

Near-infrared Detection of Straw Cellulose by Orthogonal Signal Correction and Partial Least Squares

Qingming Kong, Zhongbin Su, Weizheng Shen, Mingming Han, Bohan Li and Xudong Wang

*School of Electronic Engineering and Information Northeast Agricultural University
Harbin, 150030, China
wzshen@neau.edu.cn*

Abstract

Orthogonal signal correction (OSC) and partial least squares (PLS) were used during the pretreatment of straw to reduce environmental noise and prediction models were established for near-infrared detection of straw cellulose. Tests were run with soybean stalk as the object of study. Test results indicated that compared to a model established using a traditional denoising method, the determination coefficients for calibration set models established by second derivative+smoothing and OSC were 0.9318595 and 0.9328905 respectively while the root mean square error for calibration (RMSEC) were 0.6762902 and 0.6696454. For an OSC-PLS regression model with a factor of 8, the relative standard deviation of a prediction model was less than 5%. In the OSC denoising process, the root mean square error fluctuated with the increasing number of PLS factors. Compared to the second derivative-smoothing denoising, OSC-PLS denoising removed the non-correlated variation from spectra and improved interpretation ability of variation while the analysis and convergence were expedited. It was therefore concluded that OSC-PLS denoising could be used to realize the rapid and accurate near-infrared detection of straw cellulose.

Keywords: *Near-infrared; Orthogonal Signal Correction; Partial Least Squares*

1. Introduction

As a traditional agricultural country, China has an annual output of more than 640 million tons of straw and stalk [1]. But most of them is burned or discarded [2]. It is therefore of great significance to find ways and means to make full use of these straw and stalk available, for example, to use them to produce bio-fuel and ethanol alcohol [3]. The chemical composition of biomass has a direct impact on the yield and quality of bio-fuel and ethanol alcohol produced. The existing methods used in day-to-day production are based on chemical determination of straw cellulose, hemicellulose and lignin, which are time-consuming and expensive. And so, it is very difficult to accurately determine the chemical composition of straw [4-9]. Near-infrared spectroscopy attracted much attention from the research community because of its high speed and accuracy and much work has been done in recent years on detection of crop residues. For example, Lu Liu evaluated two important biomass feedstock for different concentrations and put forward a broad-based hypothesis that a single FT-NIR prediction model could be developed as a promising candidate for analysis of different biomass feedstock [10]. Shen Hou and Laigeng Li studied poplar and eucalyptus materials by NIR spectrum and determined the chemical composition of Klason lignin, α -cellulose, holocellulose, lignin syringyl/guaiacyl (S/G) ratio and enzymatic digestibility. A group of trained NIR models were established by fitting the chemical properties and digestibility by partial least-squares (PLS) regression, which means NIR could be used for rapid characterization of biomass

properties [11]. Donghai Wang compared the chemical method with the NIR technology in detecting the chemical composition of biomass in different groups [12-15].

However, near-infrared spectra contains a lot of sample information as well as bandwidth overlapping, which has its effect on the qualitative analysis. So, it is very important to enhance signal specificity by reducing environmental noise and eliminating random errors. Therefore, Orthogonal signal correction (OSC) and partial least squares (PLS) were used during the pretreatment of straw to reduce environmental noise and prediction models were established for near-infrared detection of straw cellulose.

2. Theory

Orthogonal signal correction is a method first proposed by Svante which can be used for the establishment of quantitative calibration model with spectral matrix and chemical values arranged in an orthogonal array to filter out unrelated signals such as baseline drift, light scattering for multivariate calibration.

OSC method consists of the following:

- (1) Definition of Original spectral matrix X and density matrix Y
- (2) Assignment of the component orthogonal to Y as t
- (3) Calculation using $t_{new} = (1 - Y(Y'Y)^{-1}Y')t$ with $\omega = X - t_{new}$
- (4) Calculation of orthogonal principal component t , $t = X\omega$ using the new calculations to get X and ω
- (5) Verification of convergence when $\|t - t_{old}\| / \|t\| < 10^{-6}$ by switching to (8) or (5)
- (6) Computation of load vector $p' = t'X / (t't_{new})$
- (7) Subtraction of orthogonal component of Y from X , $E = X - tp'$
- (8) Calculation of orthogonal principal with E as new X until an optimal value is obtained. Usually, two orthogonal principal components are needed to achieve a given accuracy.
- (9) Establishment of model using X corrected by orthogonal signal as a new spectral matrix
- (10) Prediction of unknown samples with residual vector e after correction by orthogonal signal

$$t_1 = x_{new}'\omega \quad e_1' = x_{new}' - t_1p_1 \quad (1)$$

$$t_2 = e_1'\omega_2' \quad e_2' = e_1' - t_2p_2 \quad (2)$$

3. Material and Methods

3.1 Collection and Preparation of Samples

A total of 159 pieces of soybean stalk was collected from the soybean planting areas in Heilongjiang province in the northeastern part of China.

