Wavelet Denoising Method Research of Soybean Straw Cellulose Near Infrared Rapid Detection

Weizheng Shen, Nan Ji, Haoran Du, Hongbin Li, Sida Ma and Qingming Kong*

School of Electronic Engineering and Information Northeast Agricultural University

Harbin, 150030, China

E-mail: wzshen@neau.edu.cn

** Corresponding author*

Abstract

In this paper, we made a research for soybean straw hemicellulose rapid detection by establishing a quantitative analysis model based on near-infrared spectroscopy. At first, 146 samples were collected from varieties of soybean straws are gathered in different areas of Heilongjiang province, then made chemical testing of components and spectral scanning to soybean straw, the 140 samples were classified to two groups, in which 100 samples were chosen as calibration set and the remaining 40 samples were chosen as verification set. Wavelet transform was used to deal with the noise spectrum, selected DBN wavelet, Haar wavelet and Symlet wavelet in different layers under penalty threshold, Bridge-massart threshold, and default global threshold for spectral signal decomposition and reconstruction, compared with other traditional noise reduction methods, Symlet2-2 layer decomposition wavelet basis for hemicellulose spectral processing possessed better effect with the determination coefficient of validation set rising from 0.462524 to 0.6314158 after processing.

Keywords: *Near-infrared; DBN wavelet; Haar wavelet; Symlet wavelet*

1. Introduction

As the agricultural country, there are a large number of crop residues in China, such as corn, sugar cane, rice, soybean, wheat, sugar beet, potato, cassava, rapeseed, cottonseed, etc, however, the utilization of the crop straws in China is less than 40% each year. Recently, many scholars at home and abroad are working on how to improve biofuel technology processes and technology tools to maximize the quality and the quantity of biofuels, in which the quantitative analysis of crop straw components, precise matching of raw material will be significant to improve biomass yield and quality[1-4]. At present, the main component detection of straw are based on chemical methods, and the detection cycle is about seven days with cumbersome testing process, high costs, and a lot of manpower and resources[5,6]. Therefore, in this paper we proposed to use near infrared spectroscopy technology to achieve the rapid detection of components by establishing the model between straw components and spectrum, then achieved low-cost, high-precision, online composition detection for plant straw to promote the development of biomass energy industry. In the paper, we made a research for near infrared spectrum de-noising by establishing a model with feature selection results in different kinds of wavelet basis and threshold methods for different de-noising methods.

2. Theory

In addition to containing its own information, NIR also contains many other irrelevant information and noise, for example, electrical noise, sample background interference, stray light and so on. Therefore, during the model-correcting and model-building, how to eliminate irrelevant information and spectral data and noise has become very critical pretreatment solution.

2.1 Principle of Wavelet De-noising

Gabor made his famous Gabor transform, which developed into the short time Fourier transform later. The basic idea is to create a window function, then the signal windowing function analysis, and then the Fourier transform of the window signal. Because the transformation can reflect local characteristics of the signal, STFT has been popularized in practice. With the development of science and technology, due to the size and shape of STFT window function, people encountered bottlenecks in the signal processing. At this point, Wavelet transform emerged. The size of Wavelet transform window is unchanged, but the shape of the window can be changed. So it is frequency analysis whose time window and frequency window can be changed. At low frequencies, it has a relatively high frequency resolution and low temporal resolution, while high-frequency part enjoys a high time resolution and a low frequency resolution. In the time domain and frequency domain, it can be a good characterization of the local features of the signal[7-9].

The basic model of the noisy signal is generally expressed as follows:

$$s(n) = f(n) + \sigma e(n) \quad (2-1)$$

$f(n)$ is the true signal, $e(n)$ is the noise, σ is defined as noise intensity, $s(n)$ is the noisy signal. We assume that $e(n)$ is the Gaussian white noise and σ is 1. The purpose of the noise cancellation is to suppress $e(n)$ and recover the true signal $f(n)$.

2.2 Wavelet De-noising Threshold

Johnstone and professor Donoho of the Stanford University proposed wavelet de-noising threshold in 1992. It is a non-linear de-noising method, different from the previous wavelet de-noising. It can be approximated optimal at the minimum mean square error sense. The de-noising method is practical and simple.

