Determination of manganese content in cottonseed meal using near-infrared spectrometry and multivariate calibration

YU En[†], ZHAO Rubing[†], CAI Yunfei, HUANG Jieqiong, LI Cheng, LI Cong, MEI Lei, BAO Lisheng, CHEN Jinhong and ZHU Shuijin[*]

Abstract

Background: Manganese (Mn) is an essential microelement in cottonseeds, which is usually determined by the techniques relied on hazardous reagents and complex pretreatment procedures. Therefore a rapid, low-cost, and reagent-free analytical way is demanded to substitute the traditional analytical method.

Results: The Mn content in cottonseed meal was investigated by near-infrared spectroscopy (NIRS) and chemometrics techniques. Standard normal variate (SNV) combined with first derivatives (FD) was the optimal spectra pre-treatment method. Monte Carlo uninformative variable elimination (MCUVE) and successive projections algorithm method (SPA) were employed to extract the informative variables from the full NIR spectra. The linear and nonlinear calibration models for cottonseed Mn content were developed. Finally, the optimal model for cottonseed Mn content was obtained by MCUVE-SPA-LSSVM, with root mean squares error of prediction (RMSEP) of 1.994 6, coefficient of determination (R^2) of 0.949 3, and the residual predictive deviation (RPD) of 4.370 5, respectively.

Conclusions: The MCUVE-SPA-LSSVM model is accuracy enough to measure the Mn content in cottonseed meal, which can be used as an alternative way to substitute for traditional analytical method.

Keywords: Cottonseed meal, Manganese content, Near-infrared spectroscopy, Multivariate calibration

Introduction

Manganese (Mn) is an essential microelement for plant growth. For example, Mn participated in the water-splitting system of photosystem II (PSII) and provided electrons necessary for photosynthetic electron transport. In addition, a group of four Mn atoms (Mn cluster) was associated with the oxygen-evolving complex (OEC) bound to the reaction center protein (D1) of PSII in water photolysis (Goussias et al. 2002). Mn also intervened in activating enzyme-catalyzed reactions, including phosphorylation, decarboxylation, reduction, and hydrolysis reaction. These reactions could affect the processes such as respiration, amino acid synthesis, lignin biosynthesis and the level of hormones in plants (Millaleo et al. 2010). Although Mn is an important

inorganic element for plant growth and development, plant disorders would be happen if soils contained extremely high amounts of Mn or acid soils had moderate Mn content (Robinson 1919). For example, Mn toxicity could cause the crinkle leaf disease of cotton.

Cottonseed is an important by-product of cotton production, which has high contents of protein (27.83% ~ 45.60%) and oil (28.24% ~ 44.05%). Cottonseed could be used as livestock food and edible oil. However, high Mn contents in the cottonseeds will restrict the utilization of cottonseed, as ingestion of excess Mn can cause toxic effects for human or animals. For example, exposure to Mn in childhood at concentrations exceeding the homeostatic range can cause a neurotoxic syndrome that affects dopamine balance and behavior control (Ericson et al. 2007; Zoni and Lucchini 2013).

Although Mn is one of the most important microelements for cotton growth, high Mn content will be a limiting factor of the utilization of cottonseed, especially

* Correspondence: shjzhu@zju.edu.cn
[†]En YU and Rubing ZHAO contributed equally to this work.
Department of Agronomy, Zhejiang Key Lab of Crop Germplasm, Zhejiang University, Hangzhou 310058, China

when cottonseed is used as feed for livestock. Therefore, it's very important to measure the cottonseed Mn content. Cottonseed Mn content is generally determined by atomic absorption spectrometry (AAS), inductively coupled plasma optical emission spectrometry (ICP-OES), or inductively coupled plasma mass spectrometry (ICP-MS). However, due to relying on hazardous re-agents and complex pretreatment procedures of samples, these methods are quite expensive and time-consuming. In contrast, near-infrared spectroscopy (NIRS) is a rapid, non-destructive, pretreatment-simple, low-cost, effective, and reagent-free analytical method, which can supply an alternative way to substitute the traditional analytical method to measure cottonseed Mn content.

