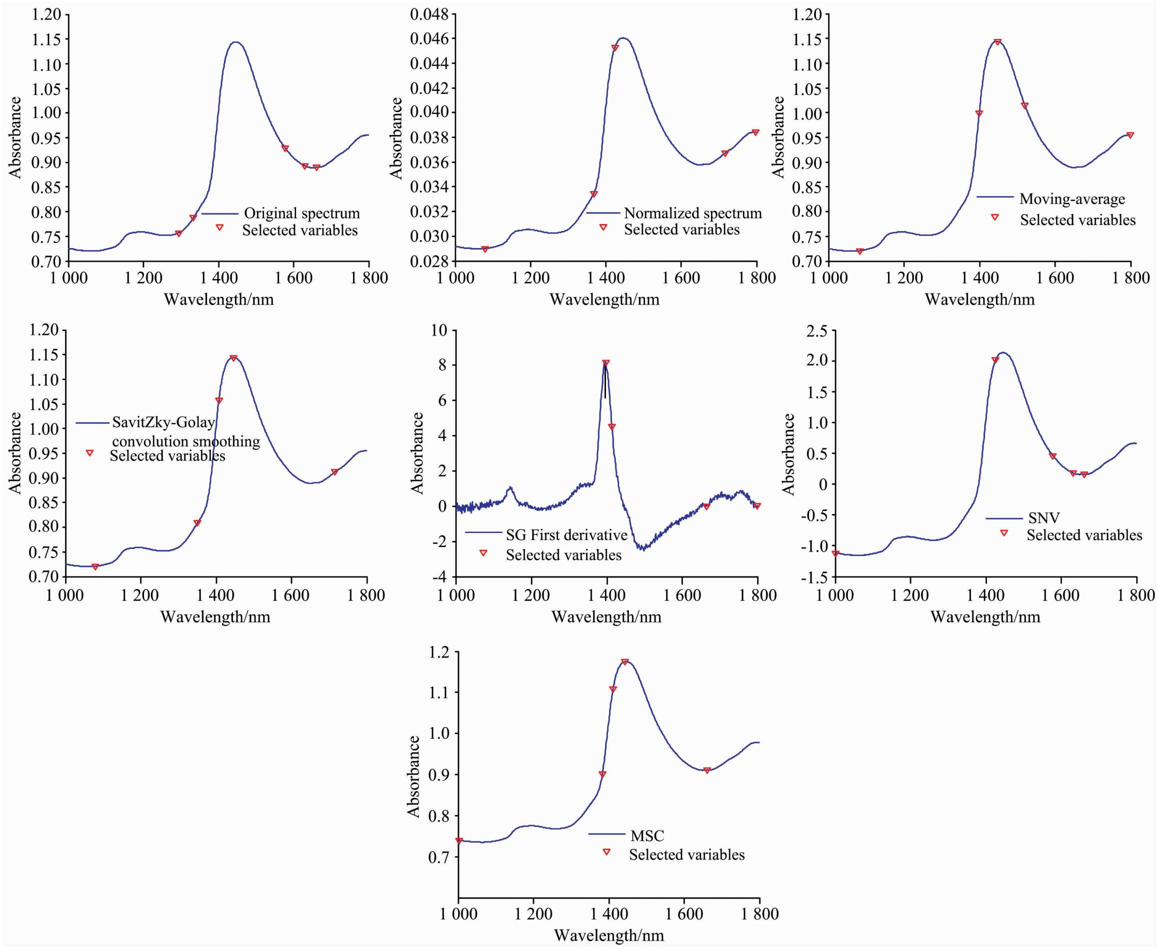

Fig. 2 Selected variables from different pre-treatment methods

at the position of wave peak and local minimum so the characteristic wavelength was slightly superior to SNV method.

Based on the principal components analysis of the preprocessed spectral data, the MSC pre-treatment method presented the best performance to reduce the scattering noise this preprocessing. The principal features of the spectrum were 1 002, 1 383, 1 411, 1 443 and 1 661 nm, among which, the latter two bands could be considered as the main sensitive wavelength to estimate the water content of jujube leaves.