Wavelet decomposition of signal, select a wavelet bases, and determine the wavelet decomposition scale N. Then decompose the signal into N scale wavelet. Decomposition process is shown in Figure(such as three scale decomposition). Noise is mainly in the high-frequency coefficients of wavelet decomposition. Low-frequency coefficients of wavelet decomposition is mainly including energy signal.

2.3 Determination of Threshold Method in the Wavelet De-noising

In threshold wavelet transform, the threshold value is mainly determined by the following several mathematical models.

① The default threshold values:

$$th = \sqrt{2 \lg n} \sigma \quad (2-2)$$

n is the length of the signal itself.

The threshold of Birge-Massart determines Birge-Massart strategy obtained the determined threshold by the following rules.

For a given wavelet decomposition scale j , $j+1$ layer and the higher layers, we should retain all the wavelet coefficients. For the i -layer wavelet coefficients, we retain the maximum absolute value n_i . n_i is determined by the following equation.

$$n_i = M(j+2-i)^\alpha \quad (2-3)$$

Formula M and α are empirical coefficients. When it is default, we take $M = L(1)$. This is the length of the first layer after decomposition coefficients. Generally, M is in $L(1) \leq M \leq 2L(1)$. The α has different values due to different uses. When reduce noise, α is 3.

② Penalty threshold

So t^* is the value which causes the function $crit(t) = -\sum_{k \leq t} c_k^2 + 2\sigma^2 t(\alpha + \log(n/t))$ to obtain the minimum value of t . c_k is the first large coefficient values in the wavelet coefficients. σ is noise intensity. α is the empirical coefficient. Penalty global threshold expression.

3. Material and Methods

3.1 Collection and Preparation of Samples

(1) The Sample Collection

Collected from Heilongjiang province, a total of 146 varieties of soybean straw, each variety collection one individual plant straw, Label to be used, a total of 150 samples.

The method is as follows

① Under the condition of natural dried to remove the surface moisture.

② Using HC - 152 crusher soybean straw to pieces and by 1 mm sieve.

The mixing soybean straw powder respectively into two sealing bag, and label.

(2) Standard Chemical Calibration

Chemical calibration sample mainly aims at the main ingredients of soybean straw cellulose, hemicellulose, and lignin chemical calibration, ash content. The entire process takes about 35 h, boiling hydrolysis of 3h, filter 1h, drying 7h, ashing for 24 h[10], the whole process is shown in Fig1.

Figure 1. Preparation of Samples Using Van Soest Method

The ideal calibration sample set should satisfy the following conditions: correct the concentration of sample should be included all possible chemical composition of the sample under test in the future, the scope(or properties) of its concentration should be exceed in the future might have happened to the sample under test based on the determination of Van Soest method soybean straw hemicellulose. The Kennard-Stone algorithm[11,12], selecting the most representative samples as the calibration set, the conditions of sample distribution such as table.Can be seen from table1 that calibration sets and validation sets are uniform distribution, The samples of validation sets are within the scope of the calibrating samples, verification sets and calibration sets are very close to the average,That average is 14.59 and 14.74.

Table 1. Hemicellulose Samples Statistics of Soybean Straw

Sample/Classification	Component	Samples No.	Range/(%)	Mean/(%)
Soybean Straw	Hemicellulos	140	12.83-16.3 8	14.64
Calibration sets	Hemicellulos	100	12.83-16.3 8	14.59
Validation sets	Hemicellulos	40	13.35-16.0 3	14.74

Note: the value accounts for the percentage of the whole soybean straw quality

3.2 Acquisition of Spectra

The soybean straw samples were scanned twice with Thermo's Antaris II over the wavelength ranging from 830nm to 2500nm (12000cm^{-1} - 4000cm^{-1}) with a resolution of 8cm^{-1} for a total number of 32 scans in a single operation. The average spectra were then automatically recorded in scanning process. A number of samples were packed in a rotating sample cup as appropriate. The spectrum of soybean straw is shown in Fig2.

Figure 2. NIR Spectrum of Soybean Straw

In this article we respectively compared three kinds of wavelet denoising effect, that are DBN series wavelet, Haar wavelet and Symlet wavelet.