Theoretically, there were no absorption bands for inorganic elements in the NIR region. However, inorganic elements could chelate with organic compounds, so it indirectly reflected in near-infrared spectra (Kumagai et al. 2013; Chen et al. 2010). Recently, NIRS has been applied to analyze inorganic elements concentration in different plant species, including cadmium and arsenic in rice (Kumagai et al. 2013; Font et al. 2005; Zhu et al. 2015), arsenic and lead in red paprika (Moros et al. 2008). In addition, inorganic elements concentration in sediment (Xia et al. 2007), soil (Moros et al. 2009), and water samples (Ning et al. 2012; Kleinebecker et al. 2013) were also determined by NIRS. However, no reports have been published on the usage of NIRS technique to measure the microelements content in cottonseed meal.

In order to set up a fast and accurate method to measure cottonseed Mn content, partial least squares (PLS) and least-squares support vector machine (LSSVM) regression are used to develop the calibration models. In addition, the variable selection methods, including Monte Carlo uninformation variables elimination (MCUVE) and successive projections algorithm (SPA), are employed to improve the performance of models.

Materials and methods
Plant materials
A total of 288 cottonseed samples were collected from 10 cultivar regional experiments located in Yangtze River cotton production region of China in 2013, including Hangzhou (30°16′N, 120°09′E), Jiangshan (28°74′N, 118°61′E), Jinhua (29°12′N, 119°64′E), Lixian (29°65′N, 111°75′E), Wuhu (30°52′N, 114°31′E), Wulin (29°05′N, 111°69′E), Yancheng (33°38′N, 120°13′E), Jiujiang (29°71′N, 115°97′E), Yueyang (29°37′N, 113°09′E), and Hefei (31°86′N, 117°27′E). There were 11 cultivars or lines in each experiment with a randomized block design and three replicates. All agronomic managements, including weed and disease control, were the same as those of local cotton production. The cotton-seeds materials were sampled at harvest, then stored at 4 °C for Mn analysis.

Mn measurement
Each sample was ground by an auto milling-machine and passed through a 0.4 mm screen. A total of 0.40 g of cottonseed powder was measured and digested at 80 °C for 30 min in a tube containing 6 mL HNO_3 and 0.2 mL H_2O_2 (30%, v/v). The tubes were then digested in a microwave digestion oven (Microwave 3000, Anton paar, Austria) for another 90 min. The element concentration in the digested solution was determined by inductively coupled plasma-mass spectrometer (Elan DCR-e Peki-nElmer USA) after appropriate dilution. All reagents were of the highest purity and all solutions were prepared in ultrapure water produced by Millipore Milli-Q system (Bedford, MA, USA) with a resistivity of 18.2 MʊΩ·cm.

NIRS measurement
About 3.5 g cottonseed meal was loaded in a circle sample cup (35 mm in diameter and 18 mm in depth) and pressed moderately to obtain similar packing density. In order to get an ideal working condition for NIR machine, the temperature and humidity were strictly controlled within 25 °C and 40%, respectively. The loading time was controlled as fast as possible to avoid excessive moisture absorption. The spectra were collected in the wavelength range of 1 100~2 498 nm, and were recorded as log (1/R) at 2 nm interval using the WinISI II (InfraSoft International, USA) software. Each sample was loaded and scanned 4 times, and the average spectrum was used for NIR analysis.

Spectral data analysis
The raw spectral data needed to be preprocessed because lots of systematic noises and slope-background information existed in NIR original spectral (Li et al. 2012). In our experiment, different pre-processing methods were used to increase the relationship between the chemical composition and spectral signal, including Savitzky-Golay (SG) smoothing, the first or second derivative (FD, SD; the value of polynomial and the number of points in the window were 1 and 5, respectively), multiplicative scatter correction (MSC), linear baseline correction, spectroscopic transformation (ST), standard normal variate (SNV), and some of their combinations. All these pre-processing methods were carried out according to the instructions of the Unscrambler V9.7 (CAMO PROCESS AS, Oslo, Norway). All chemometric algorithms were performed in Matlab (Version 7.12.0.635, the MathWorks. Inc. US) under Windows 7.