2.4 Prediction model

PLSR method was used to establish prediction models based on different pre-treatment methods. The different models were evaluated by using the pre-diction determination coefficient (R^2), the prediction root mean square error (RMSEP), and the cross validation root mean square error (RMSEPCV). The prediction results based on the studied pre-treatment methods were shown in Table 2. The RMSEP values obtained for the different models were small, especially for the models based on the Normalized, SNV and MSC methods which reached an error inferior to 4%. The best determination coefficients were 0.71 and 0.75 obtained respectively with the SNV and MSC methods. With respect to the cross-validation, the SG convolution smoothing and the MSC methods resulted in a RMSE lower than 3%. It was noticed that the MSC method presented the best values for the three statistical indicators.

Table 2 Statistical results of prediction models based on different spectral pre-treatment methods

Pre-treatment methods	Leaf moisture content		
	R^2	RMSEP	RMSEPCV
None	0.510 4	0.040 7	0.067 9
Normalized	0.627 4	0.036 8	0.041 4
Moving average smoothing	0.595 0	0.042 5	0.046 7
SG convolution smoothing	0.680 4	0.043 6	0.024 3
SG first derivative	0.620 4	0.040 7	0.031 3
SNV	0.708 9	0.036 7	0.040 7
MSC	0.750 4	0.034 3	0.021 5

3 Conclusions

In dry regions such as Southern Xinjiang, the leaves of vegetation are often covered by sand and dust particles leading to significant scattering noise in spectral measurements. The denoising ability of different pre-treatment methods were analyzed and compared by means of the sensitive wavelength selection and the performance of prediction models. The results showed that both the SNV and MSC methods presented a good rejection of scattering noise and baseline drift. In particular, the spectral pre-processing based on the MSC method obtained the best performance for the prediction of the jujube leaf moisture content. This method allowed selecting a set of five characteristic wavelengths (1 002, 1 383, 1 411, 1 443, 1 661 nm) of which the strong leaf reflectance peak and the local minimum was identified, 1 443 and 1 661 nm, respectively. The R^2, RSMEP and RMSEPCV of the prediction model based on the MSC pretreatment was 0.750 4, 0.034 3 and 0.021 5, respectively. The results can provide an effective method for non-destructive testing of jujube leaf moisture content in sandy and dusty areas of Xinjiang. Since the number of samples was relatively small, the prediction accuracy of the model couldbe improved by adding jujube leaf samples with different moisture gradient in future work. In addition, the unknown random noise and background noise in the sample will also show diversity with the difference of measurement environment and instrument equipment. More effective eliminating methods also need to be further studied.

References

[1] Wang Shuwen, Zhao Yue, Wang Lifeng, et al. Transactions of the Chinese Society of Agricultural Engineering, 2016, 32(20): 187.

[2] Bai Tiecheng, Wang Yaming, Zhang Nannan, et al. Spectroscopy and Spectral Analysis, 2017, 37(11): 3419.

[3] Li Minzan, Sun Hong, Yang Wei, et al. Transactions of the Chinese Society of Agricultural Engineering, 2017, 33(Supp. 1): 137.

[4] Hu Zhenzhu, Pan Cunde, Pan Xin, et al. Scientia Silvae Sinicae, 2016, 52(12): 39.

[5] Chu Wanlin, Qi Yanbing, Chang Qingrui. Journal of Northwest A&F Unibersity(Nat. Sci. Ed.), 2016, 44(9): 65.

[6] Liu Chang, Sun Pengsen, Liu Shirong. Forest Research, 2017, 30(1): 88.

[7] Ajanto K Hazarika, Somdeb Chanda, Santanu Sabhapondit, et al. NIR News, 2018, 29(1): 9.

[8] PazJosé, Visconti Fernando, Chiaravalle Mara, et al. Anal. Bioanal. Chem., 2016, 408(13): 3537.

[9] Maria F Buitrago, Thomas A Groen, Christoph A, et al. International Journal of Applied Earth Observation and Geoinformation, 2018, 69: 237.

[10] Anastassia Baxevani, Krzysztof Podgórski. Communications in Statistics-Theory and Methods, 2018, 48(2): 448.

[11] Nikonov A V, Davletshin R V, Iakovleva N I, et al. Journal of Communications Technology and Electronics, 2017, 62(9): 1048.

[12] Tong Peijin, Dua Yiping, Zheng Kaiyi, et al. Chemometrics and Intelligent Laboratory Systems, 2015, 143: 40.

[13] Li Wenbing, Lu Jidong, Dong Meirong, et al. Energy & Fuels, 2018, 32(1): 24.