4. Results and Discussion

4.1 DBN Series Wavelet

All of the signal of wavelet decomposition are based on the length of the signal to determine the maximum number of decomposition layers, Using wavelet toolbox, by matlab statements $N = \text{wmaxlev}(N, \text{'wname'})$. In this paper the calibration set sample length is 692 points, If we use the DB2 wavelet decomposition, the maximum number of layer is 7. That means the soybean straw spectral data in denoising decomposition layers can be used in 2, 3, 4, 5, 6, 7 different wavelet decomposition layers such as a table. For how many layers of decomposition to achieve the best effect, this paper has carried on the different decomposition layers and different threshold method of wavelet denoising, Figure 3, 4 respectively second, six floor DB2 wavelet decomposition, Selects the Brige-Massart layered threshold respectively, Penalty global threshold and the default threshold diagrams of the results of the three ways of data processing.

Using DB2 wavelet spectrum reconstruction, spectrum reconstruction as shown in Tab2.

Table 2. Decomposition Layers Statistics of Different Wavelet

Wavelet			Wavelet		
basis	Variable	Decompositio n	basis	Variable	Decomposition
DB2	692	7	Sym2	692	9
DB3	692	7	Sym3	692	7
DB4	692	6	Sym4	692	6
DB5	692	6	Sym5	692	6
DB6	692	5	Sym6	692	5
DB7	692	5	Sym7	692	2
Haar	692	9			

And can be seen from the diagram with the increase of decomposition layers. Choose Penalty to denoising threshold global threshold spectrum of a significant distortion, signal standard deviation with the increasing of decomposition layers.

Figure 3. 2 Layer Wavelet Denoising Results of DB2

Due to the DBN series can choose DB2-DB7 series wavelet, each of the wavelet is 2 to 7 layers can be broken down, and each decomposition to choose three types of threshold decomposition method, the reconstructed many spectrum are produced.

In this article we takes only DB2 wavelet spectrum reconstruction for example, spectrum reconstruction information as shown in Tab3.

Table 3. Standard Deviation of DB2 Wavelet between Original and Construction Signal

Wavelet basis	Decomposition	Threshold method	Standard deviation of signal
DB2	2	Penalty	0.02215
	2	Brige-Massart	0.01269
	2	Default	0.02941
DB2	3	Penalty	0.05806
	3	Brige-Massart	0.0137
	3	Default	0.0381
DB2	4	Penalty	0.1359
	4	Brige-Massart	0.0141
	4	Default	0.0421
DB2	5	Penalty	0.2616
	5	Brige-Massart	0.0142
	5	Default	0.0448
DB2	6	Penalty	0.4369
	6	Brige-Massart	0.0143
	6	Default	0.0465
DB2	7	Penalty	0.6024
	7	Brige-Massart	0.0144
	7	Default	0.0473

For there are too much data in DB3-DB6 choice has the lowest standard deviation of the signal decomposition. Table4 only show the best wavelet correction verification results.

Table 4. Standard Deviation of DB3-DB6 Wavelet between Original and Construction Signal

Wavelet basis	Decomposition	Threshold method	Standard deviation of signal
DB3	3	Penalty	0.0508
	3	Brige-Massart	0.0085
	3	Default	0.0358
DB4	4	Penalty	0.1143
	4	Brige-Massart	0.0069
	4	Default	0.0391
DB5	2	Penalty	0.0162
	2	Brige-Massart	0.0063
	2	Default	0.0247
DB6	2	Penalty	0.0162
	2	Brige-Massart	0.0055
	2	Default	0.0248

4. 2 Haar Wavelet Transform

Haar transform which is the easiest mean that can reflect the time-variant spectrum was first proposed in 1909 by the A.Haar[13]. It has the same design concept with same Fourier Transform.

The mother wavelet of Haar wavelet can be expressed as:

$$\psi(t) = \begin{cases} 1 & 0 \leq t < 1/2 \\ -1 & 1/2 \leq t < 1 \\ 0 & \text{otherwise} \end{cases} \quad (2-4)$$

Its corresponding scaling functions can be expressed as:

$$\phi(t) = \begin{cases} 1 & 0 \leq t < 1 \\ 0 & \text{otherwise} \end{cases} \quad (2-5)$$

Its filter is defined as:

$$h[n] = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } n=0, 1 \\ 0 & \text{otherwise} \end{cases} \quad (2-6)$$

when $n=0$ and $n=1$, it contains two nonzero coefficients, therefore, it can also be written as:

$$\begin{aligned} \frac{1}{\sqrt{2}} \psi\left(\left(\frac{t}{2}\right)\right) &= \sum_{n=-\infty}^{\infty} (-1)^{1-n} h[1-n] \phi(t-n) \\ &= \frac{1}{\sqrt{2}} (\phi(t-1) - \phi(t)) \end{aligned} \quad (2-7)$$

In all orthonormal wavelet transformation, Haar wavelet is one of the simplest

conversions, it yet is not suitable for smoother function, because it has only one vanishing moment.