Results
Reference data and reflectance spectra analysis
In modeling, 288 samples were divided into two sets with a ratio of 3:1 according to Kennard-Stone algorithm

YU *et al. Journal of Cotton Research*   (2019) 2:12

Page 3 of 7

based on Euclidean distances (Kennard and Stone 1969). Seventy-two samples consisted of the validation set for the prediction, and 216 samples formed calibration set for modeling (Table 1). The cottonseed Mn content in calibration set ranged from 10.251 9 to 48.991 8 mg·kg^{-1}, and that of validation sets ranged from 11.031 6 to 41.392 2 mg·kg^{-1}. The range of calibration set covered the whole range of validation set and the value of RSD varied obviously in these two sets (Table 1), which indicated that Mn distribution in these two sets was appropriate and had the ability to develop reliable calibration equations (Bao et al. 2007). The calibration model was checked by full cross-validation (Gómez et al. 2006). In addition, the validation set, an external test set, was also used to validate the actual prediction ability of calibration model (Esteban-Díez et al. 2007).

In our study, the regression modeling technique and different pre-treatment methods were used to optimize NIR spectra data for cottonseed meal. Compared with the raw data as a control, SNV, MSC, AN, TB, FD and ST pre-treatment methods decreased the value of root mean square error (RMSE), while increased the value of coefficient of determination (R^2) (Table 2), which indicated that these methods improved the quality of regression model for cottonseed Mn content. The optimal spectra preprocessing method was obtained by the combination SVN with FD, with a lowest RMSE and highest R^2 values (Table 2).

The raw spectra revealed three prominent absorption bands at 1 500, 1 750, and 1 950 nm, as well as four small absorption bands at 1 200, 2 050, 2 300, and 2 350 nm (Fig. 1a). However, the reflectance spectra changed significantly after using the optimal preprocessing method (Fig. 1a, b). There were also three prominent absorption bands, but the number of small absorption bands reached eight (Fig. 1b), and all absorption bands became much sharper and clearer than that of raw spectra, which indicated that the optimal spectra pre-processing method improved the spectra quality for modeling.

Development of full-spectra PLS and LSSVM model

Before developing the full-spectra PLS model, the latent variables (LVs) should be optimized. In this study, the

Table 1 The statistic of Mn content of sample in two sets by KS algorithm

Subsets	Sample number	Mn contents / (mg·kg^{-1})		SD	RSD / %
		Range	Mean		
Total	288	10.251 9 ~ 48.991 8	19.583 1	8.328 4	42.53
Calibration set	216	10.251 9 ~ 48.991 8	18.709 5	8.026 7	42.90
Prediction set	72	11.031 6 ~ 41.392 2	22.204 0	8.717 4	39.26

SD standard deviation, *RSD* relative standard deviation

Table 2 Evaluation of different pre-treatment methods for Mn contents in cottonseed meal

Pre-treatment	RMSE	R^2
Raw data	3.219 4	0.851 1
SNV	3.167 5	0.855 9
MSC	3.167 9	0.855 8
AN	3.141 9	0.858 2
FD	2.938 4	0.876 0
SD	4.156 0	0.751 9
TB	3.178 2	0.854 9
SG	3.223 7	0.850 7
ST	3.130 3	0.859 2
SG + FD	3.010 3	0.869 8
SNV + FD	2.901 8	0.879 0
SNV + SG	3.170 9	0.855 5

SNV standard normal variate, *MSC* multiplicative scatter correction, *AN* area normalize, *FD* first derivatives, *SD* Second derivatives, *TB* transform baseline, *SG* Savitzky-Golay smoothing, *ST* spectroscopic transformation, *RMSE* root mean square error, R^2 prediction coefficient of determination

optimal number of LVs for PLS was determined by prediction residual error sum of squares (PRESS) value from the one-out cross validation procedure. The PRESS value descended greatly with the increasing of LVs, to the lowest level at LVs = 10, and then slowly increased when LVs were > 10 (Additional file 1: Figure S1). Therefore, LVs = 10 was considered as the optimal value for PLS model. The predictive results of PLS model are shown in Table 3.