Fresh straw samples were dried for 48 hours in open air prior to NIRS scanning to determine dry matter content. All soybean stalk samples were milled with high-speed grinder for 1 minute to keep all these samples in the same condition. The crushed samples were sealed in ziplocked bags for future use. Neutral detergent fiber (NDF) and acid detergent fiber (ADF) were analyzed using Van Soest method as shown in Fig. 1 below. All chemical analyses were done in duplicate and expressed on a dry weight basis.

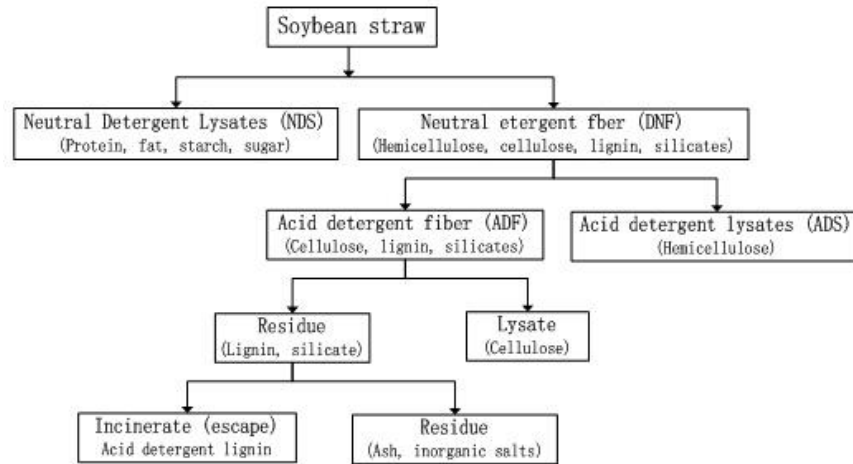


Figure 1. Preparation of Samples Using Van Soest

Table 1. Cellulose Content in Soybean Stalk

Sample components	Number	Min(%)	Max(%)	Average(%)	SD(%)
cellulose	159	37.7410	49.45694	43.1256	3.00

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.

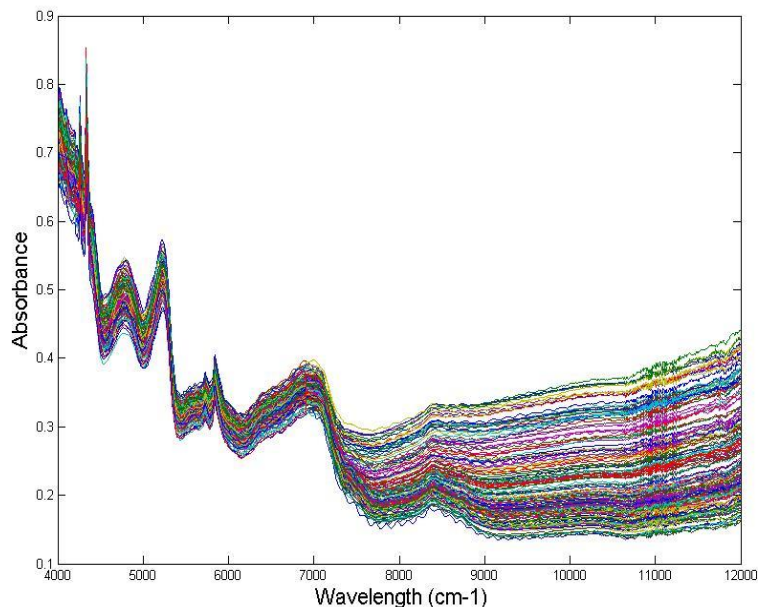


Figure 2. Absorption Spectra of Soybean Stalk

4. Results and Discussion

4.1 S-G Convolution Smoothing Algorithm

S-G Smoothing algorithm is a method used to do the multi-point processing of spectra using S-G function. Its main idea is in essence to use polynomials to fit data within the window by least squares. It is in essence a weighted average method, which can be expressed as shown below.

