Journal of Cereal Science 69 (2016) 145–150

Contents lists available at ScienceDirect

Journal of Cereal Science

journal homepage: www.elsevier.com/locate/jcs

Feasibility of analyzing frost-damaged and non-viable maize kernels based on near infrared spectroscopy and chemometrics

Shiqiang Jia [a], Liguo Yang [b], Dong An [a, *], Zhe Liu [a], Yanlu Yan [a], Shaoming Li [a], Xiaodong Zhang [a], Dehai Zhu [a], Jiancheng Gu [c]

[a] *College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China*
[b] *Beijing Agricultural Machinery Test Identification Promotion Station, Beijing, 100079, China*
[c] *Beijing Kings Nower Seed S&T CO., LTD., Beijing, 100080, China*

ARTICLE INFO

Article history:
Received 13 June 2015
Received in revised form
24 February 2016
Accepted 29 February 2016
Available online 2 March 2016

Keywords:
Seed damage
Viability
Near infrared spectroscopy
Discrimination

ABSTRACT

The current detection methods for maize seed damage and viability assessment are usually time-consuming, tedious, and costly. In this investigation, near infrared spectroscopy (NIRS), a nondestructive and rapid analytical method, was explored to analyze frost-damaged and non-viable seeds. Principal component analysis (PCA), partial least squares (PLS) and orthogonal linear discriminant analysis (OLDA) were combined to extract feature from near infrared (NIR) spectra. PLS + OLDA can extract difference characteristics of normal and frost-damaged seeds more efficiently than PCA + OLDA. Three classification algorithms were utilized and compared: Support vector machine (SVM), biomimetic pattern recognition (BPR), and mahalanobis distance (MD). BPR can classify normal and frost-damaged seeds better and achieved the highest average accuracy of 97%. Discrimination model of viable and non-viable seeds in frost-damaged seeds based on SVM, BPR and MD were built and achieved accuracies of 94%, 97.25% and 89.5% respectively. BPR model yielded the most correct value in predicting germination rate of validation set, and improved the germination rate of validation set dramatically from 27.5% to 100% by screening out non-viable seeds. NIRS and chemometrics as demonstrated in this paper can provide a novel method which can assess and improve quality of maize seed quickly and inexpensively.

1. Introduction

Maize is an important food crop that can be processed into various foods, fodders, and industrial products. Maize production has strict requirements on the quality of seeds to guarantee favorable yield. Damage degree and viability are two vital indicators of seed quality grading. Seeds will exhibit low viability under adversities (e.g., insect pest, frost, rain, and so on) or damage during threshing, drying, storage, transformation, and other harvest processes. Frost damage caused by low temperature can also damage seeds to different degrees and thus decrease seed viability. Low temperature can cause the intercellular freezing of plant tissues, excessive cytoplasm dehydration, and protein structure damage (Nan et al., 2010; Woltz et al., 2006). Discriminating damaged seeds in the production process is necessary to guarantee seed quality. The decrease of viability would lead to the decline of germination rate and do harm to seed production companies. To improve the germination rate of a batch of seed, those non-viable seeds must be screened out. This would be of great significance in improving seed quality.

Ordinary methods used for seed damage and viability detection include seedling growth characteristic measurement, adversity resistance measurement, physiological and biochemical determination, as well as physical and chemical methods, among others (Hu et al., 2009). Seedling growth characteristic and adversity resistance measurement methods required a germination test to determine seed viability according to seedling growth. Physiological and biochemical determination involves conductivity measurement, staining, and enzyme activity determination.

Abbreviations: NIR, near infrared; NIRS, near infrared spectroscopy; PCA, principal component analysis; PLS, partial least squares; OLDA, orthogonal linear discriminant analysis; SVM, support vector machine; BPR, biomimetic pattern recognition; MD, mahalanobis distance; FT-NIR, Fourier transform near infrared; TTC, triphenyltetrazolium chloride; ACR, average correct rate.

* Corresponding author. College of Information and Electrical Engineering, China Agricultural University, Box 63, No. 17, Qinghuadonglu, Haidian district, Beijing 100083, China.

E-mail address: andong@cau.edu.cn (D. An).

http://dx.doi.org/10.1016/j.jcs.2016.02.018
0733-5210/

Triphenyltetrazolium chloride (TTC) is a widely-used staining method in physiological and biochemical determination. Seed viability is determined by whether the embryo can be stained and based on the stained parts of the embryo (Hu et al., 2009). Physical and chemical methods include soft X-ray ultra weak illumination intensity, among others (Gao, 1998; Cheng et al., 2001). These measurement methods can accurately detect seed viability but are incapable of satisfying the demand for lossless and rapid discrimination of damaged and nonviable seeds in mass production and processing because of the complicated, time-consuming, and high cost of the operation, as well as the possible damage sustained by seeds.