The parameters, γ and $\mathcal{x}^2$, should be optimized in the radius basis function (RBF) kernel function to develop the full-spectra LSSVM model. In this study, genetic algorithm (GA) approach and tenfold cross-validation were applied for global optimization of these two parameters, and the optimal values of γ and $\mathcal{x}^2$ were 2.060 1 and 2.255 1, respectively. Compared with the full-spectra PLS model, the values of R^2 and RPD were increased, while that of RMSEC, RMSEP, and RMSECV were decreased in full-spectra LSSVM model (Table 3), which indicated that non-linear regression model (LSSVM) was superior to linear regression model (PLS) to develop the calibration model for cottonseed Mn content.
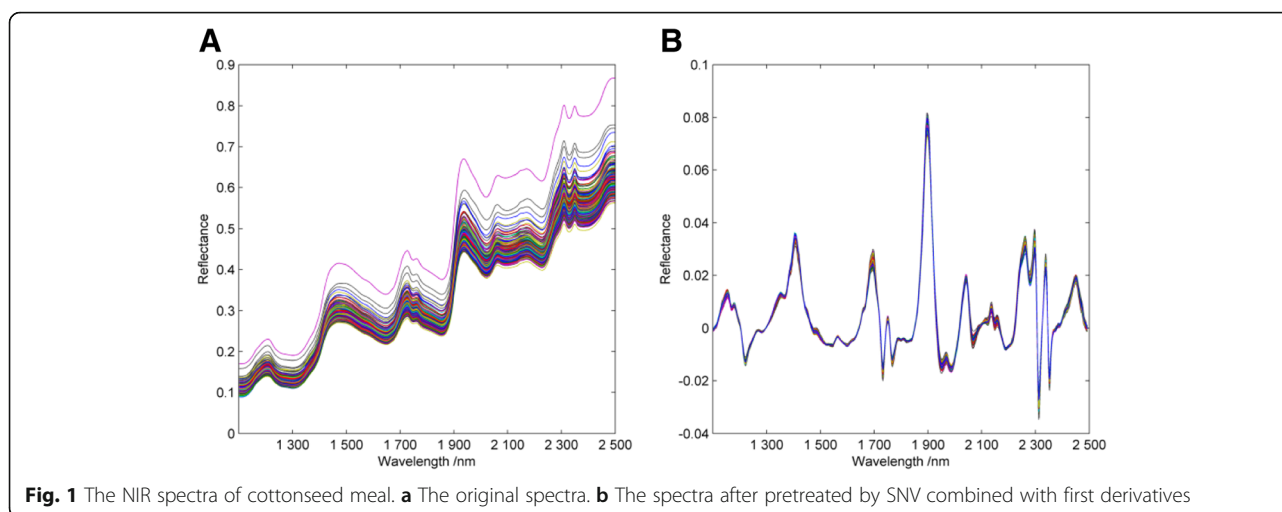
Development of LSSVM model using variables selection methods

There were 700 variables in the original spectra of cottonseed meal and most of them were typically consisted of broad, weak, nonspecific, and extensively overlapped bands (Blanco et al. 1994). In order to improve the predictive precision and eliminate the influence of uninformative variables on the robust of LSSVM model, Monte Carlo uninformative variable elimination (MCUVE) and successive projections algorithm (SPA) method were proposed for

Fig. 1 The NIR spectra of cottonseed meal. **a** The original spectra. **b** The spectra after pretreated by SNV combined with first derivatives

variable selection. The stability of each variable in the wavelength from 1 100 to 2 498 nm was evaluated by MCUVE method (Fig. 2a). The stability of any variable between the dot lines would be identified as uninformative variable and should be eliminated. The root mean square error of cross validation (RMSECV) value changed relied on the cutoff value and the minimal RMSECV value could get the optimal cutoff value (Fig. 2b). The optimal cutoff value was set as 1.2, then 233 variables were selected by MCUVE to establish MCUVE-LSSVM model, which were given in Table 3. In order to further optimize the MCUVE-LSSVM model, SPA method was used to reduce the number of uninformative variables further. Finally, 49 variables were obtained to develop the MCUVE-SPA-LSSVM model. The predictive results of MCUVE-SPA-LSSVM model were shown in Table 3.