[14] Taddeuccia T N, Haighta R C, Leea H Y, et al. Nuclear Data Sheets, 2015, 123: 135.

[15] Liang Liang, Zhang Lianpeng, Lin Hu, et al. Scientia Agricultura Sinica, 2013, 46(1): 18.

[16] Sun Jun, Cong Sunli, Mao Hanping, et al. Transactions of the Chinese Society of Agricultural Engineering, 2017, 33(5): 178.

[17] Zhang Qianxuan, Li Qingbo, Zhang Guangjun. Spectroscopy: An International Journal, 2012, 27(2): 93.

南疆沙尘区骏枣叶片水分含量检测的近红外光谱预处理方法对比

白铁成[1,2]，王　涛[2]，陈佑启[3]，MERCATORIS Benoît[1*]

1. TERRA Teaching and Research Centre，Biosystems Dynamics and Exchanges，Gembloux Agro-Bio Tech，Liège University，5030，Gembloux，Belgium
2. 塔里木大学信息工程学院，新疆 阿拉尔　843300
3. 中国农业科学院农业资源与农业区划研究所，北京　100081

摘　要　南疆地区沙尘多、灰尘大，枣树叶片表面经常覆盖一定程度的粗颗粒度沙尘，为了有效去除沙尘、灰尘在枣树叶片水分光谱测量过程中产生的散射噪声和基线漂移，研究一种适用于风沙较大地区的枣树叶片水分含量的快速检测方法，以不同灌溉梯度下的枣树叶片为研究对象，通过近红外光谱仪获取 120 个叶片样本的 1 000～1 800 nm 的光谱数据，并同步测量叶片水分含量，采用归一化、移动窗口平滑、SavitZky-Golay(SG)卷积平滑、SG 求导、标准正态变量校正(SNV)和多元散射校正(MSC)等方法对原始光谱进行预处理，分析对比不同方法对散射噪声的处理能力，采用偏最小二乘回归分析方法筛选了敏感波段和建立预测模型。实验结果表明，枣树叶片水分含量强吸收峰为 1 443 nm，波谷为 1 661 nm；归一化光谱并未消除 1 000～1 400 nm 波段的散射噪声；移动窗口平滑和 SG 卷积平滑并未改进光谱曲线，散射噪声仍然存在；SG 导数光谱的光谱特征峰和特征谷明显左移，光谱曲线不够平滑，噪声明显；SNV 和 MSC 方法具有较好的散射噪声消除能力。偏最小回归分析方法筛选特征波长的结果表明(设置筛选波长数量为 5)，基于原始光谱未筛选到 1 443 nm 的强波峰和 1 661 nm 的波谷附近的波段；基于归一化光谱在 1 450 nm 波峰附近筛选的波长有一定的偏差，在 1 661 nm 波谷附近的筛选的波长明显高于 1 700 nm；基于移动窗口和 SG 卷积平滑光谱在 1 443 nm 具有一定的筛选能力，但并未筛选到1 661 nm 附近的波长；导数光谱并未筛选到 1 443 和1 661 nm 波段；SNV 和 MSC 在波峰和波谷位置附近均筛选了敏感的光谱波段，其中 MSC 略优于 SNV 方法恰好在波峰和波谷位置，共筛选了 1 002，1 383，1 411，1 443 和 1 661 nm 五个特征波段，也证明了 MSC 方法散射噪声和基线漂移处理能力最优，提高了敏感波长的筛选能力。偏最小二乘回归模型结果表明，不同预处理方法的 RMSE 值均较低，SNV 和 MSC 方法改进了模型的预测结果，R^2 高于 0.7，其中基于 MSC 方法的模型具有最高的 R^2 和最低的 RMSEP 和 RMSEPCV，$R^2=0.750\ 4$，RMSEP$=0.034\ 3$，RMSECV$=0.021\ 5$，预测结果较优。证明 MSC 方法对沙尘和颗粒度引入的散射噪声具有较好的去除能力，可改进波长的筛选、提高预测模型精度，为南疆沙尘区的枣树叶片水分含量的无损检测提供了有效方法。

关键词　叶片水分含量；近红外光谱；散射噪声；预处理方法；多元散射校正

（收稿日期：2018-03-24，修订日期：2018-07-29）

* 通讯联系人