$$X_{k,smooth} = \bar{X}_k = \frac{1}{H} \sum_{i=-\omega}^{+\omega} x_{k+i} h_i \quad (3)$$

$$\text{And } H = \sum_{i=-\omega}^{+\omega} h_i$$

S-G Smoothing can be used to smooth high-frequency noise to improve the signal to noise ratio, but it also smoothes the valid data within the window, which results in signal distortion. It is very important to choose the right window size for the smoothing operation. The study monitored the prediction set by selecting different window sizes ranging from 7 to 21 points. It could be seen from Table 2 that R^2 improves steadily with the increase in window size. Meanwhile, R^2 were to become optimal value when the window size reached 19. R^2 decreased steeply with the increasing window size.

Table 2. Smoothing Window Size with the Optimal PLS Models for R^2

Smoothing window size	Factors	RMSECV	R^2
7 points	8	1.0307031	0.8422679
9 points	9	1.0283575	0.8409819
11 points	10	0.9725069	0.8582605
13 points	10	0.9394095	0.868045
15 points	9	0.8902916	0.8808613
17 points	10	0.8643142	0.88822294
19 points	8	0.8356439	0.8952178
21 points	10	1.0532745	0.8337787

4.2 Treatments of Derivative

Derivative spectral pretreatment methods are commonly used to eliminate baseline drift and band overlapping, besides the reinforcement of band features. The derivatives contain first and second derivatives. First derivative can be used to remove with a wavelength-independent drift, while the second derivative can be used to remove a wavelength-dependent drift. The high-frequency noise is amplified while low-frequency baseline is eliminated. Spectra must be smoothed before they are actually used. It could be seen from Table 3 that the best calibration results were obtained for the first derivative by selecting the window size with 23 points while $R^2=0.9259052$ and $RMSECV=0.724261$. The best prediction accuracy could be obtained for the second derivative+ SG smoothing by selecting the window size with 31 points while $R^2=0.86045$ and $RMSECV=0.939095$. The best prediction accuracy could be obtained for the second derivative by selecting the window size with 17+9 points while $R^2=0.9318595$ and $RMSECV=0.6762902$.

$$\text{1}^{\text{st}} \text{ derivative : } X_{k,1st} = \frac{X_{k+g} - X_{k-g}}{g} \quad (4)$$

$$2^{\text{nd}} \text{ derivative : } X_{k,2\text{nd}} = \frac{X_{k+g} - 2X_k + X_{k-g}}{g^2} \quad (5)$$

Table 3. Results of 1st Derivative with the Optimal PLS Models in Different Window Size

1st derivative window size	Factors	RMSECV	R ²
13 points	5	0.7882226	0.9068677
15 points	4	0.7441875	0.9166693
17 points	4	0.7250665	0.9208594
19 points	4	0.7287187	0.9201731
21 points	5	0.7233103	0.921207
23 points	8	0.7024261	0.9259052
25 points	7	0.7099903	0.9244087

Table 4. Results of 2nd Derivative with the Optimal PLS Models in Different Window Size

2nd derivative	Factors	RMSECV	R ²
15 points	29	1.1165124	0.8121259
17 points	22	1.0997503	0.8170651
19 points	18	1.0532745	0.8337787
21 points	11	1.0307031	0.8402679
23 points	19	1.0371665	0.8365808
25 points	12	1.0283575	0.8409819
27 points	7	0.9725069	0.8582605
29 points	6	1.0060037	0.8473686
31 points	5	0.9394095	0.86045

Table 5. 2nd Derivative +SG Smoothing Results with Optimal PLS Models for Different Window Size

2nd derivative + SG smoothing	Factors	RMSECV	R ²
15 points + 9 points	8	0.69081650	0.928464
17 points + 9 points	9	0.6762902	0.9318595
19 points + 9 points	10	0.6837488	0.90247
21 points + 9 points	10	0.6785429	0.9308974
23 points + 9 points	9	0.6899397	0.9284221
25 points + 9 points	10	0.6805919	0.9306082
27 points + 9 points	8	0.6896414	0.9286152
29 points + 9 points	10	0.6806929	0.9302197
31 points + 9 points	8	0.69081650	0.928464