Comparison of accuracy of four kinds of regression models

There were some important criteria to evaluate the performance of regression models, such as the coefficient of determination (R^2) between the measured and predicted parameters, the residual predictive deviation (RPD) which was calculated as the ratio between the SD of
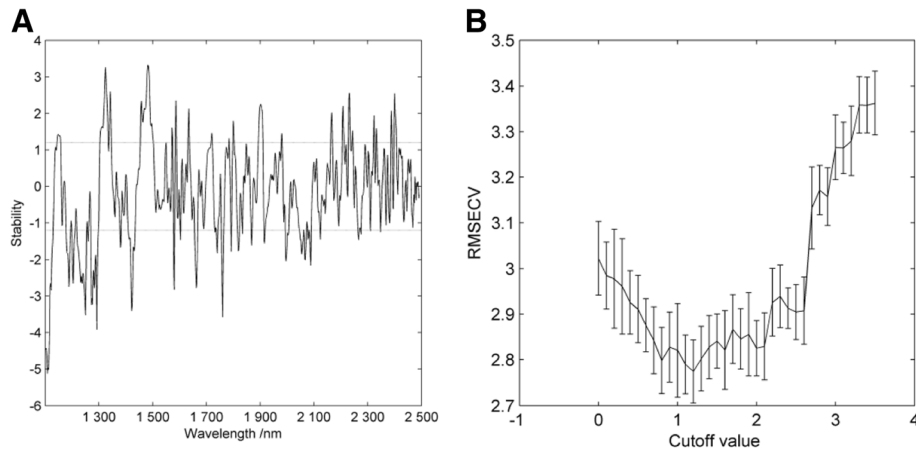
reference value and the standard error of cross-validation (SECV). RPD was indicative of the usefulness of the calibration model; if the ratio exceeded 3, the calibration model was excellent, whereas the ratio below 2, its applicability was limited (Rosales et al. 2011). RMSECV and the root mean square error of prediction (RMSEP) were other two indexes. The model with a low RMSECV and RMSEP, and a high RPD and R^2 values was considered as a good one, and a good model had a reliable ability to predict the chemical composition (Arana et al. 2005).

Four kinds of regression models, namely PLS, LSSVM, MCUVE-LSSVM, and MCUVE-SPA-LSSVM were built in our study. The four calibration models were set for the same optimal parameters, and the criteria to evaluate the performance of different regression models were shown in Table 3. It revealed that LSSVM model had better performance than PLS model in measurement of cottonseed Mn content. While MCUVE-LSSVM model with 233 variables had better quality than the full spectra LSSSVM model, as the values of R^2 and RPD were increased, while the RMSEP and RMSECV values were decreased. Furthermore, the MCUVE-SPA-LSSVM model had the best prediction ability, as only 49 useful variables were selected to

Table 3 The performance of four kinds of regression models

Models	Varialbes	Calibration			Validation			RPD	RMSECV
		Num	R^2_c	RMSEC	Num	R^2_v	RMSEP		
PLS	700	216	0.894 7	2.598 3	72	0.926 9	2.438 7	3.574 6	3.171 9
LSSVM	700	216	0.984 9	1.001 7	72	0.944 1	2.108 4	4.134 7	2.491 0
MCUVE-LSSVM	233	216	0.976 3	1.242 0	72	0.948 6	2.010 8	4.335 4	2.309 1
MCUVE-SPA-LSSVM	49	216	0.972 9	1.330 8	72	0.949 3	1.994 6	4.370 5	2.101 3