The decomposition and reconstruction of the signal is shown in Tab5, for Harr wavelet, its decomposition level of the signal can be 2-9 layers, which can make Penalty threshold, Brige-Massart and default threshold processing respectively. As the data is numerous, therefore it is shown as tabular format. Recording to the signal standard deviation, Haar-2 Brige-Massart wavelets is chosen.

Table 5. Standard Deviation of Harr Wavelet between Original and Construction Signal

Wavelet	Signal standard deviation			
Haar-2/ 3/ 4/ 5/ 6 /7/ 8/ 9 layer decomposition	0.0503	0.1093	0.2069	0.3254
Penalty Threshold	0.5914	0.8367	1.4908	2.0293
Haar-2/ 3/ 4/ 5/ 6 /7/ 8/ 9 layer decomposition	0.0324	0.0352	0.0363	0.0369
Brige-Massart	0.0371	0.0372	0.0372	0.0372
Haar-2/ 3/ 4/ 5/ 6 /7/ 8/ 9 layer decomposition	0.0431	0.0508	0.0552	0.0576
Default threshold	0.0587	0.0593	0.0595	0.0596

4. 3 Symlets Wavelet Transform

Symlets wavelet is the minimum frequency of the four kinds between Daubechies wavelet and Biorthogonal wavelet function system. Meanwhile Symlets wavelet is obtained by the improvement of Daubechies wavelet. Not only has Biorthogonality itself, but also has compact support and approximate symmetry. Symlet wavelet has also achieved good results in practical applications and is particularly advantageous in details[14,15].

Symlet wavelet has more small wavelets form as wavelet DBN series, has also become SymN series, in this paper we choose Sym2-Sym8 to demonstrate, since the data is large, ultimate expressing it with a trend graph form. Figure5 layers of signal decomposition and reconstruction result by Sym2 in selecting the global threshold, Bridge-massart threshold and default threshold conditions. Fig 4 is Sym2-Sym9 signal reconstruction standard deviation contrast.

Figure 4. Standard Deviation of SymN Wavelet between Original and Construction Signal

From the above figures, tables and theoretical analysis, it is known that Wavelet

de-noising results in different levels and different threshold chosen mode with the less decomposition level, after denoising the higher the proportion of signal energy in the original signal and the smaller the signal standard deviation with the original signal, the more well preserved early development of high-frequency characteristics of the signal. This is because the soybean straw is a solid powder, the near-infrared spectral absorption is not strong, so the wavelet decomposition level become small and denoising performance is relatively better. In summary, this object chooses Symlet2-2 layer decomposition Brige-Massart threshold denoising to denoise the hemicellulose. To verify the validation sample set by establishing calibration models and the results are shown in Table 7. The corrects validate model shown in Fig 5.

Table 7. Model Validation Results after Different Wavelet Denoising

Wavelet	Factor s	RMSECV	RMSEP	SEP	R ²
DB2-4 layer-Brige-Massart	7	0.4605151	0.489528 2	0.712034 8	0.6284816
Haar-2 layer-Default threshold	7	0.4563149	0.491613 8	0.685276 9	0.6253091
Symlet2-2 layer-Brige-Massart	8	0.4486225	0.487591 2	0.668319 8	0.6314158

Figure 5. Sym2-2 Layer Wavelet De-noising Cross Validation Results

5. Conclusions

We studied the application of near infrared spectrum denoising method in soybean straw hemicelluloses in this paper, and proposed a new method to get rid of the noise based on wavelet transform in signal processing area, made comparison in DBN, Harr, Symlet respectively in the condition of different decomposition layers and threshold methods, the results showed that the wavelet transform method possessed optimum effect with symlet 2-2 decomposing layers and Brige-Massart threshold method, the determinate coefficient of verification set R² rose from 0.4436922 to 0.6330684, RMSEP dropped from 0.608179 to 0.4486225. The research and application of this method provided a reliable theoretical foundation and technical support to model denoising and the formation of rapid detection system.