4.3 Orthogonal Signal Correction

Orthogonal signal correction method is mainly used to filter out such noise as baseline drift, light scattering from the near-infrared spectrum matrix and its main idea is to filter raw spectral matrixes X and Y which are irrelevant to the test quality. The data were analyzed with UnscramblerV9.5 and matlab7.0. In addition, the feasibility of PLS models could be verified for different OSC factors. The raw spectra of soybean stalk could be pretreated with an OSC model. It could be seen from Fig. 3 that RMSECV was the lowest after pretreatment when the PLS factor was 2. RMSEV increased greatly with the

increasing number of PLS factors and it reached its maximum for an OSC factor of 10. Compared to several other denoising methods, the whole OSC denoising process remained volatile and unstable. It could be used to accelerate the fitting with the model, although it had not much effect on the prediction accuracy of the model established.

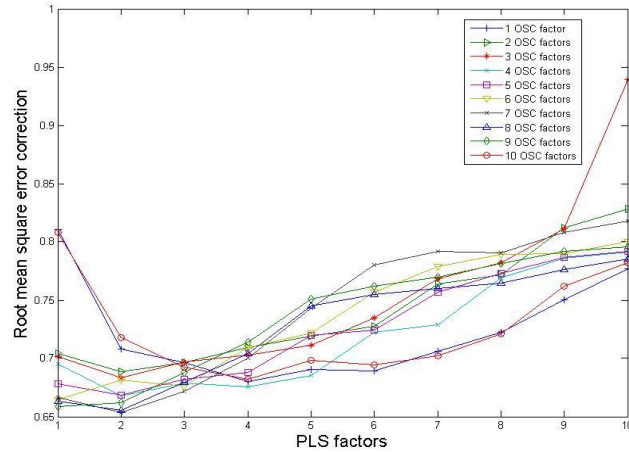


Figure 3. PLS Factors for OSC Pretreatment

It was found through analysis of different de-noising methods that second derivative + smoothing and OSC could be used to obtain the optimal correction results. Smoothing by denoising was apt to the losses of useful spectral information due to different smoothing window size used, so it was worthwhile to spend much time in selecting right window size. Normalization could be used to eliminate the variation of light path and sample dilution. Normalization had little effect on the denoising results because it does not change much. SNV and MSC were mainly used to erase scattering and to improve the repeatability of spectra, but the size and uniformity distribution of solid particles caused significant interference on the spectral information. The wavelength-independent shift could be removed by first derivative, and the second derivative could be used to remove a wavelength-dependent drift. However, the high-frequency noise would be amplified while low-frequency baseline was eliminated by the second derivative. So, raw spectra must be smoothed before they were actually used. It could be seen from Figure 4 that the denoising effect of model was improved greatly after a second derivative smoothing operation.