Num Number of samples used to perform the calibration and validation, R^2_c Coefficient of determination in calibration, R^2_v Coefficient of determination in calibration, *RMSEC* The root mean square error of calibration, *RMSEP* The root mean square error of prediction, *RPD* The residual predictive deviation, *RMSECV* The root mean square error of cross-valid
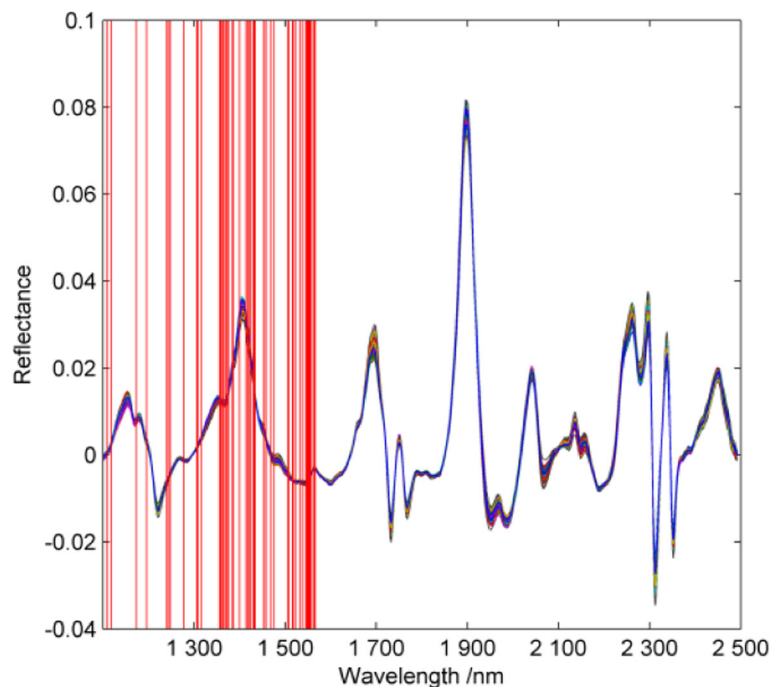
Fig. 2 Variables selection in Mn NIR spectral modeling. **a** The stability distribution of each variable by MCUVE method. The two dotted lines indicate the lower and upper cutoff values, respectively. **b** The trend of RMSECV with the descending of the cutoff value based on MCUVE

develop the calibration model (Fig. 3) and 651 uninformative variables were eliminated by MCUVE-SPA method.
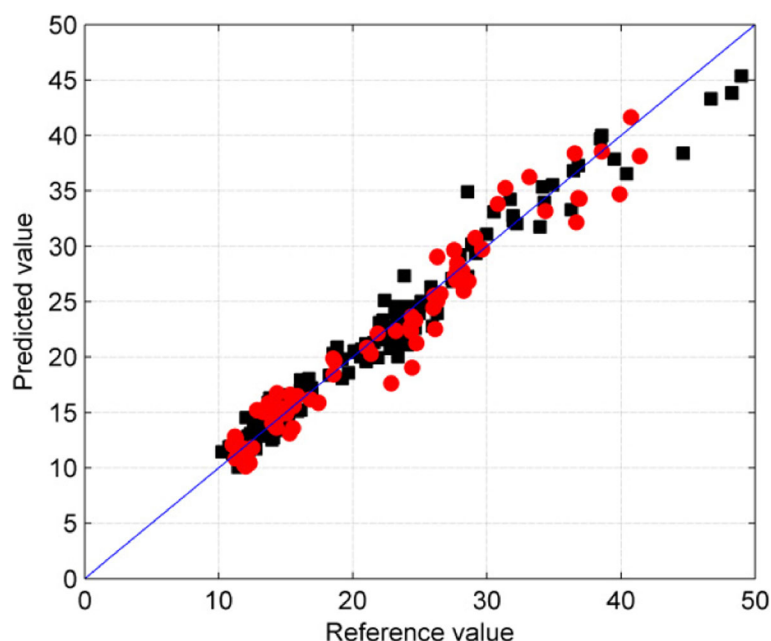
Discussion

In present work, the full-spectra PLS and LSSVM algorithm were implemented to build regression models for cottonseed Mn content. Compared with full-spectra PLS model, the values of R^2 and RPD were increased, while the values of RMSEC, RESEP and RMSECV were decreased in full-spectra LSSVM model (Table 3), which

indicated that nonlinear full-spectra LSSVM model was superior to classical linear full-spectra PLS model to build the calibration model for cottonseed Mn content. Since not all of the variables were related to cottonseed Mn in the original spectra, the variables selection methods, including MCUVE and SPA, were used to eliminate the uninformative variables. Finally, 49 informative variables were selected to build the MCUVE-SPA-LSSVM model (Fig. 3). The scatter plot of the correlation between the reference and predicted values from