Near infrared spectroscopy (NIRS) has been widely applied for the quality analysis of crop seeds for its rapid analysis speed, low sample preparation requirements, and capability to preserve samples. Studies on maize kernels mainly focused on quantitative analyses of oil (Armstrong, 2006), protein, and starch content (Spielbauer et al., 2009). A few studies on NIRS have also focused on seed viability detection. Tigabu and Oden, (2004) applied NIRS to the viability detection of masson pine seeds. Seeds of various vitalities were obtained through the accelerated aging method and a combination of NIRS with SIMCA and PLS-DA analysis methods. The un-aged masson pine seeds and artificially aged masson pine seeds were identified with an accuracy of 100%. Furthermore, seeds with different aging times (3, 7, and 9 days) were identified with an accuracy of 80%. The reported high accuracy in discrimination enabled the discrimination of aged seeds from a large number of seeds. Yang et al. (2013) utilized NIRS and BP neural network to classify artificially aged maize seeds into three degrees and to establish a seed viability detection model with an accuracy of 85%.

Studies on seed damage detection based on NIRS are limited. Wang et al. (2001) found that the reflectance spectra between 400 and 1700 nm can accurately discriminate normal soybeans and wheat kernels from those that are heat-damaged with greater than 95% accuracy. Wang et al. (2002) employed the NIR spectral region from 400 nm to 1700 nm and the neural network model to classify sound soybean seeds and those damaged by frost and mold with accuracy of 98% and 86%, respectively. Agelet et al. (2012) discriminated normal from damaged (frost and heat damage) corn and soybean kernels seeds using NIR spectra (850 nm–1650 nm). They achieved a discrimination success rate of 99% for normal and heat damaged seeds. However, the discrimination of frost-damaged from normal seeds failed. NIRS-based discrimination of damaged (frost and mold damage) and nonviable maize seeds and a NIRS waveband of 1700 nm–2500 nm have not been systematically studied.

Our study explored the feasibility of analyzing frost-damaged and non-viable maize seeds by using the 1110 nm–2500 nm waveband in near infrared (NIR) spectra. Spectra feature extraction methods such as principal component analysis (PCA), partial least squares (PLS) and orthogonal linear discriminant analysis (OLDA) were studied. Three classification algorithms were tested and compared: Support vector machine (SVM), biomimetic pattern recognition (BPR) (Wang et al., 2002) and mahalanobis distance (MD).

2. Material and methods

2.1. Seed samples

Maize seeds were provided by Beijing Kings Nower Seed S&T Co., Ltd. Up to 400 normal, 400 frost-damaged seeds of hybrid NH101 were used in this study. Normal seeds have a moisture content of 12%–13% and germination rate above 95%. The frost-damaged seeds were obtained through manual processing. The seeds were directly threshed without equilibrium water (moisture content of 30%) immediately after harvest. Some seeds were treated with low temperature. The seeds were placed in a low-temperature environment (−19.2 °C) and dried in natural state until their moisture content reached 12%–13%. The seeds were stored in a −5 °C environment.

2.2. Spectra measurement and germination test

Fourier transform near infrared (FT-NIR) spectrometer (MPA spectrometer, Bruker Co., Germany) in diffuse reflectance mode fitted with a tungsten lamp and a cooled PbS detector was employed. Spectra were obtained at a resolution of 8 cm^{-1} in the range of 9000 cm^{-1} to 4000 cm^{-1} (1110 nm–2500 nm, a total of 633 wavelength points) at room temperature (approximately 25 °C). Each spectrum was the average of 20 scanned interferograms to reduce noise signal from non-uniform bulk. All spectra were recorded as log (1/R) with respect to the golden reference standard. The software OPUS 6.5 (Bruker Co., Germany) was used to modify spectrometer set-up and store spectral data. The data analysis software used was Matlab 8.2 (the USA, Mathworks Company). A total of 800 spectra of normal, frost-damaged seeds (400 spectra for each kind of seed) were measured with the embryo of the seed facing the light source and detector.

Germination test was conducted under a standard experimental environment by professional staff to detect seed viability. Out of 400 frost-damaged seeds, 155 were viable, and 245 nonviable.

2.3. Spectral processing and discrimination model

2.3.1. Spectral processing

The original spectral data size was relatively large, and strong correlations existed among the data from different wavelengths. Dimensionality reduction on the original spectra was applied to reduce information redundancy and improve computation speed. This study comprehensively used PCA, PLS, and OLDA to process the spectra.

PCA is a multivariate technique that analyzes a data table in which observations are described by several inter-correlated quantitative dependent variables. Its goal is to extract the important information from the raw data to represent it as a set of new orthogonal variables called principal components, and to display the pattern of similarity of the observations and of the variables as points in maps. PLS is an analysis method that relates changes in spectral data. As a form of PCA, PLS makes use of the information of the NIR spectrum and the established analyze values associated with the spectrum (Dai et al., 2006; Xie et al., 2009). In PLS, data reduction was conducted creating latent variables which are orthogonal with each other but at the same time describe the response variable (in this case, the class labels). Class labels were entered as a logical array, with each class represented by a column of zeros and ones (this last indicating the membership to one of the classes).

OLDA is a specific algorithm for linear discriminant analysis (LDA), the key property of OLDA is that the discriminant vectors of OLDA are orthogonal to each other, i.e., the transformation matrix of OLDA is orthogonal. LDA is a classical statistical approach for feature extraction and dimension reduction. LDA computes the optimal transformation (projection), which minimizes the within-class distance (of the dataset) and maximizes the between-class distance simultaneously, thus achieving maximum discrimination.

However OLDA requires the total scatter matrix to be non-singular. In spectra analysis, all scatter matrices in question can be singular since the data points are from a very high-dimensional