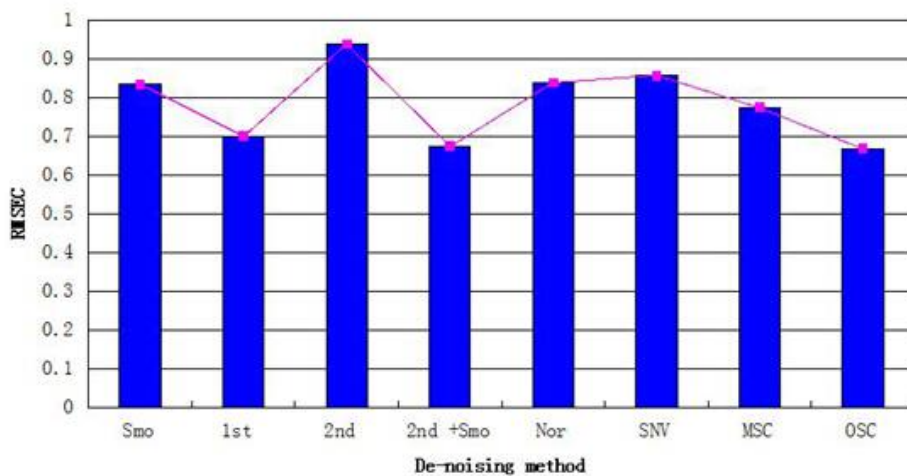


Figure 4. RMSEC in Different De-noising Method

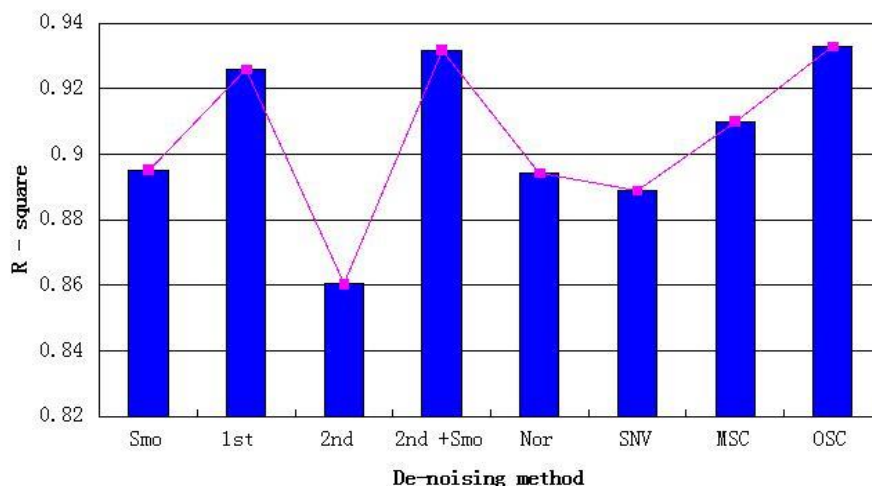


Figure 5. R^2 for Different De-noising Method

5. Conclusions

The study discussed denoising approach to the cellulose of soybean straw in different way. A variety of de-noising methods were compared to choose the optimal method including SG smoothing (Savitzky-Golay), SNV (Standard Normal Variate), derivative processing (1st derivative; 2nd derivative), normalization and orthogonal signal correction (OSC). The spectra were pretreated with the method of S-G smoothing, 1st derivative, 2nd derivative, orthogonal signal correction, normalization, SNV as well as MSC. Statistical analysis showed that RMSEC had 0.8356439 for smoothing, 0.7024261 for 1st derivative, 0.9394095 for 2nd derivative, 0.6762902 for 2nd +Smoothing, 0.839797 for Normalization, 0.8579724 for SNV, 0.7757527 for MSC and 0.6696454 for OSC. It turns out that using orthogonal signal correction is feasible to correct reasonably filtering out independent signal in the spectral density matrix and the matrix. This way is valid for reducing the number of model factors, which decreases model complexity and enhance robustness of model. It has a significant role in detecting the composition of the straw finely and guiding the manufacturing process of biomass in future.

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] X. Liu and L. Han, "Evaluation of near-infrared reflectance spectroscopy (NIRS) for predicting chemical composition of straw silage", *Journal of Animal and Feed Sciences*, vol. 15, (2006), pp. 329-336.
- [2] MA long-long, "Process Technology of Bio- Energy Utilization and Its Development", *Chemical Industry*, vol. 25, (2007), pp. 9-14.
- [3] B. Liu, "Problems and countermeasures of new energy technology in our country", *Journal of Liaoning University of Technology*, vol. 11, no. 2, (2009).
- [4] A. Belanche, M. R. Weisbjerg, G. G. Allison, C. J. Newbold, J. M. Moorby, "Estimation of feed crude protein concentration and rumen degradability by Fourier-transform infrared spectroscopy", *Journal of Dairy Science*, vol. 96, (2013), pp. 7867-7880.
- [5] S. Jin and H. Chen, "Near-infrared analysis of the chemical composition of rice straw", *Industrial Crops and Products*, vol. 26, (2007), pp. 207-211.

- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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. 1, no. 2, (2010), pp. 321-326.
- [10] L. Liu, X. Philip, Y. Alvin, 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.
- [11] 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.
- [12] 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.
- [13] F. Xu and D. Wang, "Rapid determination of sugar content in corn stover hydrolysates using near infrared spectroscopy", *Bioresource Technology*, (2013), pp. 293-298.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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 2wsupport vector data description", *Microchemical Journal*, vol. 102, (2012), pp. 68-74.
- [19] G. W. Mathison, H. Hsu, R. Soofi-Siawash, G. Recinos-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.
- [20] C. Huang, L. Han, Z. Yang and X. Liu, "Prediction of heating value of straw by proximate data, and nearinfrared spectroscopy", *Energy Conversion and Management*, vol. 49, (2008), pp. 3433-3438.

Author



Zhongbin Su, (1965-), male, he Ph.D., professor, mainly engaged in the research and application of information technology in agriculture.