Fig. 3 Variables selected by MCUVE-SPA method. The red lines indicate variables selected by MCUVE-SPA method and a total of 49 variables were selected to develop the calibration model

Fig. 4 The correlation between predicted and reference values for MCUVE-SPA-LSSVM model. The blank panes and red circles represent the calibration samples and validation samples, respectively. Samples on diagonal line indicate their predicted values which are equal to reference values

MCUVE-SPA-LSSVM model for calibration and prediction sets was shown in Fig. 4. The samples distribution in both calibration and prediction set were near the diagonal line, which suggested that MCUVE-SPA-LSSVM model for cottonseed Mn had excellent correlation between predicted and reference values. Theoretically, there were no direct absorption bands for inorganic Mn in cottonseed, while the calibration model of Mn was accurate to determinate cottonseed Mn content (Fig. 4; Table 3), which indicated that Mn could be chelated with some substances in cottonseed, from which the absorption bands of Mn was indirectly reflected in near-infrared spectroscopy.

In order to know which kinds of substances could be chelated with Mn in cottonseed meal, we try to analyze the 49 informative variables selected in MCUVE-SPA-LSSVM model. The results showed that the wavelength of these selected variables were mainly concentrated at 1 110, 1 118, 1 174, 1 196, 1 240, 1 244, 1 248, and 1 278 nm, as well as some variables at 1 306 ~ 1 386 nm, 1 400 ~ 1 476 nm, and 1 506 ~ 1 566 nm (Fig. 3). It was reported that wavelength between 1 100 ~ 1 672 nm was resulted from reduced intensity of the water bangs with the increased total protein contents (Hacisalihoglu et al. 2009). As we know, cottonseed is rich in proteins and oils. In addition, proteins are complex nutritional parameters including many chemical bonds such as C-H, O-H, N-H and S-H, which are the four main types of bands in organic compounds. These bonds have strong absorbance in near-infrared region (Zhu et al. 2015), and may be specially related to cottonseed Mn content. This may give an

explanation of why inorganic Mn content could be detected by NIRS technique. However, which kinds of organic compounds can chelate with Mn in cottonseed is still unknown. We need to isolate the target organic compounds related to Mn in further study.

Conclusions
The calibration and validation statistics obtained in the current work showed the potential of NIRS to predict microelement Mn content in cottonseed meal. The best results were obtained by using MCUVE-SPA LSSVM method, with RMSEP of 1.994 6, R^2 of 0.949 3, and RPD of 4.370 5, respectively. This model was accurate enough to measure the cottonseed Mn content, and supplied an alternative way to substitute for traditional analytical method.

Additional file

Additional file 1: Figure S1. Variation of prediction residual error sum of squares value (PRESS value) with different latent variables (LVs) for Mn full-spectra PLS model. (DOCX 46 kb)

Authors' contributions
Yu E and Zhu SJ designed the experiments and wrote the manuscript. Yu E, Zhao RB and Cai YF analyzed the data. Yu E, Zhao RB, Huang JQ and Li C (Cheng) participated in experiment. Li C (Cong), Mei L, Bao LS and Chen JH assisted in editing the article. Zhu SJ and Chen JH conducted and supervised the experiments. All authors read and approved the final manuscript.

References
Arana I, Jarén C, Arazuri S. Maturity, variety and origin determination in white grapes (*Vitis Vinifera* L.) using near infrared reflectance technology. J Near Infrared Spectrosc. 2005;13:349–57.

Bao J, Wang Y, Shen Y. Determination of apparent amylose content, pasting properties and gel texture of rice starch by near-infrared spectroscopy. J Sci Food Agric. 2007;87:2040–8.

Blanco M, Coello J, Iturriaga H, et al. Analysis of cotton-polyester yarns by near-infrared reflectance spectroscopy. Analyst. 1994;119:1779–85.