Acknowledgements

This research was supported by the National High-tech R&D Program of China(863 Program)(2013AA102303) and the Natural Science Foundation of Heilongjiang Province of China(F201402), the Key Technologies R&D Program of Harbin(2013AA6BN010), and Northeast Agricultural University Innovation Foundation For Postgraduate (yjscx14002).

References

- [1] S. Hou and L. Li, "Rapid Characterization of Woody Biomass Digestibility and Chemical Composition Using Near-infrared Spectroscopy", *Journal of Integrative Plant Biology*, vol. 53, no. 2, (2011), pp. 166-175.
- [2] C. J. Lomborg , M. H. Thomsen, E. S. Jensen and K. H. Esbensen, "Power plant intake quantification of wheat straw composition for 2nd generation bioethanol optimization", *A Near- Infrared Spectroscopy (NIRS) feasibility study*, *Bioresource Technology*, vol. 101, (2010), pp. 1199-1205.
- [3] D. Wang, F. Dowell and D. P. Chung, "Assessment of heat-damaged wheat kernels using near-infrared spectroscopy", *Cereal Chemistry*, vol. 78, no. 5, (2001), pp. 625-628.
- [4] D. W. Templeton, A. D. Sluiter, T. K. Hayward, B. R. Hames and S. R. Thomas, "Assessing corn stover composition and sources of variability via NIRS", *Springer Science+Business Media B.V.*, vol. 16, no. 4, (2009), pp. 621-639.
- [5] F. Xu, J. Yu, T. Tesso, F. Dowell and D. Wang, "Qualitative and quantitative analysis of lignocellulosic biomass using infrared techniques: A mini-review", *Applied Energy*, vol. 104, (2013), pp. 801-809.
- [6] H. Jiang, G. Liu, X. Xiao, C. Mei, Y. Ding and S. Yu, "Monitoring of solid-state fermentation of wheat straw in a pilot scale using FT-NIR spectroscopy and support vector data description", *Microchemical Journal*, vol. 102, (2012), pp. 68-74.
- [7] F. Xu, J. Yu, T. Tesso, F. Dowell and D. Wang, "Qualitative and quantitative analysis of lignocellulosic biomass using infrared techniques: A mini-review", *Applied Energy*, vol. 104, (2013), pp. 801-809.
- [8] S. Bruun, J. W. Jensen, J. Magid, J. Lindedam and S. B. Engelsen, "Prediction of the degradability and ash content of wheat straw from different cultivars using near infrared spectroscopy", *Industrial Crops and Products*, vol. 2, no. 321-326, (2010).
- [9] P. Kaparaju and C. Felby, "Characterization of lignin during oxidative and hydrothermal pre-treatment processes of wheat straw and corn stover", *Bioresource Technology*, vol. 101, (2010), pp. 3175-3181.
- [10] F. E. Dowell, D. Wang, X. Wu and K. M. Dowell, "Detecting the Antimalarial Artemisinin in Plant Extracts Using Near-infrared Spectroscopy", *American Journal of Agricultural Science and Technology*, vol. 2, no. 1, (2014), pp. 1-7.
- [11] C. Huang, L. Han, Z. Yang and X. Liu, "Prediction of heating value of straw by proximate data, and near infrared spectroscopy", *Energy Conversion and Management*, vol. 49, (2008), pp. 3433-3438.
- [12] F. Xu and D. Wang, "Rapid determination of sugar content in corn stover hydrolysates using near infrared spectroscopy", *Bioresource Technology*, vol. 7, (2013), pp. 293-298.
- [13] C. He, L. Chen, Z. Yang, G. Huang, N. Liao and L. Han, "A rapid and accurate method for on-line measurement of straw-coal blends using near infrared spectroscopy", *Bioresource Technology*, (2012), pp. 314-320.
- [14] G. W. Mathison, H. Hsu, R. S. Siawash, G. R. Diaz, E. K. Okine, J. Helm and P. Juskiw, "Prediction of composition and ruminal degradability characteristics of barley straw by near infrared reflectance spectroscopy", *Canadian Journal of Animal Science*, vol. 79, no. 4, (1999), pp. 519-523.
- [15] L. Lu, X. Philip, A. Ye, R. Womac and S. Sokhansanj, "Variability of biomass chemical composition and rapid analysis using FT-NIR techniques. *Carbohydrate Polymers*", vol. 81, (2010), pp. 820-829.

Authors

Weizheng Shen, (1977-), male, He is a Ph.D., professor, mainly engaged in the research and application of information technology in agriculture.