Chen GP, Mei Y, Tao W, et al. Micro near infrared spectroscopy (MicroNIRS) based on on-line enrichment: determination of trace copper in water using glycidyl methacrylate-based monolithic material. Anal Chim Acta. 2010;670:39–43.

Ericson JE, Crinella FM, Clarke-Stewart KA, et al. Prenatal manganese levels linked to childhood behavioral disinhibition. Neurotoxicol Teratol. 2007;29:181–7.

Esteban-Díez I, González-Sáiz JM, Sáenz-González C, et al. Coffee varietal differentiation based on near infrared spectroscopy. Talanta. 2007;71:221–9.

Font R, Vélez D, Del Río-Celestino M, et al. Screening inorganic arsenic in rice by visible and near-infrared spectroscopy. Microchim Acta. 2005;151:231–9.

Gómez AH, He Y, Pereira AG. Non-destructive measurement of acidity, soluble solids and firmness of Satsuma mandarin using Vis/NIR-spectroscopy techniques. J Food Eng. 2006;77:313–9.

Goussias C, Boussac A, Rutherford A W. Photosystem II and photosynthetic oxidation of water: an overview. Philos Trans R Soc Lond B Biol Sci. 2002;357: 1369–81.

Hacisalihoglu G, Larai B, Settles AM. Near-infrared reflectance spectroscopy predicts protein, starch and seed weight in intact seeds of common bean (*Phaseolus vulgaris L.*). J Agric Food Chem. 2009;58:702–6.

Kennard RW, Stone LA. Computer aided design of experiments. Technometrics. 1969;11:137–48.

Kleinebecker T, Poelen M, Smolders A, et al. Fast and inexpensive detection of total and extractable element concentrations in aquatic sediments using near-infrared reflectance spectroscopy (NIRS). PLoS One. 2013;8:e70517.

Kumagai M, Ohisa N, Amono T, et al. Canonical discriminant analysis of cadmium content levels in unpolished rice using a portable near-infrared spectrometer. Anal Sci. 2013;19:1553–5.

Li S, Zhu X, Zhang J, et al. Authentication of pure camellia oil by using near infrared spectroscopy and pattern recognition techniques. J Food Sci. 2012;77:C374–80.

Millaleo R, Reyes-Díaz M, Ivanov A G, et al. Manganese as essential and toxic element for plants: transport, accumulation and resistance mechanisms. J Soil Sci Plant Nutr. 2010;10:470–81.

Moros J, Llorca I, Cervera ML, et al. Chemometric determination of arsenic and lead in untreated powdered red paprika by diffuse reflectance near-infrared spectroscopy. Anal Chim Acta. 2008;613:196–206.

Moros J, Martínez-Sánchez MJ, Pérez-Sirvent C, et al. Testing of the region of murcia soils by near infrared diffuse reflectance spectroscopy and chemometrics. Talanta. 2009;78:388–98.

Ning Y, Li J, Cai W, et al. Simultaneous determination of heavy metal ions in water using near-infrared spectroscopy with preconcentration by nano-hydroxyapatite. Spectroc Acta Pt A-Molec Biomolec Spectr. 2012;96:289–94.

Robinson W. The water soluble manganese of soils. Science. 1919;50:423–5.

Rosales A, Galicia L, Oviedo E, et al. Near-infrared reflectance spectroscopy (NIRS) for protein, tryptophan, and lysine evaluation in quality protein maize (QPM) breeding programs. J Agric Food Chem. 2011;59:10781–6.

Xia XQ, Mao YQ, Ji JF, et al. Reflectance spectroscopy study of Cd contamination in the sediments of the Changjiang River, China. Environ Sci Technol. 2007;41:3449–54.

Zhu X, Li G, Shan Y. Prediction of cadmium content in brown rice using near-infrared spectroscopy and regression modelling techniques. Int J Food Sci Technol. 2015;50:1123–9.

Zoni S, Lucchini R. Manganese exposure: cognitive, motor and behavioral effects on children: a review of recent findings. Curr Opin Pediatr. 2013;25:255.